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[4,5], Vojtěch Vacek[1], Ondřej Brzoň[1], Petr Soukal[1], Laura Eme[2], Joel B. Dacks[4,6], Anna Karnkowska[3], Marek Eliáš[7] & 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] School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; Swiss Institute of Bioinformatics, Lausanne, Switzerland.

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

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

* Contributed equally

[$] Corresponding author

ABSTRACT

Until recently, mitochondria were considered essential organelles impossible to truly lose in a lineage. This view changed in 2016, with the report that the oxymonad *Monocercomonoides exilis*, was the first known eukaryote without any mitochondrion. Questions remain, however, about whether this extends to the entire lineage and how this transition took place. Oxymonadida are a group of gut endobionts of insects, reptiles, and mammals. They are housed in the Preaxostyla (Metamonada), a protistan group that also contains free-living flagellates of genera *Trimastix* and *Paratrimastix*. These latter two taxa harbor conspicuous mitochondrion-related organelles (MROs), while no mitochondria were reported for any oxymonad. Here we report genomic data sets 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 corroborate both the absence of MRO in Oxymonadida and the nature of MROs present in other Preaxostyla and to decipher the evolutionary transition towards amitochondriality and endobiosis. Our results provide insights into the metabolic and endomembrane evolution, but most strikingly the data confirm the complete loss of mitochondria and every protein that has ever participated in the mitochondrion function for all three oxymonad species (*M. exilis*, *B. nauphoetae*, and *Streblomastix strix*) extending the amitochondriate status to the whole Oxymonadida.

INTRODUCTION

Multiple eukaryotic lineages have adapted to low-oxygen and/or endobiotic lifestyles by modifying their mitochondria into a wide range of mitochondrion-related organelle (MRO) types via the process of reductive and adaptive evolution (Roger et al. 2017). The most radically modified MROs are traditionally categorized as hydrogenosomes, producing ATP by extended glycolysis, and as mitosomes with no role in energy metabolism (Müller et al. 2012). However, exploration of a broader diversity of MRO-containing lineages makes it clear that such categorization is artificial, and MROs of various organisms form a functional continuum (Stairs et al. 2015; Klinger et al. 2016; Leger et al. 2017; Brännström et al. 2022).

The extreme outcome of mitochondrial reductive evolution is the complete loss of the organelle, but so far only one organism has been conclusively shown to have reached this point, the chinchilla gut symbiont *Monocercomonoides exilis* (Oxymonadida, Preaxostyla, Metamonada). A genomic project has thoroughly corroborated the amitochondriate status of *M. exilis*, in which it failed to identify any mitochondrion-associated genes while showing multiple other eukaryotic cellular systems to be well represented (Karnkowska et al. 2016; Karnkowska et al. 2019; Treitli et al. 2021). The existence of such an organism implies that mitochondria are not necessary for the thriving of complex eukaryotic organisms, which also has important bearings to our thinking about the origin of eukaryotes (Hampl et al. 2018).

Oxymonadida contain approximately 140 species of morphologically divergent and diverse flagellates exclusively inhabiting digestive tracts of metazoans, of which none has been shown to possess a mitochondrion by cytological investigations (Hampl 2017). It is, therefore, possible that the entire Oxymonadida group is an amitochondriate lineage, and that the mitochondrion was lost in its common ancestor. Oxymonads belong to Preaxostyla, one of the five primary clades of the

67 phylum Metamonada consisting of exclusively anaerobes or microaerophiles; the other clades
68 being represented by Fornicata (e.g., *Giardia intestinalis*), Parabasalia (e.g., *Trichomonas*
69 *vaginalis*), barthelonids (Yazaki et al. 2020) and Anaeramoebidae (e.g., *Anaeramoeba*
70 *flamelloides*; Stairs et al. 2021). The two additional known branches of Preaxostyla, classified as
71 the genera *Trimastix* and *Paratrimastix*, split out at two successive points off the trunk leading to
72 Oxymonadida. They are both comprised of free-living, bacterivorous flagellates exhibiting
73 a typical "excavate" morphology and ultrastructure and thriving in low oxygen environments
74 (Zhang et al. 2015).

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

84 To start answering questions about the timing and circumstances of this single evolutionary
85 experiment in which the mitochondrion was lost from the cell, a denser sampling of "omics",
86 functional and microscopic data from Preaxostyla is needed. Currently, the available data
87 encompasses a genome draft of *M. exilis*, transcriptomes of *P. pyriformis* and *T. marina* with
88 variable completeness, and a fragmentary single-cell genome assembly of an oxymonad
89 *Streblomastix strix*. As a key step towards this end, we present high-quality genomic assemblies

90 for another oxymonad, *Blattamonas nauphoetae*, and for the free-living *P. pyriformis*. The former
91 has enabled us to test the hypothesis of amitochondriality as a shared feature of oxymonads in
92 general, while the latter provided deeper insights into the metabolic capabilities of free-living
93 Preaxostyla represented by *P. pyriformis*. The distribution of mitochondrion hallmark proteins and
94 comparisons of gene repertoires and metabolic functions among five Preaxostyla species of various
95 ecology and MRO status were used to illuminate the adaptations connected to amitochondriality
96 and the origin of endobiotic lifestyle within the group.

RESULTS AND DISCUSSION

De novo genome assemblies are highly contiguous and complete

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

The *B. nauphoetae* genome was assembled into 879 contigs spanning 88,537,989 bp, with an N50 = 199,589 bp and a GC content of 44.96%. Automatic and manual gene prediction resulted in 25,221 predicted protein-coding genes. Using BUSCO v3 (Simão et al. 2015) with the eukaryota_odb9 dataset, the genome completeness was estimated to be 76.6%. The *P. pyriformis* genome was assembled into 633 scaffolds spanning 56,627,582 bp, with an N50 = 276,605 bp and a GC content of 60.94%. Manual and automatic gene prediction resulted in 13,466 predicted protein-coding genes. Using the same method as for the *B. nauphoetae* genome, we estimated the genome completeness to be 82.1%, which is the highest value among all Preaxostyla genomes analyzed (Table 1). It should be noted that in the case of protist genomes, the estimates are not expected to reach 100% simply due to high divergence or true absence of some of the marker genes. For example, various Euglenozoa representatives, including *Trypanosoma brucei*, have completeness estimates in the range of 71-88% (Butenko et al. 2020) and metamonads in the range

119 60-91% (Salas-Leiva et al. 2021), both similar to the values for the two Preaxostyla genomes
120 reported here.

121 **Comparison of gene repertoires within Preaxostyla and contribution of LGT**

122 The data generated in this study together with the previously published data from *M. exilis*
123 (Karnkowska et al. 2019), *S. strix* (Treitli et al. 2019), and *T. marina* (Leger et al. 2017) were
124 exploited to define the common genetic toolkit of Preaxostyla and specific toolkits of its subgroups.
125 We note that the analyses described below rely on a slightly outdated version of the *M. exilis*
126 genome, since they had been initiated before a more complete version (Treitli et al. 2021) became
127 available, and we also acknowledge the limitations of employing the inherently incomplete data
128 from *S. strix* (single-cell genome assembly) and *T. marina* (transcriptome assembly only). We
129 created orthologous groups (OGs) from the protein-coding gene inventories of the five Preaxostyla
130 species and 14 other metamonads (Supplementary file 1) and estimated the core gene set for
131 Metamonada and its clade-specific increments in Preaxostyla and Oxymonadida (Fig. 1,
132 Supplementary file 2). Please note that due to the incomplete nature of all data sets we use the term
133 core gene set in a relaxed sense and included genes present in at least one representative of all main
134 subclades of the taxon (e.g., Fornicata, Parabasalia, and Preaxostyla in the case of Metamonada
135 core gene set), see Material and Methods for details.

136 Each core set can in principle be composed of three different categories of genes: 1) genes inherited
137 by the respective group from its earlier eukaryotic ancestors outside Metamonada, i.e., ancestral
138 eukaryotic genes; 2) genes gained by lateral gene transfer (LGT); and 3) group-specific genes that
139 originated in the respective group, i.e., truly novel genes or extremely diverged descendants of
140 ancient genes, the homology of which was not discerned. To roughly distinguish between these
141 categories within each ancestral set, we have searched for protein homologs (e-value < 10^{-15}) in the

142 NCBI non-redundant protein sequence database and in the EukProt database (Richter et al. 2022),
143 manually investigated phylogenies of proteins with prokaryotic hits, and in parallel employed an
144 LGT detection pipeline. The latter revealed hundreds of putative LGT candidate genes in
145 Preaxostyla (Supplementary file 3), which probably originated from at least 190 LGT events
146 involving prokaryotic donors and affecting different branches of the Metamonada phylogeny
147 (Fig. 1). Of these, 19, 15, 19, 1, and 3 LGT events were mapped to the common ancestor of
148 Metamonada, the Fornicata+Preaxostyla clade, Preaxostyla, the *P. pyriformis*+Oxymonadida
149 clade, and Oxymonadida, respectively, i.e., contributed to the core gene sets of these clades. The
150 numbers of the ancestral LGTs might be underestimated, as they require sufficient sequence
151 similarity between extant proteins to form a clade in a phylogenetic tree and this condition was
152 often not fulfilled.

153 Of the 1,399 OGs from the Metamonada core gene set, 1,398 have homologues outside
154 Metamonada and just a single orthologous group, q2019831, did not have any significant hit from
155 NCBI nr or EukProt database at e-value < 10^{-15}. Interestingly, this protein identified by
156 InterProScan as a member of Major Facilitator Superfamily of transporters is present only in the
157 data set of the parabasalid *Tritrichomonas foetus*, the oxymonad *S. strix* (Streblo_st30458), and the
158 fornicates *Chilomastix cuspidata* and *Dysnectes brevis*. Although it may represent an innovation
159 arising in the common ancestor of Metamonada, its function is clearly not essential. We identified
160 19 LGT events that contributed to the ancestral gene set of Metamonada and further 15 LGT events
161 specific for the Preaxostyla+Fornicata clade (resolved in a recent phylogenomic analysis and at
162 odds with the traditional view placing Preaxostyla as a sister lineage of all other metamonads; Stairs
163 et al. 2021). These include metabolic enzymes used in glycolysis, amino acid metabolism and other
164 functionally unrelated pathways. Some of these cases, e.g., enzymes involved in the arginine

165 deiminase pathway, have been investigated in detail elsewhere (Liapounova et al. 2006; Novák et
166 al. 2016; Karnkowska et al. 2019), while others are proposed here for the first time (Supplementary
167 file 3).

168 11,854 OGs of all investigated metamonads are present only in Preaxostyla and 516 of these are
169 widespread in this group and represent a putative Preaxostyla specific increment to the
170 Metamonada core gene set. 71 OGs seem to be the group-specific genes common for the
171 Preaxostyla lineage or genes too divergent, as they have no close eukaryotic BLAST hits (e-value
172 $< 10^{-15}$) in any other organism. The vast majority of these have vague annotations and their function
173 is simply not known. The remaining 445 OGs have homologs in eukaryotes outside Metamonada.
174 We identified 19 LGT events that contributed to the core gene set of Preaxostyla (Fig. 1,
175 Supplementary file 2 and 3).

176 A well-known example of LGT into Preaxostyla is represented by the proteins involved in the SUF
177 pathway for FeS cluster synthesis (Vacek et al. 2018). Another potential case worth mentioning
178 represents the divergent paralogue of the glycine cleavage system (GCS) L protein (GCS-L). Five
179 GCS-L homologs were identified in the Preaxostyla investigated here: two in *P. pyriformis* and
180 one in *T. marina*, *M. exilis*, and *B. nauphoetae* each. Given the fact that the GCS is
181 a mitochondrion-specific pathway, the presence of the isolated L protein in the amitochondriate
182 oxymonad *M. exilis* draws attention. The identified GCS-L sequences formed two clusters in the
183 phylogenetic analysis (Supplementary file 4). *T. marina* and one of the *P. pyriformis* sequences
184 branch together with sequences from Parabasalia and *Carpediemonas membranifera* with
185 moderate statistical support, suggesting a common metamonad origin. These proteins were
186 previously hypothesized to function inside the MRO (Zubáčová et al. 2013). The other
187 *P. pyriformis* sequence (PAPYR_1328) and the two oxymonad sequences branch with high

188 statistical support together with sequences from two Archamoebae (*Mastigamoeba*
189 *balamuthi* and *Pelomyxa schiedti*) and another eukaryotic anaerobe, *Breviata anathema*
190 (Breviatea), and are here referred to as GCS-L2. Interestingly, in *M. balamuthi* it has been
191 suggested that GCS-L2 protein could function outside the context of the glycine cleavage system
192 (Nývltová et al. 2015), which might be the case also for the three related Preaxostyla sequences,
193 potentially explaining the presence of GCS-L2 protein in the amitochondriate oxymonads. It should
194 be noted that, besides GCS, GCS-L is known to be part of three other protein complexes, pyruvate
195 dehydrogenase, branched-chain amino acid dehydrogenase, and 2-oxoglutarate dehydrogenase
196 (Spalding and Prigge 2010), none of them being present in Preaxostyla. Outside of these typical
197 roles or under specific conditions, GCS-L was shown to have a moonlighting proteolytic activity
198 (Babady et al. 2007) or a diaphorase activity by which it oxidizes NADH using labile ferric iron
199 (Petrat et al. 2003), nitric oxide (Igamberdiev et al. 2004), or ubiquinone (Xia et al. 2001). It is,
200 therefore, possible that one of these moonlighting activities may represent the primary role of GCS-
201 L2 in Preaxostyla. The well-supported relationship between the GCS-L2 sequences from
202 Archamoebae and Preaxostyla may be explained by a eukaryote-to-eukaryote LGT.

203 Putting aside the 19 putative Preaxostyla-specific LGTs, the remaining 425 Preaxostyla-specific
204 increments of the core gene set were likely inherited vertically from the eukaryotic common
205 ancestor but lost or not detected in the other metamonad lineages, and as such they represent an
206 interesting addition to the complexity of the common ancestor of Metamonada revealed thanks to
207 the sampling of Preaxostyla genomes. To provide a single specific example, we focus on the small
208 GTPase Rsg1 (Supplementary file 5). This protein was previously characterized in mammalian
209 cells as a ciliogenesis factor (Agbu et al. 2018; Langousis et al. 2022), but its evolutionary origin
210 and taxonomic occurrence have not been investigated. Having noticed its putative homolog in

211 *P. pyriformis*, we searched publicly available sequence databases to demonstrate that Rsg1
212 orthologs indeed occur outside Metazoa, exhibiting a patchy distribution across distantly related
213 protist lineages (Supplementary file 5). In Metamonada, Rsg1 was found in *P. pyriformis* and
214 *T. marina*, while no orthologs were identified in oxymonads, fornicates, *Barthelona* sp.,
215 parabasalids and *Anaeramoeba* spp., apparently due to multiple independent losses of the Rsg1
216 gene. Notably, Rsg1 orthologs occur only in eukaryotes that can build cilia (Supplementary file 5),
217 indicating that the cilium-associated role defined for the mammalian Rsg1 is likely a general and
218 ancestral functional attribute of the protein. The occurrence of Rsg1 in the free-living Preaxostyla
219 thus points to specific differences in the functioning of their flagella as compared to other
220 metamonads. Furthermore, considering the unsettled position of Metamonada with respect to the
221 root of the eukaryote phylogeny (Derelle et al. 2015), the identification of Rsg1 in metamonads,
222 along other phylogenetically diverse eukaryotes, strengthens the case that the protein is an ancestral
223 component of the eukaryotic cell functioning already in the last eukaryote common ancestor
224 (LECA).

225 Of the 9,513 OGs present only among oxymonads, 8,271 seem to be species-specific and 1,242
226 represent a putative oxymonad-specific increment to the Preaxostyla core gene set (Fig. 1,
227 Supplementary file 2). Of these, 630 OGs seem to be oxymonad-specific or divergent, without
228 close BLAST hits (e-value < 10^{-15}), and may thus represent novel genes evolved in this group
229 providing new functionalities. Quite expectedly, 580 of them, although supported by transcriptome,
230 have no or vague annotation and possibly represent genetic dark matter. The remaining 50 received
231 annotations usually from KEGG or InterProScan, which were likely based on higher e-values than
232 we used as a threshold for our BLAST and comprise a mixture of putative functions noticeably
233 enriched for Fe-S cluster containing proteins and proteins involved in endomembrane transfer. The

234 other 612 OGs have a homolog in eukaryotes outside Metamonada in the NCBI non-redundant
235 protein sequence database or EukProt database (e-value < 10^{-15}). We identified three LGT events
236 that contributed to the ancestral gene set of oxymonads and brought proteins with unclear or
237 mutually unrelated functions. One of them is UTP—glucose-1-phosphate uridylyltransferase
238 involved in glycogenesis and cell wall synthesis (Fig. 1, Supplementary file 3).

239 Based on orthology grouping, most of the genetic innovation is observed on the terminal branches
240 of Preaxostyla rather than in the ancestral nodes. Putting aside the highly incomplete transcriptome
241 of *T. marina* and the incomplete and fragmented genome of *S. strix*, we see that the number of
242 species-specific singletons or OGs varies between 906 in *P. pyriformis* to 2,493 in *B. nauphoetae*.
243 Also, 46 LGT events were mapped to the terminal branches and 2 to internal nodes of Preaxostyla,
244 suggesting that the acquisition of genes is an ongoing phenomenon, and it might be adaptive to
245 particular lifestyles of the species.

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

247 The relatively complete and contiguous genomic and transcriptomic datasets for *P. pyriformis*
248 allowed to compose a more complete and more accurate inventory of the *P. pyriformis* genetic
249 toolkit than the previous studies (Hampl et al. 2008; Zubáčová et al. 2013). Given the fact that this
250 species represents the first free-living representative of the Preaxostyla lineage studied at the
251 genome level, it is not that surprising that our exploration of the *P. pyriformis* genome data revealed
252 a number of unusual or unexpected features of its metabolism, some of which are presented in
253 detail below.

254 *P. pyriformis* encodes a complete pathway required for the biosynthesis of *p*-cresol from tyrosine
255 (Supplementary file 6), which consists of three steps of the Ehrlich pathway (Hazelwood et al.

256 2008) converting tyrosine to 4-hydroxyphenyl-acetate and the final step catalyzed by a fusion
257 protein comprised of 4-hydroxyphenylacetate decarboxylase (HPAD) and its activating enzyme
258 (HPAD-AE). The *P. pyriformis* HPAD-AE protein is closely related to its homolog in the free-
259 living archamoebid *M. balamuthi* (Supplementary file 4), the only eukaryote reported so far to be
260 able to produce *p*-cresol (Nývltová et al. 2017). Interestingly, we identified additional eukaryotic
261 homologs of this fusion protein in the free-living fornicates *Kipferlia bialata* and *Trepomonas* sp.
262 This distribution may be explained by LGT from prokaryotes followed by one or more LGTs
263 between the eukaryotic lineages facilitated by a shared anaerobic habitat. Interestingly, the close
264 endobiotic relatives of these taxa (oxymonads, *Entamoeba histolytica*, parasitic diplomonads such
265 as *Giardia intestinalis* or *Spironucleus salmonicida*) lack this protein. Biosynthesis of *p*-cresol by
266 *M. balamuthi* was hypothesized to confer a competitive advantage against prokaryotes inhabiting
267 the same environment, because this compound inhibits the growth of certain prokaryotes, while no
268 such inhibition was observed in *M. balamuthi* (Nývltová et al. 2017).

269 The presence of demethylmenaquinone methyltransferase / 2-methoxy-6-polyprenyl-1,4-
270 benzoquinol methylase (UbiE) in *P. pyriformis* (Supplementary file 6) is surprising, since the
271 organism lacks any other enzymes involved in the ubiquinone and menaquinone biosynthesis.
272 Interestingly, this enzyme has been shown to have an alternative function, a capability to mediate
273 the synthesis of volatile methylated selenium (Se) compounds, namely dimethyl selenide and
274 dimethyl diselenide, in bacteria and plants (Swearingen et al. 2006; Zhou et al. 2009). Selenium is
275 a micronutrient that becomes toxic in high concentrations. Biological Se volatilization is an
276 important detoxification process that converts inorganic Se compounds into volatile organic
277 compounds. We propose that Se volatilization might be the physiological function of the
278 *P. pyriformis* UbiE. This enzyme is another noteworthy case of prokaryotic LGT within the

279 *P. pyriformis* genome, most likely from a verrucomicrobial source (Supplementary file 4).
280 Interestingly, UbiE homologs occur also in some additional metamonads, including the oxymonad
281 *B. nauphoetae* and certain fornicates (Supplementary file 4), but these seem to represent
282 independent LGT events from different bacterial sources. The recurrent acquisition of UbiE by
283 different eukaryotic anaerobes attests to the physiological significance of the enzyme.

284 *P. pyriformis* has a complete pathway for biosynthesis of NAD^+ from L-aspartate and
285 dihydroxyacetone phosphate consisting of L-aspartate oxidase (NadB), a nicotinate-nucleotide
286 pyrophosphorylase-quinolinate synthase fusion protein (NadC-NadA), nicotinate-mononucleotide
287 adenylyltransferase (NadD), DNA-binding transcriptional repressor/NMN adenylyltransferase
288 (NadR), and NAD^+ synthase (NadE); at least some of these enzymes are also present in *T. marina*
289 (Supplementary file 6). The presence of this pathway in anaerobes is surprising because the
290 standard NadB reaction involves molecular oxygen (O_2). However, NadB can alternatively use
291 fumarate instead of oxygen (Sakuraba et al. 2002). If this is the case also in *P. pyriformis*, it would
292 explain the presence of fumarate hydratase (FH) in this organism: fumarate produced by FH from
293 malate could serve as an electron acceptor for NadB. Phylogenetic analyses of NadA, NadB, NadC,
294 and NadR show the Preaxostyla sequences branch separately from other eukaryotic sequences and
295 have an affinity to Chloroflexi (NadA, NadC), and Lentisphaerae (NadB) hinting at LGT origin
296 from prokaryotes. In the case of NadB, this LGT is likely shared with *M. balamuthi*, which occupies
297 a similar position inside Lentisphaerae. Only the Preaxostyla sequences of NadD and NadE branch
298 together with other eukaryotes (Supplementary file 4).

299 Uniquely among eukaryotes, *P. pyriformis* may possess enzymes for mercury (Hg) methylation.
300 Production of highly toxic methyl mercury is known in some Bacteria and Archaea (Podar et al.
301 2015; Cooper et al. 2020). The exact reason why these microorganisms methylate Hg is unknown,

302 however, it is widely accepted that it is not a detoxification process as these are not less susceptible
303 to Hg (Gilmour et al. 2011). In prokaryotes, the Hg methylation is performed by two enzymes
304 HgcA and HgcB encoded by two independent genes, although they are often found adjacent to
305 each other in a single operon (Christensen et al. 2016). In contrast, *P. pyriformis* contains a single
306 fusion protein HgcAB (Supplementary file 6). As the order of the two parts reflects the arrangement
307 of the genes in bacteria, both were likely acquired *en bloc* by lateral transfer of a whole operon.

308 The structure of the prokaryotic HgcAB complex was recently determined (Cooper et al. 2020).
309 Comparison of PAPYR_7512 to HgcA and HgcB from *Pseudodesulfovibrio mercurii* and
310 a selection of other prokaryotes (Fig. 2) revealed remarkable similarity and conservation in the
311 HgcA part, which is predicted to contain a corrinoid binding domain (CMD) including the well-
312 conserved corrinoid binding alpha-helix and cobalamin binding residues (Thr60, Thr66, Val91,
313 Cys93, and Ala153 in *P. mercurii*), an N-terminal transmembrane domain (with four
314 transmembrane domains in contrast to five in *P. mercurii*), and a conserved motif
315 (NVWCAAGKG) at the positions 97 – 105. The *P. mercurii* HgcB shares with the *P. pyriformis*
316 sequence the same twofold pseudosymmetry as the bacterial 4Fe-4S ferredoxin and two strictly
317 conserved CxxCxxCxxxCP 4Fe-4S binding motifs, conserved C-terminal cysteine residues, and
318 a cysteine residue (Cys73 in *P. mercurii*) at the beginning of an alpha-helix. Overall PAPYR_7512
319 shows high similarity to HgcAB complex and contains all important residues necessary for
320 methylation of Hg, a function that has not been reported in any other eukaryote (Fig. 2).

321 Another intriguing gene fusion was found between genes for two histidine-processing enzymes:
322 histidine ammonia-lyase and histidyl-tRNA synthase (HAL-HARS). HAL is responsible for the
323 first step in the catabolic pathway converting histidine to glutamine. Histidyl-tRNA synthases with
324 an N-terminal extension homologous to HAL have been previously found in land plants and green

325 algae (Saga et al. 2020). However, the plant HAL-HARS sequences lack a crucial catalytic site of
326 the HAL enzyme (MIO cofactor formed from three conserved amino acids) and are therefore
327 considered non-functional. The *P. pyriformis* sequence (Supplementary file 6), on the other hand,
328 has these three amino acids conserved. The existence of this fusion of the two enzymes that both
329 have histidine as substrate in two distant eukaryotic lineages suggests there might be a functional
330 connection between histidine ammonia-lyase and histidyl-tRNA synthase, which deserves further
331 investigation.

332 Altogether, the gene inventory of *P. pyriformis* demonstrates that this protist possesses several
333 metabolic capabilities which are very rare or so far unseen in eukaryotes and that LGT is a common
334 means of their acquisition.

335 **Membrane transporter complement may reflect adaptation to an endobiotic lifestyle in**
336 **oxymonads**

337 Though most attention is paid to the divide between MRO-possessing and putatively MRO-lacking
338 Preaxostyla, at least two other transitions can be examined using this set of species. The first is
339 from a free-living to a symbiotic lifestyle. As an assessment of this, we examined proteins
340 responsible for the transport of metabolites and other chemicals across the plasma membrane and
341 other cell membranes, as changes in their repertoire and capabilities may reflect evolutionary
342 transitions in both metabolic capabilities and the lifestyle of the organism. We searched for a broad
343 spectrum of transmembrane transporters (except for transporters of MROs) using homology-based
344 methods in order to compare the repertoire of functional types as well as the number of paralogues
345 between the five species of Preaxostyla (Supplementary file 7).

346 Just like in *T. vaginalis*, the most gene-rich group of membrane transporters identified in
347 Preaxostyla is the ATP-binding cassette (ABC) superfamily represented by MRP and pATPase
348 families. Altogether, representatives of 19 transporter families have been identified in Preaxostyla.
349 All of them are present in *P. pyriformis* and *T. marina*, while four families (PotE, SPNS, RFC, and
350 TauE) are missing in all three oxymonad species investigated. On the other hand, transporters of
351 nucleosides, sugars, amino acids, choline, and phospholipids have consistently higher numbers of
352 paralogues in the genomic datasets of the oxymonads *M. exilis* and *B. nauphoetae* than in
353 *P. pyriformis*. These two observations parallel findings of functional domain loss and expansion
354 by gene duplication of transporter families in Microsporidia (Nakjang et al. 2013), possibly hinting
355 at a broader evolutionary pattern at the transition to the endobiotic lifestyle.

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

358 Another evolutionary transition addressable by our results is the presence of a morphologically
359 identifiable Golgi body in *P. pyriformis* + *T. marina* versus the lack in oxymonads (O'Kelly et al.
360 1999; Zhang et al. 2015; Treitli et al. 2018). Nonetheless, a substantial complement of proteins
361 associated with Golgi membrane-trafficking and transport has been previously reported,
362 confirming at least the genomic signal for a Golgi body, in the oxymonad *M. exilis* (Karnkowska
363 et al. 2016; Karnkowska et al. 2019). Therefore, our dataset represents one of the tightest samplings
364 to date where genomic data is available for the closest known species on both sides of the divide
365 between stacked and unstacked Golgi morphology. Consequently, we searched in our datasets to
366 assess the Golgi complement in the additional Preaxostyla representatives, particularly to see
367 whether the complement was more extensive in the organisms possessing stacked Golgi bodies.

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

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

391 adhesive golgin-type proteins that regulate stacking rather than the identity of any given Golgi-
392 stacking protein (Barlow et al. 2018).

393 **Carbon and energy metabolism adaptation in the common Preaxostyla ancestor**

394 The most prominent evolutionary transition addressed in our dataset of course is that of the MROs.
395 Before examining this directly, we delved into the associated cellular metabolism in the
396 preaxostylan species in order to establish a baseline metabolic context for function of the different
397 organelles. Anaerobically living eukaryotes often use a modified glycolytic pathway that
398 incorporates pyrophosphate (PPi)-dependent instead of ATP-dependent enzymes. This alternative
399 form of glycolysis can produce three ATP molecules and five high-energy phosphate bonds per
400 one glucose molecule, in comparison to only two ATP and two high-energy phosphate bonds in
401 the "classical" pathway known from aerobes (Mertens 1993). Multiple alternative glycolytic
402 enzymes have been previously identified in *P. pyriformis*, *M. exilis*, and *S. strix* (Liapounova et al.
403 2006; Slamovits and Keeling 2006; Stechmann et al. 2006).

404 Here, we report the identification of both phosphofructokinase (PFK) and diphosphate-fructose-6-
405 phosphate 1-phosphotransferase (PFP) catalyzing the 3rd step of glycolysis in *P. pyriformis* and
406 *T. marina*, while only PFP is present in the three oxymonad species (Fig. 4, Supplementary file 6),
407 suggesting that the alternative enzyme PFP has been acquired by a common ancestor of
408 Preaxostyla. Both pyruvate kinase (PK) and pyruvate-phosphate dikinase (PPDK) catalyzing the
409 last step have been identified in all five species studied. In the case of the 4th and 8th steps,
410 alternative versions of the enzymes catalyzing these reactions, namely Class II fructose
411 biphosphate aldolase (FBA) and 2,3-bisphosphoglycerate independent phosphoglycerate mutase
412 (iPGM), respectively, were identified in all five studied species, while no "classical" enzymes for
413 these reactions have been found. This further emphasizes that the glycolytic pathway in Preaxostyla

414 is an evolutionary mosaic composed of enzymes of different origins (Fig. 4), likely acquired in
415 adaptation to the anaerobic lifestyle.

416 Many eukaryotic anaerobes, including Preaxostyla, generate ATP using the extended glycolysis
417 pathway, which produces acetate, CO_2, and H_2 from pyruvate while performing substrate-level
418 phosphorylation of ADP to ATP (Lindmark and Müller 1973). The pathway uses pyruvate as
419 a substrate, which is either directly sourced from the cytosolic glycolysis or produced by
420 decarboxylation of malate through the activity of malic enzyme (ME). This enzyme has been
421 identified in all five Preaxostyla species. All Preaxostyla apparently rely on oxidative
422 decarboxylation of pyruvate to acetyl coenzyme A (acetyl-CoA) and CO_2 in a reaction catalyzed
423 by pyruvate:ferredoxin oxidoreductase (PFO; Williams et al. 1987), owing to the identification of
424 three to six PFO isoforms in each species analyzed (Fig. 4, Supplementary file 6). On the other
425 hand, none of the species encodes any of the two other alternative enzymes mediating the
426 conversion of pyruvate to acetyl-CoA in various other eukaryotic anaerobes, pyruvate:NADP+
427 oxidoreductase (PNO) and pyruvate formate lyase (PFL). Both the decarboxylation of malate by
428 ME and of pyruvate by PFO are oxidative processes that release electrons, producing NADH and
429 reduced ferredoxin, respectively.

430 The final fate of the electrons carried by ferredoxin often lies in the reduction of protons to
431 molecular hydrogen through the activity of [FeFe] hydrogenases (HydA; Payne et al. 1993). In
432 addition to the "simple" hydrogenases, which are present in all species of Preaxostyla,
433 [FeFe] hydrogenases with N-terminal homology to the NuoG subunit of NADH-quinone
434 oxidoreductase and a C-terminal homology to NADPH-dependent sulfite reductase (CysJ), were
435 identified in *T. marina* and *P. pyriformis* (Fig. 5). Similar "fused" hydrogenases (Fig. 5) have been
436 previously reported in other eukaryotic anaerobes, including *T. vaginalis* (Tachezy and Doležal

437 2007), the breviate *Pygsuia biforma* (Stairs et al. 2014), and the jakobid *Stygiella incarcerata*
438 (Leger et al. 2016). Although they do not belong to the group of A3 trimeric hydrogenases
439 (Greening et al. 2016) known to be capable of NADH oxidation via electron confurcation (Schut
440 and Adams 2009), they were hypothesized to catalyze NAD(P)H-dependent formation of H_2
441 (Tachezy and Doležal 2007). However, experimental evidence is still missing. Ferredoxin (Fd)
442 serves as an electron intermediate between PFO and HydA. Both Fd and unique
443 flavodoxin-ferredoxin fusion proteins (Fld-Fd) were identified in all five species (Supplementary
444 file 6).

445 The second reaction of the extended glycolysis, which yields ATP, acetate, and CoA, can be
446 catalyzed either by a two-enzyme system consisting of acetate:succinate CoA-transferase (ASCT)
447 and succinyl CoA synthetase (SCS) like in *T. vaginalis*, or by a single enzyme acetyl-CoA
448 synthetase (ACS) like in *G. intestinalis*. All five Preaxostyla species contain ACS, while neither
449 ASCT nor SCS were identified. ACS has a complicated evolutionary history in Metamonada
450 characterized by multiple LGT events and gene losses (Leger et al. 2017; Yazaki et al. 2020;
451 Vargová et al. 2022). The majority of ACS homologs in Metamonada are predicted to function in
452 the cytosol, with the only exception of the ACS2 isoform from *S. salmonicida* which functions in
453 the MRO (Jerlström-Hultqvist et al. 2013; Leger et al. 2017). Phylogenetic analysis of Preaxostyla
454 ACSs (Supplementary file 4) shows four unrelated clades, none in close relationship to the
455 *S. salmonicida* MRO homolog, suggesting a cytosolic localization of these enzymes in Preaxostyla.

456 **Amino acid metabolism is more complete in *P. pyriformis* and *T. marina* than in oxymonads**

457 While some aspects of the core organellar metabolism seemed similar across Preaxostyla, some
458 clear differences were observed, particularly in amino acid metabolism. *P. pyriformis* predicted
459 proteome contains 53 enzymes putatively involved in amino acid metabolism and seven more

460 enzymes putatively involved in folate metabolism and one-carbon pool, which are closely
461 connected to the amino acid metabolism (Ducker and Rabinowitz 2017). This is a higher number
462 than in *T. vaginalis*, which has the most complex predicted amino acid metabolism among
463 metamonads studied so far, containing 39 enzymes (Carlton et al. 2007). The *T. marina* predicted
464 proteome contains 31 enzymes of amino acid metabolism and six of folate and one-carbon
465 metabolism, but the lower number compared to *P. pyriformis* may at least partly reflect the different
466 nature of the sequence data available (transcriptome assembly only). The number of enzymes
467 involved in the amino acid metabolism in oxymonads is 31 in *M. exilis*, 24 in *B. nauphoetae*, and
468 22 in *S. strix* (Supplementary file 6).

469 Hypothetical metabolic maps of the amino acid metabolism were reconstructed based on the
470 metabolic maps in the KEGG database (Kanehisa et al. 2014) and catalytic activities of enzymes
471 reported from other metamonads (Carlton et al. 2007; Xu et al. 2014). The reconstructed
472 metabolism (Supplementary figs. 1-3) shows a possibility for *de novo* biosynthesis from glycolytic
473 intermediates of at least five protein-building amino acids in *P. pyriformis* (cysteine, serine,
474 glycine, threonine, and selenocysteine), only one in *T. marina* (cysteine), which is probably caused
475 by incomplete data, and three in *M. exilis* (cysteine, serine, and selenocysteine).

476 Methionine (in the form of S-adenosylmethionine, SAM) consumed in reactions catalyzed by
477 SAM-dependent methyltransferases can be regenerated in *P. pyriformis* and *T. marina* thanks to
478 the methionine cycle, which is absent in oxymonads. The selenocysteine biosynthesis pathway
479 present in *P. pyriformis*, *T. marina*, and *M. exilis* is notable, as the capacity to synthesize this amino
480 acid has been reported only in *S. salmonicida* among other metamonads studied so far (Xu et al.
481 2014).

482 Like many other metamonads (Novák et al. 2016), *M. exilis* has been suggested to utilize arginine
483 for ATP production via the arginine deiminase pathway consisting of three enzymes: arginine
484 deiminase (ADI), carbamate kinase (CK), and ornithine transcarbamylase (OTC). This important
485 metabolic capability has been probably formed in the common Metamonada ancestor by the
486 acquisition of genes for ADI and OTC via LGT (Supplementary file 2). Here we show the presence
487 of the complete arginine deiminase pathway also in *T. marina* and *B. nauphoetae*, while
488 *P. pyriformis* and *S. strix* lack ADI and CK, respectively. This suggests that ATP production via
489 arginine catabolism is widespread among Preaxostyla, and it is present in both free-living and
490 endobiotic representatives. Other amino acids can be used in energy metabolism as well: cysteine,
491 serine, and tryptophan can be converted to pyruvate, while methionine can be converted to α-keto-
492 butyrate. Both products can be used by PFO and ACS to generate ATP (Anderson and Loftus
493 2005).

494 The presence of a complete glycine cleavage system (GCS) in *P. pyriformis* and *T. marina* is
495 connected to the presence of complete folate and methionine cycles (Zítek et al. 2022). The methyl
496 residue liberated from glycine by the activity of GCS enters the connected folate and methionine
497 cycles and can be later utilized in a multitude of metabolic pathways requiring one-carbon units,
498 e.g., remethylation of homocysteine to form the amino acid methionine via MetH (Ducker and
499 Rabinowitz 2017). The transsulfuration pathway, associated with the folate and methionine cycles
500 in mammals (Stipanuk 1986), was not found in any Preaxostyla species. It is possible that the loss
501 of the folate and methionine cycles in oxymonads is related to the loss of mitochondria because it
502 is connected with mitochondria via GCS, an exclusively mitochondrial pathway, which was indeed
503 found to be localized in the MRO of *P. pyriformis* (Zubáčová et al. 2013; Zítek et al. 2022).

504 **Oxymonads predominantly use 4Fe-4S clusters in their iron-sulfur clusters containing**
505 **proteins**

506 Fe-S clusters are one of the most ancient and versatile inorganic cofactors and are present in
507 virtually all living organisms. Thanks to their ability to delocalize electrons over the Fe-S bond
508 (Noodleman and Case 1992; Glaser et al. 2000) they are ideally suited for mediating electron
509 transfer and therefore are part of important biological pathways, such as the respiratory chain,
510 photosynthesis, DNA metabolism, ribosome biogenesis, and many others. In a typical eukaryotic
511 cell, Fe-S clusters are synthesized by a combination of the mitochondrial ISC and the cytosolic
512 CIA pathway, but members of Preaxostyla uniquely combine the CIA pathway with the SUF
513 pathway secondarily acquired from bacteria to perform the same task (Vacek et al. 2018; Braymer
514 et al. 2021). Although functional details of the Fe-S cluster biogenesis in this group are mostly
515 unknown, it was shown to run wholly in the cytosol (Zítek et al. 2022), and this major shift was
516 likely a preadaptation for the loss of the mitochondrion in the lineage leading to *M. exilis*
517 (Karnkowska et al. 2016).

518 To assess how much the change of the Fe-S cluster assembly pathway affected the inventory of Fe-
519 S cluster-containing proteins in Preaxostyla, we predicted these proteins from *in silico* proteomes
520 of *P. pyriformis*, *B. nauphoetae, T. marina,* and *S. strix* and compared them with the previously
521 predicted (Karnkowska et al. 2019) Fe-S proteins of *M. exilis* (Supplementary file 8). The numbers
522 of Fe-S cluster-containing proteins identified in individual species varied from 48 in *T. marina* to
523 93 in *P. pyriformis* and are thus not decreased in comparison to other heterotrophic eukaryotes
524 (Andreini et al. 2016). Predicted Fe-S proteins fell into distinct 164 OGs. Numbers of these OGs
525 shared among Preaxostyla species are shown in Supplementary Fig. S4.

526 Most of the predicted Fe-S proteins in oxymonads contain 4Fe-4S clusters, with the exception of
527 xanthine dehydrogenase (XDH), which contains both 2Fe-2S and 4Fe-4S clusters and is present in
528 all Preaxostyla in two (*S. strix, B. nauphoetae, T. marina*) or three copies (*M. exilis* and *P.*
529 *pyriformis*). The free-living *P. pyriformis* furthermore contains several proteins with 2Fe-2S
530 clusters only, such as [FeFe] hydrogenases and 2Fe-2S ferredoxin. The higher number of 2Fe-2S
531 proteins in *P. pyriformis* compared to the oxymonads may reflect the presence of the MRO in this
532 organism.

533 The most widespread types of Fe-S cluster-containing proteins in Preaxostyla fall into expected
534 functional groups acting in processes of extended glycolysis and electron transfer (PFO, HydA,
535 ferredoxins, and flavodoxin-ferredoxin domains containing proteins), DNA replication and repair,
536 transcription and translation (e.g., DNA and RNA polymerases, Rli1p), Fe-S cluster assembly itself
537 (SufDSU, SufB, Nbp35, and NAR-1), nucleotide (XDH), and amino acid metabolism (L-serine
538 dehydratase) etc. An exception represents a radical SAM enzyme (OG q2001674) that is
539 homologous to a member of a three-gene system AmmeMemoRadiSam (Balaji and Aravind 2007;
540 Burroughs and Aravind 2014; Burroughs et al. 2019). Notably, homologs of the other two members
541 of this system are also present in all Preaxostyla and most metamonads, suggesting a functional
542 significance. Each AmmeMemoRadiSam protein of Preaxostyla is affiliated with homologs from
543 other eukaryotes (Supplementary file 4). Given the fact that the function of these genes in
544 prokaryotes is unclear, we do not know what function they could perform in metamonads.

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

547 The uniqueness of Preaxostyla lies in the fact that it includes species harboring MROs, exemplified
548 by *T. marina* and *P. pyriformis* (Hampl et al. 2008; Zubáčová et al. 2013; Leger et al. 2017), and

549 at least one species lacking any form of a mitochondrion, i.e., *M. exilis* (Karnkowska et al. 2016).
550 The hypothesis on the absence of the organelle in *M. exilis* was postulated based on the absence of
551 mitochondrion-related proteins in its genome and transcriptome and led to the twist in the paradigm
552 of mitochondria as ubiquitous among eukaryotes (Karnkowska et al. 2016). Obviously, the
553 amitochondriate status of *M. exilis* is immediately falsifiable by finding any evidence of a putative
554 organelle in this species. Careful inspection of the predicted proteomes of other oxymonads for the
555 presence of mitochondria is an obvious next step that may further support, or weaken, this
556 hypothesis.

557 We used several different methods to identify candidates for mitochondrial proteins (results for all
558 searches are shown in Supplementary file 9 and summarized in Fig. 6). In all cases, proteins
559 identified in the first step ("hits") were reciprocally searched against the MitoMiner (Smith et al.
560 2012; Smith and Robinson 2018) and NCBI databases. We considered proteins yielding results of
561 reciprocal blasts as "candidate" mitochondrial proteins, which were further manually verified with
562 extensive searches (see Materials and Methods). Proteins passing this third step were classified as
563 validated mitochondrial proteins.

564 We searched all five investigated species for homologs of nuclear genome-encoded proteins
565 typically associated with mitochondria or MROs in other eukaryotes. Initially, we used profile
566 Hidden Markov Models (HMMs) to search for components of the mitochondrion protein import
567 and maturation machinery, considered one of the most conserved mitochondrial features (Pfanner
568 et al. 2019). Homology searches resulted in 22 (*T. marina*) up to 63 (*M. exilis*) hits, which were
569 further evaluated by reciprocal searches against the MitoMiner and NCBI databases resulting in 57
570 candidates. Manual verification indicated that all the candidates recovered in oxymonad data sets
571 are false positives (Fig. 6, Supplementary file 9). For *P. pyriformis* and *T. marina*, the situation

572 was different. Most of the previously identified components of the *P. pyriformis* mitochondrion
573 protein import and maturation machinery (Zubáčová et al. 2013; Zítek et al. 2022) were found in
574 the predicted proteomes: the β-barrel outer membrane translocases Tom40 and Sam50, the inner
575 membrane translocase Tim17 and its associated protein Pam18, the α and β subunit of the
576 mitochondrial processing peptidase (MPP), and the chaperone protein Cpn60 and Cpn10. Tim17
577 and Hsp70, were also identified in the *T. marina* dataset, corroborating previous findings (Leger et
578 al. 2017). The successful identification of protein translocon components in these two species
579 validated our approach.

580 MROs export or import ATP and other metabolites typically using transporters from the
581 mitochondrial carrier family (MCF) or sporadically (Tsaousis et al. 2008; Major et al. 2017) by the
582 bacterial-type (NTT-like) nucleotide transporters. We did not identify any homolog of genes
583 encoding known mitochondrial metabolite transport proteins in any of the three oxymonads
584 investigated. In contrast, the mitochondrial carrier (MCF) proteins were recovered in the number
585 of four for each *P. pyriformis* and *T. marina* (Supplementary file 6).

586 The sensitive homology-based HMM searches were complemented by an extensive search for
587 putative homologues of known mitochondrial proteins using a pipeline based on the MitoMiner
588 database (Supplementary file 9), which was enriched with identified mitochondrial proteins of
589 diverse anaerobic eukaryotes with MROs (see Materials and Methods). As already shown for *M.*
590 *exilis*, the specificity of the pipeline in organisms with divergent mitochondria is low (Karnkowska
591 et al. 2016). Many of the selected candidates were assigned as proteins that are obviously not
592 mitochondrial (e.g., histones or ribosomal proteins) or belonging to very general GO categories
593 (e.g., protein-binding). The search recovered a comparable number of proteins as candidates for
594 functions in a putative mitochondrion in all investigated species, from 636 candidates in *S. strix* up

595 to 1,025 candidates in *P. pyriformis*. While no reliable mitochondrial protein was found among the
596 oxymonad candidates, for *P. pyriformis* and *T. marina* we identified a list of proteins previously
597 known to be localized in their MROs as well as some new candidates, such as mitochondrial fission
598 protein 1 (Fis1; Gene.12785::gnl|Trimastix_PCT|4043), which is a model TA protein mediating
599 mitochondrial fission (van der Bliek et al. 2013). The phylogenetic pattern of dynamin-related
600 proteins corroborates the finding of Karnkowska et al. (2019) showing a specific
601 *Paratrimastix*+*Trimastix* clade potentially mediating the division of MROs in these species
602 (PAPYR_3413 and Gene.668::gnl|Trimastix_PCT|268, Supplementary file 4). In addition, both *T.*
603 *marina* (Gene.3674::gnl|Trimastix_PCT|1191) and *P. pyriformis* (PAPYR_2826) but none of the
604 oxymonads exhibit an ortholog of the mitochondrial outer membrane-anchored protein MIRO
605 ("mitochondrial Rho"; Eberhardt et al. 2020) broadly conserved in eukaryotes but rare in MRO-
606 containing taxa (Vlahou et al. 2011; Gentekaki et al. 2017); recently it was confirmed by
607 proteomics to be MRO-associated in the latter species (Zítek et al. 2022).

608 As an alternative to homology searches, we also searched for several types of signature sequences
609 typical for mitochondrion-targeted proteins. The matrix proteins of mitochondria and MROs are
610 expected to contain characteristic N-terminal targeting signals (NTS) needed for the targeted
611 import into mitochondrion-related organelles (Dolezal et al. 2006). As we have previously shown,
612 even in the amitochondriate *M. exilis*, a fraction of proteins (though less than 1%) contains NTS
613 recognized by the subcellular localization predicting tools (Karnkowska et al. 2016). It has been
614 previously recognized that the presence of a predicted NTS by itself does not prove the targeting,
615 as such amino acid sequences can also appear by chance (Lucattini et al. 2004). Here we used up-
616 to-date tools for targeting signal prediction and identified from two hits in *S. strix* up to 248 hits in
617 *B. nauphoteae* (Supplementary file 9). Oxymonad species contain a lower proportion of proteins

618 with the putative mitochondrial NTS (always below 1%) than *Trimastix* and *Paratrimastix* and
619 most of them gave no hits in MitoMiner and the NCBI nr database; the few candidates with good
620 hits matched to proteins unlikely to function in mitochondria. If the NTS in the protein is functional
621 and not formed by chance, it is expected that a NTS will be detected also in its orthologues from
622 other species. With this assumption, we have focused on candidates of which more Preaxostyla
623 orthologues were predicted to contain NTS. Only four proteins from oxymonads (ribosomal protein
624 L21, L23a, and L34e, and tRNA-dihydrouridine synthase) fulfilled this criterion, none of them
625 representing a reasonable mitochondrial protein in an organism that for sure lacks a mitochondrial
626 genome. Based on those results, we assume that all NTS predictions on oxymonad proteins are
627 false positives. In *P. pyriformis* and *T. marina* the proportion of proteins with predicted NTS is
628 higher (1.5% and 2.7% respectively), which might reflect the presence of the organelle. Indeed, we
629 identified among them, for example, aminoadipate-semialdehyde dehydrogenase and the GCS-H
630 protein in *T. marina*, both previously suggested to localize to the MRO (Leger et al. 2017).

631 The outer mitochondrial membrane accommodates two special classes of proteins, tail-anchored
632 (TA) proteins and mitochondrial β-barrel membrane proteins (MBOMPs), the former using
633 specific C-terminal signals (Borgese et al. 2007; Denic 2012; Rada et al. 2019). We identified up
634 to 475 TA hits in the predicted proteome of *S. strix*, and around 100 for the rest of the species, with
635 only 25 for *T. marina* (Supplementary file 9). Depending on the species only five to twelve were
636 considered as candidates with homologs in the MitoMiner database and the majority of these were
637 Golgi apparatus/ER-related. A subset of the TA proteins corresponded to 22 OGs that were present
638 in an oxymonad and at least one more species of Preaxostyla. These were mostly without
639 MitoMiner hits and in most cases annotated as SNARE or other proteins involved in vesicle
640 trafficking. This suggests that the TA prediction did produce true positives, but they all likely

641 represent non-mitochondrial TA proteins. The only probable mitochondrial candidate is Fis1
642 identified in the *T. marina* predicted proteome.

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

653 Finally, given the limitations of the previous searches, namely, high false-positive rates, uncertainty
654 about mitochondrial localization of proteins included in the MitoMiner database, and the inability
655 to distinguish between the cytosolic and mitochondrial isoforms, we have used also a more curated
656 approach and searched for orthologues of proteins from the experimentally well-established
657 mitochondrial proteome of *T. brucei* (Panigrahi et al. 2009; Dean et al. 2017; Peikert et al. 2017).
658 Reciprocal BLAST searches of the *T. brucei* proteins against the predicted proteins of Preaxostyla
659 revealed ~200 putative groups of orthologues and those were investigated manually. Careful
660 inspection of the raw localization data in *T. brucei* and protein phylogenies (Supplementary file 10)
661 rejected most cases by disputing either the mitochondrial localization in *T. brucei* or the orthology.
662 The gene orthology was considered uncertain if *T. brucei* and Preaxostyla proteins were separated

663 by prokaryotic or non-mitochondrial eukaryotic homologs in the phylogenetic trees. The 30 cases
664 surviving the manual scrutiny (Supplementary file 9) were divided into two groups.

665 The high confidence group comprises 17 proteins that likely share mitochondrial origin as they
666 form a monophyletic cluster with other mitochondrial homologs. These were present only in
667 *P. pyriformis* and/or *T. marina* and never in oxymonads. Nine out of these 17 proteins were already
668 predicted to be mitochondrial previously (Zubáčová et al. 2013; Zítek et al. 2022). The remaining
669 eight proteins, such as L-threonine 3-dehydrogenase (TDH) or 2Fe-2S ferredoxin (Fd) thus have a
670 high potential for being novel mitochondrial proteins in those two protists.

671 The 13 putative mitochondrial proteins of the low confidence group cluster with mitochondrial
672 proteins in other eukaryotes; however, the mitochondrial origin is not well supported. They either
673 cluster also with cytosolic and peroxisomal isoforms, or the enzymes are known to be dually
674 localized. An example of the former is isocitrate dehydrogenase, for which the yeast cytosolic,
675 peroxisomal, and mitochondrial proteins form a clade in the tree. Examples of the latter are
676 aconitase and alanine aminotransferase, for which a dual cytosolic/mitochondrial location was
677 reported for various eukaryotes including trypanosomatids (Duschak and Cazzulo 1991; Saas et al.
678 2000). This low confidence group contains nine proteins present also in oxymonads
679 (Supplementary file 9), and these are discussed in more detail below.

680 Aspartate and alanine aminotransferases and the two tRNA synthetases in this category are all
681 dually localized in the mitochondrion and cytosol of *T. brucei*. In the case of tRNA synthetases,
682 the dual localization arose specifically in trypanosomatids (Peikert et al. 2017) and, consistently
683 with this hypothesis, Preaxostyla and *T. brucei* proteins cluster together with cytosolic tRNA
684 synthetases of other eukaryotes. In the case of aspartate aminotransferase, the dual localization
685 seems to be of deeper evolutionary origin (Yagi et al. 1990). The two mevalonate pathway enzymes

686 (hydroxymethylglutaryl-CoA synthase and 3-hydroxy-3-methylglutaryl-CoA reductase) are
687 localized in the mitochondrion of kinetoplastids (Peña-Diaz et al. 2004), however, this is again a
688 specialty of this lineage, as eukaryotes typically run this pathway in the cytosol and the ER. In all
689 these cases, we conservatively assume that these enzymes were localized in the cytosol in the
690 common ancestor of Preaxostyla and *T. brucei* and were partially or fully moved to the
691 mitochondrion in the lineage leading to *T. brucei*. The last candidate is the malic enzyme. The
692 evolution of this enzyme family is complicated, and there are two homologs of the enzyme in
693 kinetoplastids, one cytoplasmic and one mitochondrial, which cluster together (Supplementary
694 file 4). Again, it cannot be inferred what was the situation in the common ancestor with Preaxostyla.

695 In summary, our systematic searches for MRO proteins allowed us to update the predicted set of
696 MRO-localized proteins of *P. pyriformis* and *T. marina* and to provide updated predictions of their
697 MRO proteomes (Supplementary file 6). Notably, no reliable candidate for an MRO protein was
698 detected in any of the oxymonad data sets, supporting the hypothesis of the absence of the
699 mitochondrion in oxymonads as a group (Fig. 6).

700 **No evidence for subcellular retargeting of ancestral mitochondrial proteins in oxymonads**

701 Several studies demonstrated the presence of genes of alphaproteobacterial origin in non-
702 mitochondrial compartments of eukaryotes. For example, it was proposed that more than 50% of
703 the peroxisomal proteome is of alphaproteobacterial origin (Bolte et al. 2015). Similarly, some
704 cytosolic enzymes such as the cytosolic uroporphyrinogen synthase participating in heme
705 synthesis, or several glycolytic enzymes are also proposed to be of alphaproteobacterial origin
706 (Martin and Müller 1998). A widely accepted explanation for these observations is that these genes
707 originated by endosymbiotic gene transfer from an ancestor of mitochondria during eukaryogenesis
708 (Martin 2010). This assumes that relocation of proteins from endosymbiont/mitochondria into

709 other cell compartments is a simple process, in which the loss of protein-localization information
710 either by the gene transfer to the nucleus or by the removal of the targeting peptide on the nuclear-
711 encoded mitochondrial protein represents the only obstacle.

712 While a substantial effort was made to assess the degree of the "early" endosymbiotic gene transfer
713 before last eukaryotic common ancestor (Thiergart et al. 2012; Pittis and Gabaldón 2016) relocation
714 of proteins from established mitochondria into the cytosol in the individual eukaryote crown
715 lineages has not been systematically studied. A careful literature review has revealed only a single
716 case of a protein for which the relocation from mitochondria into the cytosol has been sufficiently
717 documented, mHCF101. This protein was recently shown to be ubiquitous in mitochondria of
718 Cryptista, Haptista, and the SAR clade. However, in the SAR member *Toxoplasma gondii*
719 (Apicomplexa), it has been relocated to the cytosol (Barylyuk et al. 2020; Pyrih et al. 2021).

720 Amitochondriate oxymonads represent a suitable model group for the detection of nucleus-encoded
721 proteins that were secondarily displaced out of the mitochondrion in the course of evolution. We
722 argue that the analyses in the previous section suggest that not only do oxymonads contain no
723 mitochondrion, but they also lost all the protein associated with this organelle in their metamonad
724 ancestors, i.e., none was relocated. Firstly, oxymonads contain no clear orthologues of proteins
725 known to localize to the MROs of *P. pyriformis* and *T. marina* (Leger et al. 2017; Zítek et al. 2022).
726 The closest case is the second copy of glycine cleavage system L-protein (GCS-L2), which likely
727 represents a cytosolic paralogue or xenologue with an unknown function in all Preaxostyla and it
728 is, moreover, suspected to have undergone LGT in its history. Noteworthy is also the situation with
729 PFO and [FeFe] hydrogenases. They are present in multiple paralogues in both oxymonads and
730 their free-living relatives, but none of them was localized into the MRO of *P. pyriformis* (Zítek et
731 al. 2022), so cytosolic localization of these proteins is assumed in the common ancestor of

732 *P. pyriformis* and oxymonads, while their localization in the metamonad common ancestor cannot
733 be inferred from their distribution and phylogenies. Finally, the comparison with the
734 experimentally verified *T. brucei* proteome indicated several partial or full relocations of originally
735 cytosolic proteins into the mitochondrion of trypanosomatids but not a single mitochondrion-to-
736 cytosol protein relocation in the lineage leading to oxymonads.

737 The fact that we do not detect any clear protein retargeted outside mitochondria in oxymonads is
738 surprising. Why not even a single of the mitochondrial proteins present in the common ancestor of
739 oxymonads and trypanosomatids has been retained and repurposed in the oxymonad cell? Is there
740 a constrain that prevents mitochondrial proteins' relocation during the reductive evolution of these
741 organelles, or are the functions of mitochondrial proteins outside the organelle in anaerobes and
742 microaerophiles useless?

743 Similar observations were made in other protists with highly reduced mitochondria, such as
744 *G. intestinalis* or *E. histolytica*, but the topic has never been addressed systematically, and the
745 evidence was not as clear-cut, since the proteomes of their MROs are still not robustly established.
746 Mitosomes of *G. intestinalis* represent an interesting and extreme example. They harbour a handful
747 of proteins and fulfil likely only one essential biological role, Fe-S cluster synthesis. To run this
748 pathway in the organelle, the cell maintains additional pathways for mitosomal biogenesis and the
749 import of proteins and metabolites. Although the relocation of the "last-standing" pathway into the
750 cytosol seems a move to a much more efficient arrangement, the evolution has never achieved it
751 by retargeting the proteins. In the few organisms that have managed to dispense with the
752 mitochondrial part of the Fe-S cluster synthesis pathway, the proteins were replaced by their
753 functional xenologues acquired by LGT: SUFs in Preaxostyla and NIFs in Archamoebae (Ali et al.
754 2004; Gill et al. 2007; Karnkowska et al. 2016; Zítek et al. 2022). Although the protein relocations

755 are generally considered relatively simple and probable events, relocations outside mitochondria
756 may not be, as our data suggest.

757 **Conclusions**

758 In this manuscript we report the genome sequences of *P. pyriformis* and *B. nauphoetae*, adding to
759 the previously published genomes of *M. exilis* and *S. strix*, thus doubling the number of available
760 annotated genomes of Preaxostyla protists. By comparing the proteomes predicted from these data
761 to available genomes and transcriptomes of metamonads we were able to define the core gene
762 inventories of Metamonada, Preaxostyla, and Oxymonadida, and to detect LGT contributions to
763 them. Based on its genetic toolkit, *P. pyriformis* is endowed with several functions unique among,
764 or rare in, eukaryotes – synthesis of *p*-cresol, mercury methylation, and Se volatilization, or
765 preforms some of the common functions in an unusual way, exemplified by the predicted fumarate-
766 dependent NAD^+ synthesis. We identified differences in inventories of membrane transporters
767 between oxymonads and free-living *P. pyriformis* and *T. marina* potentially reflecting the transition
768 from the free-living to the endobiotic lifestyle at the origin of oxymonads. Proteins involved in the
769 formation and regulation of the Golgi structure have a patchy distribution and show a trend towards
770 loss in oxymonads, which is consistent with the lack of ultrastructural evidence for the presence of
771 a stacked Golgi in oxymonads, but also with an emerging cell biological model for how Golgi
772 maintain their hallmark morphology of stacked cisternae. The switch to the SUF pathway in these
773 species has apparently not affected the number of Fe-S-containing proteins but led to a decrease in
774 the usage of 2Fe-2S clusters.

775 The findings of multiple exclusively mitochondrion-associated proteins in the two studied free-
776 living species, *T. marina* and *P. pyriformis*, corroborate the presence of a unique types of MROs
777 in these organisms. Contrary, the thorough search for mitochondrion-associated proteins has failed

778 to uncover any convincing candidates in the studied oxymonads, further supporting the hypothesis
779 that not only *M. exilis* but at least a large part of oxymonads has completely lost the mitochondrion.
780 This fact moves this unique loss to at least 100 MYA deep past, when oxymonads had been already
781 diversified (Poinar 2009), and shows that a eukaryotic lineage without mitochondria can thrive for
782 eons and undergo pronounced morphological evolution. Thorough searches revealed that
783 oxymonads not only lost the organelle but did not retain any protein acting within the
784 mitochondrion of their ancestors, indicating that the transformation from aerobic mitochondria to
785 microaerophilic MROs and further to the amitochondriate cell was not accompanied by the
786 relocations of proteins to other cell compartments, but by gradual loss of all proteins concerned.

787 **MATERIALS AND METHODS**

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

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

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

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

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

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

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

816 For *P. pyriformis* PacBio sequencing was performed at SEQme s.r.o, using a PacBio RSII
817 sequencer, whereas Illumina sequencing was performed using Illumina HiSeq and MiSeq
818 sequencers at the Institute of Molecular Genetics of the ASCR, v. v. i. For *B. nauphoetae*, one pair-
819 end and two mate-pair libraries were prepared and sequenced on Illumina MIseq PE 2x300bp at
820 Beijing Genomics Institute (BGI).

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

828 **Genome assembly binning and decontamination**

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

833 The initial assembly of the genomes was made only with the Nanopore and PacBio generated reads
834 using Canu v1.7.1 assembler (Koren et al. 2017), with corMinCoverage and corOutCoverage set
835 to 0 and 100000 respectively. After assembly, the data was binned using tetraESOM (Haddad et
836 al. 2009). The resulting eukaryotic bins were also checked using a combination of BLASTn and a
837 scoring strategy based on the identity and coverage of the scaffold as described in (Treitli et al.
838 2019). After binning, the resulted genomic bins were polished in two phases. In the first phase, the
839 scaffolds were polished with Nanopolish (Loman et al. 2015) using the raw reads generated by
840 Nanopore. In the second phase, the resulting scaffolds generated by Nanopolish were further
841 corrected using Illumina short reads with Pilon v1.21 (Walker et al. 2014). Finally, the genome
842 assembly of *P. pyriformis* was further scaffolded with raw RNA-seq reads using Rascaf (Song et
843 al. 2016).

844 **Repeat masking, prediction, and annotation of the genomes**

845 Repetitive elements in the genomic assembly were identified using RepeatModeler v1.0.11
846 (Tarailo-Graovac and Chen 2009), and only repeats that are members of known repeat families
847 were used for masking the genome assemblies prior to gene prediction. For the *P. pyriformis* and
848 *B. nauphoetae*, we used Augustus for gene prediction (Stanke and Waack 2003). For *de novo*
849 prediction of genes, Augustus was first re-trained using a manually curated set of gene models. In
850 the next step, intron hints were generated from the RNAseq data and gene prediction was performed
851 on repeat-masked genomes using Augustus 3.2.3 (Stanke and Waack 2003). Next, transcriptome
852 assemblies were mapped to the genomes using PASA (Haas et al. 2003), and the resulting
853 assembled transcripts were used as evidence for gene model polishing with EVM (Haas et al. 2008).
854 The genome completeness for each genome was estimated using BUSCO v3 with the Eukaryota
855 odb9 dataset and the genome completeness was estimated on the sets of Augustus-predicted protein
856 sequences as the input.

857 Automatic functional annotation of the predicted genes was performed using the KEGG Automatic
858 Annotation Server (Moriya et al. 2007), in parallel to similarity searches against NCBI nr protein
859 database using BLASTp. Targeted analyses of genes and gene families of specific interest were
860 performed by manual searches of the predicted proteomes using BLASTp and HMMER (Eddy
861 2011), and complemented by tBLASTn searches of the genome and transcriptome assemblies to
862 check for the presence of genes that were potentially missed in the predicted protein sets. Gene
863 models were manually refined for proteins of interest when necessary and possible. The annotated
864 genomes were submitted to NCBI under BioProject PRJNA903905, BioSample SAMN31819772
865 (*P. pyriformis*) and BioProject PRJNA887011, BioSample SAMN31149887 (*B. nauphoetae*).

866 **Defining of groups-specific gene sets**

867 We combined the gene inventories of five Preaxostyla species (two genomes presented here +
868 *M. exilis* genome under NCBI BioProject PRJNA304271 + *S. strix* genome under NCBI BioProject
869 PRJNA524138 + *T. marina* transcriptome-derived predicted proteome in EukProt EP00771,
870 Richter et al. 2022) with 14 other metamonads for which the genome or transcriptome drafts were
871 available when the study was initiated (Supplementary file 1). The complete set of proteins from
872 the 19 species comprised 337,300 items, of which 256,334 (76%) were assigned to 37,923
873 orthologous groups (OGs) by OrthoMCL (Supplementary file 1). We run the OrthoMCL pipeline
874 (Li et al. 2003) to determine orthologous groups (OGs) of proteins of selected metamonads. To
875 optimize the inflation parameter value, we ran multiple clusterings with different inflation
876 parameter values in the range 1.01 to 30.0 and calculated the number of OGs containing genes from
877 all Preaxostyla but no other taxa and from all oxymonads but no other taxa. We chose to proceed
878 with the inflation parameter 4, because under this setting the sum of these two numbers was the
879 highest, so the strength of the clustering should be optimal. Proteins belonging to the thus defined
880 OGs were automatically annotated using BLASTp searches against the NCBI nr protein database
881 (Supplementary file 1). Venn diagrams were generated using InteractiVenn (Heberle et al. 2015).
882 Relaxed core gene set for Metamonada and its Preaxostyla and Oxymonadida specific increments
883 were estimated based on the presence/absence of each OG in the analyzed species (Supplementary
884 file 2).

885 1,399 OGs present in at least one member of each of three main metamonad clades represented in
886 the analysis (Parabasalia, Fornicata, and Preaxostyla) were considered as the relaxed core
887 Metamonada gene set. The Preaxostyla specific increment was determined as a union of two sets
888 determined by two different strategies to minimize the effect of data incompleteness: i) the set of

889 285 OGs present in representatives of all three lineages of Preaxostyla (at least one oxymonad, *P.*
890 *pyriformis*, and *T. marina*), and ii) the set of 460 OGs present in any four out of the five Preaxostyla
891 species in the analysis. Out of these, 169 OGs were present in all five Preaxostyla species. The
892 Oxymonadida specific increment was determined as a set of 1,242 OGs present in at least two out
893 of the three oxymonad species yet lacking representative genes in other metamonads analyzed.

894 **Detection of lateral gene transfer candidates**

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

900 We first discarded predicted oxymonad proteins of size smaller than 80 amino acids, since the
901 corresponding phylogenies are unlikely to contain a strong and reliable signal. Given a large
902 number of remaining proteins (109,359), and in order to reduce computational burden, we carried
903 out a preliminary screen to discard sequences without detectable prokaryotic homologues. For this,
904 each metamonad sequence was used as a query for a similarity search using DIAMOND (Buchfink
905 et al. 2014) against the GTDB database (release 95; Parks et al. 2022) spanning 194,600 proteomes
906 representing known prokaryotic diversity as of July 2020 (e-value cut-off = 1e-10; identity
907 percentage >25%; query coverage >50%; subject coverage >50%). The query and subject coverage
908 criteria were chosen to ensure we did not consider spurious hit over a short part of the protein.
909 Metamonad sequences with significant hits against the GTDB were then used for a more exhaustive
910 homologue search by using them as queries for DIAMOND searches against the NCBI nr database
911 (the same cut-offs as before) in order to retrieve both prokaryotic and eukaryotic homologues. In

912 order to reduce the computational burden, we then reduced the number of sequences retrieved from
913 NCBI nr by running CD-HIT (-c 0.7; Li and Godzik 2006) to keep a single representative sequence
914 for clusters of sequences displaying >70% identity. We then retrieved the NCBI taxonomy
915 associated with each sequence.

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

920 Files with 30 sequences or fewer were discarded in order to guarantee that the directionality of the
921 LGTs can be inferred from the phylogenies. In addition, files with more eukaryotic homologues
922 than prokaryotic ones were also discarded from follow-up analyses, as they were unlikely to
923 represent a clear case of prokaryote-to-eukaryote gene transfer. The remaining files were
924 individually aligned using MAFFT (Katoh and Standley 2013) and trimmed using BMGE
925 (Criscuolo and Gribaldo 2010), and the corresponding phylogenies were inferred using FastTree
926 v.2 (Price et al. 2010).

927 To investigate the potential cases of prokaryotes-to-metamonads LGT, we used the R package
928 PhySortR (Stephens et al. 2016) in order to identify well-supported clades comprised of
929 metamonad sequences nested within prokaryotic ones. This non-trivial task was done in a stepwise
930 manner by first identifying clans with a statistical support (Shimodaira-Hasegawa test) >0.8 and
931 containing at least 90% of prokaryotes and metamonads. The reason for the latter threshold is to
932 allow for cases of taxonomic misannotation, or for the presence of a few homologues from other
933 eukaryotes, themselves potentially resulting from LGT. After this first assessment, we checked
934 which metamonad sequences were forming a well-supported monophyletic clan (i.e., a clan with

935 >90% of metamonads with >0.8 support) in order to infer if they likely corresponded to a single
936 LGT even within their last common ancestor, or if they were dispersed in the tree suggesting
937 multiple independent transfers of homologous genes.

938 However, when metamonad sequences were detected as paraphyletic, and in order to avoid
939 overestimating the number of LGT events, we manually inspected all phylogenies in order to
940 distinguish most accurately between probable independent cases of LGTs and paraphyly resulting
941 from an overall lack of phylogenetic resolution. For this, our manual inspection took into account:
942 1) the overall number of sequences in the tree (i.e., a small number can be indicative that only a
943 few homologues passed our DIAMOND threshold because this protein family is fast evolving and
944 that, consequently, the corresponding phylogeny might be poorly resolved and/or not contain
945 representatives of the true diversity of homologues); 2) the branch length separating the metamonad
946 sequences (i.e., short branch lengths can be suggestive of an unreliably resolved backbone); 3) the
947 overall taxonomic coherence of the tree (i.e., do the various clades display taxonomic homogeneity
948 or are the lineages interspersed all over the tree, suggestive of a poorly resolved phylogeny?); 4)
949 the taxonomic affiliation of the prokaryotic neighbors of each independent metamonad clade in the
950 tree (i.e., if they are similar, this is more in favor of a single LGT event ancestral to all metamonad
951 represented in the tree, which phylogenetic reconstructions have not resolved properly); 5) the
952 overall branch support across the tree (i.e., some trees might show a few well supported branches
953 suggestive of paraphyletic metamonads but if the overall backbone of tree is generally poorly
954 supported, we were cautious about the weakness of the phylogenetic signal). Taking all this into
955 account, in case of doubt, we always minimized the number of inferred LGT events by assuming a
956 single event in the last common ancestor of the metamonad sequences in a tree. All phylogenies
957 and FASTA files can be found in Supplementary file 11.

958

959 **Single gene phylogenies**

960 The phylogenies of the genes of interest were analyzed individually using the methodology
961 described below. Eukaryotic homologues of the Preaxostyla genes were gathered by taxonomically
962 constrained iterative BLAST search against publicly available sequence databases in order to
963 sample as broad eukaryotic diversity as possible. In the cases of overrepresented taxa of low interest
964 (e.g. Metazoa, land plants), only a small number of representative sequences were selected
965 arbitrarily. Prokaryotic homologues were gathered by a BLASTp search with each eukaryotic
966 sequence against the NCBI nr database with an e-value cutoff of 1e-10 and max. 10 target
967 sequences. The sequences were aligned using MAFFT v7.221 (Katoh and Standley 2013) and
968 automatically trimmed using trimAl v1.2 (Capella-Gutiérrez et al. 2009). Phylogenetic analyses
969 were performed simultaneously using IQ-Tree v2.0.5 (Nguyen et al. 2015) and RAxML-HPC2 on
970 v8.2.12 (Stamatakis 2014) on the Cyberinfrastructure for the Phylogenetic Research (CIPRES)
971 Science Gateway (Miller et al. 2010). Substitution models were inferred using IQ-Tree TESTNEW
972 function for IQ-Tree and using ModelTest-NG v0.1.5 (Darriba et al. 2019) on the CIPRES Science
973 Gateway for RAxML.

974 **Analysis of the Rsg1 gene**

975 Candidate Rsg1 orthologues were identified in a variety of databases and other sequence resources
976 (Supplementary file 5) using a combination of BLAST (BLASTp or tBLASTn, depending on the
977 target database) and HMMER searches. The taxon selection essentially followed the list of taxa
978 employed in a recent study of the distribution of the cilium-associated Ras superfamily GTPase
979 Arl16 (Dewees et al. 2022), with further expansion to include additional metamonads and recently
980 sequenced species of isolated phylogenetic position. Additional taxa were added based on

981 searching the newly released EukProt database (Richter et al. 2022) such that representatives of
982 broader taxa consistently possessing or lacking a discernible Rsg1 orthologue were selected and
983 including the final list. Significant BLAST hits retrieved from searches with reference queries (the
984 human sequence – GenBank accession number NP_112169.1, or another already validated
985 sequence from a given taxon to increase the sensitivity of the search in a taxon-specific manner)
986 were evaluated by BLAST searches against a large custom manually annotated database of Ras
987 superfamily genes to distinguish Rsg1 orthologues from other GTPase types. The discrimination
988 was facilitated by Rsg1 being rather divergent from other Ras superfamily members (such as Rab
989 GTPases). A set of confidently identified Rsg1 sequences was then used for building a custom
990 profile HMM, which was employed in searches with HMMER to identify possible divergent Rsg1
991 orthologues missed by BLAST searches. The candidates were again evaluated by BLAST-based
992 comparisons against the custom GTPase database.

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

994 Comparative genomics was carried out using HMM searches. Pan-eukaryotic query sequences
995 analyzed and curated for previous pan-eukaryotic vesicle coat, multisubunit tethering complexes,
996 and golgins were used to build profile HMMs. Query sequences were obtained from supplementary
997 material of previous papers dealing specifically with Adaptins and COPI (Hirst et al. 2014),
998 multisubunit tethering complexes (Klinger et al. 2013), and golgins (Barlow et al. 2018). Individual
999 components and proteins from each sub-complex were aligned using MUSCLE v.3.8.31 and the
1000 resulting alignment files were used to generate HMM matrices using the HMMBUILD option
1001 available through the HMMER package (Eddy 2011). HMMER searches were carried out in the
1002 predicted proteomes of *P. pyriformis*, *B. nauphoetae*, and *S. strix*, whereas for *T. marina* the
1003 transcriptome assembly was first translated in all six open reading frames using the *ab initio* gene

1004 prediction program GeneMarkS-T using the default parameters (Tang et al. 2015) and the longest
1005 resulting predicted protein sequences were used for the searches. Forward hits meeting an e-value
1006 cutoff of 0.05 were subject to reciprocal BLASTP analyses against the *Homo sapiens* and *M. exilis*
1007 predicted proteomes as well as the NCBI nr database. Any absent components were also subject to
1008 additional tBLASTn searches in the nucleotide scaffolds. Hits were deemed positive if both
1009 forward hits and at least two of three reciprocal BLAST results retrieved the correct ortholog with
1010 an e-value cutoff of 0.05. Paralogous gene families were subject to further phylogenetic analyses.

1011 Phylogenetic analyses were undertaken for the large, medium and small subunits from identified
1012 AP1-4 and CopI families, as well as Syntaxin16 and Syntaxin5. Resolved backbone alignments
1013 from dataset curated for Karnkowska et al. (2019) were used and sequences from *T. marina*,
1014 *P. pyriformis*, *B. nauphoetae*, and *S. strix* were iteratively aligned with the backbone alignment
1015 using the profile option available through MUSCLE v3.8.31 (Edgar 2004). All alignments were
1016 subsequently inspected and manually trimmed using Mesquite v. 3.5
1017 (https://www.mesquiteproject.org/) to remove heterogeneous regions lacking obvious homology
1018 and partial or misaligned sequences.

1019 Maximum likelihood analyses were carried out using RAxML-HPC2 on XSEDE v.8.2.10 for non-
1020 parametric bootstrap replicates (Stamatakis 2006). The best protein matrix model for RAxML was
1021 tested using ProtTest v.3.4.2 (Abascal et al. 2005), set to account for Gamma rate variation (+G),
1022 invariant sites (+I), and observed frequency of amino acids (+F) with default tree and a consensus
1023 tree was obtained using the Consense package, available through the Phylip v.3.66.

1024 Bayesian inference was carried out using MRBAYES on XSEDE v3.2.6 (Huelsenbeck and
1025 Ronquist 2001). Parameters specified included 10 million Markov Chain Monte Carlo generations
1026 under a mixed amino acid model with the number of gamma rate categories set a 4. Sampling

1027 frequency was set to occur every 1000 generations with a burn-in of 0.25. Tree convergence was
1028 ensured with the average standard deviation of split frequency values below 0.01. Both RAxML
1029 and MRBAYES analyses were performed on the CIPRES webserver (Miller et al. 2015). RAxML
1030 bootstrap values as well as Bayesian posterior probabilities were overlaid on the best Bayes
1031 topology with combined values of >50 and >0.80, respectively, indicating branch support. Tree
1032 visualization and rooting were carried out in FigTree v.1.4.4. Branch support value overlay and
1033 additional annotations were prepared using Adobe Illustrator CS4. All alignments are available
1034 upon request.

1035 **Prediction of proteins containing Fe-S clusters**

1036 Fe-S cluster containing proteins were predicted with MetalPredator webserver (Valasatava et al.
1037 2016). Predicted proteins were functionally annotated by BLAST searches against NCBI nr
1038 database and by InterProScan against the InterPro database (Blum et al. 2021). KEGG categories
1039 were assigned by GhostKOALA searches against the KEGG database (Kanehisa et al. 2016).
1040 Orthologous groups were created with the OrthoFinder 2 software (Emms and Kelly 2019) and the
1041 Venn diagram of OG sharing among Preaxostyla was visualized using InteractiVenn (Heberle et
1042 al. 2015).

1043 **Search for mitochondrial proteins**

1044 The comprehensive search for putative mitochondrial protein was performed for all five
1045 Preaxostyla species, including previously analyzed *M. exilis* (Karnkowska et al. 2016) as a control.
1046 Predicted proteins of each species were independently searched for similarity to known
1047 mitochondrial/MRO-related proteins and the presence of mitochondrial signature sequences. The
1048 general design of the search followed the previously described methodology (Karnkowska et al.

1049 2016). Briefly, a custom mitochondrial protein sequence database was established using the
1050 MitoMiner v4.0 database (Smith et al. 2012; Smith and Robinson 2018). Experimentally confirmed
1051 mitochondrial proteins (at least one GFP tagging experiment or three different MS experiments)
1052 coming from *H. sapiens*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, *Saccharomyces*
1053 *cerevisiae*, and *Schizosaccharomyces pombe* were used and supplemented with the MROs proteins
1054 from sixteen different organisms (Stechmann et al. 2008; Mi-ichi et al. 2009; Barberà et al. 2010;
1055 Alcock et al. 2012; Stairs et al. 2014; Noguchi et al. 2015; Nývltová et al. 2015; Leger et al. 2016;
1056 Leger et al. 2017; Pyrihová et al. 2018). Redundant homologues (90% similarity threshold) were
1057 removed from the database using CD-HIT (Li and Godzik 2006). The resulting non-redundant
1058 database contained 6,979 proteins. Reciprocal BLAST analysis was performed for each predicted
1059 proteome with an e-value threshold of 0.001. HMM searches were used to identify proteins
1060 involved in protein import and translocation, as these were shown to be often divergent (Leger et
1061 al. 2016). Searches were done in HMMER 3.1b2 (Eddy 2011) using profile HMMs previously
1062 employed in Karnkowska et al. (2016).

1063 Mitochondrial targeting signals were detected using TargetP v1.1 (Emanuelsson et al. 2007) and
1064 MitoFates v1.1 (Fukasawa et al. 2015). Both programs indicate a putative probability of
1065 mitochondrial localization of the protein. Hits with the probability of mitochondrial localization
1066 indicated to be >0.5 by both programs were considered for manual verification. To find tail-
1067 anchored proteins, transmembrane domains (TMDs) were predicted using TMHMM2.0 (Krogh et
1068 al. 2001) for all analyzed proteins. Hits with a TMD within 32 amino acid residues from the C-
1069 terminus were kept for verification. The mitochondrial β-barrel outer membrane proteins
1070 (MBOMPs) search has been conducted using the pipeline described by Imai et al. (2011). The
1071 pipeline firstly identifies a β-signal ($P_oxGh_yxH_yxH_y$ motif) in the C-terminus of protein, required

1072 for the insertion into the membrane. Subsequently, the secondary structure of the stretch of 300
1073 amino acid residues preceding the β-signal is analyzed using PSIPRED (McGuffin et al. 2000) to
1074 check for a typical β-structure. Significant hits, with at least 25% of the sequence predicted to form
1075 β-strands, no more than 10% assumed by an α-helical structure, and no more than 50% of the eight
1076 residues of the β-signal predicted as an α-helical structure, were further analyzed.

1077 In all methods, the hits were checked with additional attention paid to the proteins with hits in the
1078 MitoMiner database, herein considered "candidates". Since divergent MBOMPs have been shown
1079 to be missed by homology-based searches, for those we used NCBI nr instead, following the idea
1080 that any match with NCBI-nr should allow us to distinguish between unusual and improperly
1081 predicted proteins. All "candidates" have been thoroughly manually inspected. All proteins have
1082 been BLAST-searched against NCBI-nr and the best hits with the description not including the
1083 terms 'low quality protein', 'hypothetical', 'unknown', etc. in the description were kept. For each hit,
1084 the Gene Ontology categories were assigned using InterProScan-5.36-75.0 (Jones et al. 2014). The
1085 annotation of the *M. exilis* genome (Karnkowska et al. 2019) was transferred to orthologous
1086 proteins from the other Preaxostyla species. Based on gathered information, verified "candidates"
1087 are considered mitochondrial proteins.

1088 ***Trypanosoma brucei* mitoproteome-guided comparative analyses**

1089 Predicted proteomes of *T. marina*, *P. pyriformis*, *S. strix*, *B. nauphoetae,* and *M. exilis* were
1090 individually reverse BLAST-searched against the proteome of *T. brucei* (downloaded from
1091 Tritrypdb.org; November 2019; Aslett et al. 2010). Only reciprocal best blast hits that were
1092 identified in either of the previously published mitochondrial proteomes of *T. brucei* (Panigrahi et
1093 al. 2009; Peikert et al. 2017) were subjected to further phylogenetic analysis. For each such protein,
1094 the data set for the tree construction was composed of hits from a custom BLAST database of

1095 selected protist proteomes (Supplementary file 11, downloaded from UniProt, November 2019;
1096 Bateman 2019). Protein sequence sets were automatically aligned with MAFFT v7.453 (Katoh and
1097 Standley 2013) using the L-INS-i refinement and a maximum of 1,000 iterations, followed by
1098 trimming of sites with >70% gaps.

1099 ML trees (Supplementary file 10) were inferred by IQ-TREE v 1.6.12 using the Posterior Mean
1100 Site Frequency (PMSF) empirical model with the ultrafast bootstrapping strategy (1,000 replicates)
1101 and a LG4X guide tree. Subcellular targeting of all proteins in the tree was predicted by using
1102 TargetP-1.1 (Emanuelsson et al. 2007; https://services.healthtech.dtu.dk/service.php?TargetP-1.1);
1103 the presence of a signal peptide, a chloroplast targeting peptide or a mitochondrial targeting peptide
1104 in the respective proteins is marked by the letters S, C, or M, respectively, at the very beginning of
1105 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.

1121 **REFERENCES**

1122 Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution.
1123 *Bioinformatics* 21:2104–2105.

1124 Agbu SO, Liang Y, Liu A, Anderson K v. 2018. The small GTPase RSG1 controls a final step in
1125 primary cilia initiation. *J Cell Biol* 217:413–427.

1126 Ahat E, Li J, Wang Y. 2019. New Insights into the Golgi Stacking Proteins. *Front Cell Dev Biol*
1127 7:131.

1128 Alcock F, Webb CT, Dolezal P, Hewitt V, Shingu-Vasquez M, Likić VA, Traven A, Lithgow T.
1129 2012. A small Tim homohexamer in the relict mitochondrion of *Cryptosporidium*. *Mol Biol*
1130 *Evol* 29:113–122.

1131 Ali V, Shigeta Y, Tokumoto U, Takahashi Y, Nozaki T. 2004. An intestinal parasitic protist,
1132 *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron-
1133 sulfur cluster assembly under anaerobic conditions. *J Biol Chem* 279:16863–16874.

1134 Anderson IJ, Loftus BJ. 2005. *Entamoeba histolytica*: Observations on metabolism based on the
1135 genome sequence. *Exp Parasitol* 110:173–177.

1136 Andreini C, Banci L, Rosato A. 2016. Exploiting Bacterial Operons to Illuminate Human Iron-
1137 Sulfur Proteins. *J Proteome Res* 15:1308–1322.

1138 Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data.

1139 Aridor M. 2022. A tango for coats and membranes: New insights into ER-to-Golgi traffic. *Cell Rep*
1140 38:110258.

1141 Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP,
1142 Fischer S, Gajria B, Gao X, et al. 2010. TriTrypDB: a functional genomic resource for the
1143 Trypanosomatidae. *Nucleic Acids Res* 38:D457–D462.

1144 Babady NE, Pang YP, Elpeleg O, Isaya G. 2007. Cryptic proteolytic activity of dihydrolipoamide
1145 dehydrogenase. *Proc Natl Acad Sci U S A* 104:6158–6163.
1146 Balaji S, Aravind L. 2007. The RAGNYA fold: a novel fold with multiple topological variants
1147 found in functionally diverse nucleic acid, nucleotide and peptide-binding proteins. *Nucleic*
1148 *Acids Res* 35:5658–5671.
1149 Barberà MJ, Ruiz-Trillo I, Tufts JY a, Bery A, Silberman JD, Roger AJ. 2010. *Sawyeria*
1150 *marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties.
1151 *Eukaryot Cell* 9:1913–1924.
1152 Barlow LD, Nývltová E, Aguilar M, Tachezy J, Dacks JB. 2018. A sophisticated, differentiated
1153 Golgi in the ancestor of eukaryotes. *BMC Biol* 16:27.
1154 Barylyuk K, Koreny L, Ke H, Butterworth S, Crook OM, Lassadi I, Gupta V, Tromer E, Mourier
1155 T, Stevens TJ, et al. 2020. A Comprehensive Subcellular Atlas of the Toxoplasma Proteome
1156 via hyperLOPIT Provides Spatial Context for Protein Functions. *Cell Host Microbe* 28:752-
1157 766.e9.
1158 Bateman A. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–
1159 D515.
1160 van der Bliek AM, Shen Q, Kawajiri S. 2013. Mechanisms of mitochondrial fission and fusion.
1161 *Cold Spring Harb Perspect Biol* 5:a011072.
1162 Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-
1163 Lafosse T, Qureshi M, Raj S, et al. 2021. The InterPro protein families and domains database:
1164 20 years on. *Nucleic Acids Res* 49(D1): D344–D354.
1165 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
1166 data. *Bioinformatics* 30:2114–2120.

1167 Bolte K, Rensing SA, Maier UG. 2015. The evolution of eukaryotic cells from the perspective of
1168 peroxisomes. *BioEssays* 37:195–203.
1169 Boncompain G, Weigel A v. 2018. Transport and sorting in the Golgi complex: multiple
1170 mechanisms sort diverse cargo. *Curr Opin Cell Biol* 50:94–101.
1171 Borgese N, Brambillasca S, Colombo S. 2007. How tails guide tail-anchored proteins to their
1172 destinations. *Curr Opin Cell Biol* 19:368–375.
1173 Brännström IO, Stairs C, Campos KIA, Ettema TJG, Keeling PJ, Bass D, Burki F. 2022. A
1174 Mitosome With Distinct Metabolism in the Uncultured Protist Parasite *Paramikrocytos*
1175 *canceri* (Rhizaria, Ascetosporea). *SSRN Electronic Journal* SSRN:4065074.
1176 Braymer JJ, Freibert SA, Rakwalska-Bange M, Lill R. 2021. Mechanistic concepts of iron-sulfur
1177 protein biogenesis in Biology. *Biochim Biophys Acta Mol Cell Res* 1868(1):118863.
1178 Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND.
1179 *Nature Methods* 12:59–60.
1180 Butenko A, Opperdoes FR, Flegontova O, Horák A, Hampl V, Keeling P, Gawryluk RMR,
1181 Tikhonenkov D, Flegontov P, Lukeš J. 2020. Evolution of metabolic capabilities and
1182 molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biol* 18:1–28.
1183 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment
1184 trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
1185 Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark
1186 UCM, Besteiro S, et al. 2007. Draft genome sequence of the sexually transmitted pathogen
1187 *Trichomonas vaginalis. Science* 315:207–212.
1188 Christensen GA, Wymore AM, King AJ, Podar M, Hurt RA, Santillan EU, Soren A, Brandt CC,
1189 Brown SD, Palumbo A v., et al. 2016. Development and validation of broad-range qualitative

1190 and clade-specific quantitative molecular probes for assessing mercury methylation in the
1191 environment. *Appl Environ Microbiol* 82:6068–6078.
1192 Cooper CJ, Zheng K, Rush KW, Johs A, Sanders BC, Pavlopoulos GA, Kyrpides NC, Podar M,
1193 Ovchinnikov S, Ragsdale SW, et al. 2020. Structure determination of the HgcAB complex
1194 using metagenome sequence data: insights into microbial mercury methylation.
1195 *Communications Biology* 3:1–9.
1196 Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new
1197 software for selection of phylogenetic informative regions from multiple sequence
1198 alignments. *BMC Evol Biol* 10:210.
1199 Darriba D, Posada David, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2019. ModelTest-NG: a
1200 new and scalable tool for the selection of DNA and protein evolutionary models. *bioRxiv*
1201 612903.
1202 Dean S, Sunter JD, Wheeler RJ. 2017. TrypTag.org: A Trypanosome Genome-wide Protein
1203 Localisation Resource. *Trends Parasitol* 33:80–82.
1204 Denic V. 2012. A portrait of the GET pathway as a surprisingly complicated young man. *Trends*
1205 *Biochem Sci* 37:411–417.
1206 Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015.
1207 Bacterial proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci U S A* 112:E693–
1208 E699.
1209 Dewees SI, Vargová R, Hardin KR, Turn RE, Devi S, Linnert J, Wolfrum U, Caspary T, Eliáš M,
1210 Kahn RA. 2022. Phylogenetic profiling and cellular analyses of ARL16 reveal roles in traffic
1211 of IFT140 and INPP5E. *Mol Biol Cell* 33:ar33.

1212 Diamond LS. 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and other
1213 lumen-dwelling protozoa. *J Parasitol* 68:958–959.
1214 Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the Molecular Machines for Protein
1215 Import into Mitochondria. *Science (1979)* 313:314–319.
1216 Ducker GS, Rabinowitz JD. 2017. Cell Metabolism Review One-Carbon Metabolism in Health
1217 and Disease. *Cell Metab* 25:27–42.
1218 Duschak VG, Cazzulo JJ. 1991. Subcellular localization of glutamate dehydrogenases and alanine
1219 aminotransferase in epimastigotes of *Trypanosoma cruzi*. *FEMS Microbiol Lett* 83:131–135.
1220 Eberhardt EL, Ludlam A v., Tan Z, Cianfrocco MA. 2020. Miro: A molecular switch at the center
1221 of mitochondrial regulation. *Protein Sci* 29:1269–1284.
1222 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7(10): e1002195.
1223 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
1224 *Nucleic Acids Res* 32:1792–1797.
1225 Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using
1226 TargetP, SignalP and related tools. *Nat Protoc* 2:953–971.
1227 Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative
1228 genomics. *Genome Biol* 20:238.
1229 Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. 2015. MitoFates: Improved prediction of
1230 mitochondrial targeting sequences and their cleavage sites. *Molecular and Cellular*
1231 *Proteomics* 14:1113–1126.
1232 Gentekaki E, Curtis BA, Stairs CW, Klimeš V, Eliáš M, Salas-Leiva DE, Herman EK, Eme L,
1233 Arias MC, Henrissat B, et al. 2017. Extreme genome diversity in the hyper-prevalent parasitic
1234 eukaryote *Blastocystis*. *PLoS Biol* 11;15(9):e2003769.

1235 Gill EE, Diaz-Triviño S, Barberà MJ, Silberman JD, Stechmann A, Gaston D, Tamas I, Roger AJ.
1236 2007. Novel mitochondrion-related organelles in the anaerobic amoeba *Mastigamoeba*
1237 *balamuthi*. *Mol Microbiol* 66:1306–1320.
1238 Gilmour CC, Elias DA, Kucken AM, Brown SD, Palumbo A v., Schadt CW, Wall JD. 2011.
1239 Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding
1240 bacterial mercury methylation. *Appl Environ Microbiol* 77:3938–3951.
1241 Glaser T, Hedman B, Hodgson KO, Solomon EI. 2000. Ligand K-edge X-ray absorption
1242 spectroscopy: A direct probe of ligand - Metal covalency. *Acc Chem Res* 33(12):859–868.
1243 Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE.
1244 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H 2 is a widely
1245 utilised energy source for microbial growth and survival. *ISME Journal* 10:761–777.
1246 Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM,
1247 Rusch DB, Town CD, et al. 2003. Improving the Arabidopsis genome annotation using
1248 maximal transcript alignment assemblies. *Nucleic Acids Res* 31:5654–5666.
1249 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Robin CR, Wortman JR.
1250 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the
1251 Program to Assemble Spliced Alignments. *Genome Biol* 9:R7.
1252 Haddad I, Hiller K, Frimmersdorf E, Benkert B, Schomburg D, Jahn D. 2009. An emergent self-
1253 organizing map based analysis pipeline for comparative metabolome studies. *In Silico Biol*
1254 9:163–178.
1255 Hampl V. 2017. Preaxostyla. In: Handbook of the Protists. Cham: Springer International
1256 Publishing. p. 1139–1174.

1257 Hampl V, Čepička I, Eliáš M. 2018. Was the mitochondrion necessary to start eukaryogenesis?
1258 *Trends Microbiol* 27:1–9.
1259 Hampl V, Silberman JD, Stechmann A, Diaz-Triviño S, Johnson PJ, Roger AJ. 2008. Genetic
1260 Evidence for a Mitochondriate Ancestry in the 'Amitochondriate' Flagellate *Trimastix*
1261 *pyriformis*. *PLoS One* 3:9.
1262 Hazelwood LA, Daran JM, van Maris AJA, Pronk JT, Dickinson JR. 2008. The Ehrlich pathway
1263 for fusel alcohol production: A century of research on *Saccharomyces cerevisiae* metabolism.
1264 *Appl Environ Microbiol* 74:2259–2266.
1265 Heberle H, Meirelles VG, da Silva FR, Telles GP, Minghim R. 2015. InteractiVenn: a web-based
1266 tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169.
1267 Hirst J, Schlacht A, Norcott JP, Traynor D, Bloomfield G, Antrobus R, Kay RR, Dacks JB,
1268 Robinson MS. 2014. Characterization of TSET, an ancient and widespread membrane
1269 trafficking complex. *Elife* 3:e02866.
1270 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees.
1271 *Bioinformatics* 17:754–755.
1272 Igamberdiev AU, Bykova N v., Ens W, Hill RD. 2004. Dihydrolipoamide dehydrogenase from
1273 porcine heart catalyzes NADH-dependent scavenging of nitric oxide. *FEBS Lett* 568:146–
1274 150.
1275 Imai K, Fujita N, Gromiha MM, Horton P. 2011. Eukaryote-wide sequence analysis of
1276 mitochondrial β-barrel outer membrane proteins. *BMC Genomics* 12:79.
1277 Jerlström-Hultqvist J, Einarsson E, Xu F, Hjort K, Ek B, Steinhauf D, Hultenby K, Bergquist J,
1278 Andersson JO, Svärd SG. 2013. Hydrogenosomes in the diplomonad Spironucleus
1279 salmonicida. *Nat Commun* 4:2493.

1280 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
1281 Nuka G, et al. 2014. InterProScan 5: Genome-scale protein function classification.
1282 *Bioinformatics* 30:1236–1240.
1283 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information,
1284 knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res* 42:D199-205.
1285 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
1286 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*. 428:726-
1287 731.
1288 Karnkowska A, Treitli SC, Brzoň O, Novák L, Vacek V, Soukal P, Barlow LD, Herman EK,
1289 Pipaliya S V, Pánek T, et al. 2019. The oxymonad genome displays canonical eukaryotic
1290 complexity in the absence of a mitochondrion. *Mol Biol Evol* 36:2292–2312.
1291 Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V,
1292 Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle. *Current*
1293 *Biology* 26:1274–1284.
1294 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1295 improvements in performance and usability. *Mol Biol Evol* 30:772–780.
1296 Klinger CM, Karnkowska A, Herman EK, Hampl V, Dacks JB. 2016. Phylogeny and Evolution.
1297 In: Molecular Parasitology. Vienna: Springer Vienna. p. 383–408.
1298 Klinger CM, Klute MJ, Dacks JB. 2013. Comparative genomic analysis of multi-subunit tethering
1299 complexes demonstrates an ancient pan-eukaryotic complement and sculpting in
1300 Apicomplexa. *PLoS One* 8:e76278.

1301 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and
1302 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*
1303 27:722–736.
1304 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein
1305 topology with a hidden Markov model: Application to complete genomes. *J Mol Biol*
1306 305:567–580.
1307 Kulkarni-Gosavi P, Makhoul C, Gleeson PA. 2019. Form and function of the Golgi apparatus:
1308 scaffolds, cytoskeleton and signalling. *FEBS Lett* 593:2289–2305.
1309 Langousis G, Cavadini S, Boegholm N, Lorentzen E, Kempf G, Matthias P. 2022. Structure of the
1310 ciliogenesis-associated CPLANE complex. *Sci Adv* 8(15):eabn0832.
1311 Lee I, Tiwari N, Dunlop MH, Graham M, Liu X, Rothman JE. 2014. Membrane adhesion dictates
1312 Golgi stacking and cisternal morphology. *Proc Natl Acad Sci U S A* 111:1849–1854.
1313 Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel hydrogenosomes in the microaerophilic
1314 jakobid *Stygiella incarcerata*. *Mol Biol Evol* 33:2318–2336.
1315 Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson
1316 JO, Xu F, Yabuki A, et al. 2017. Organelles that illuminate the origins of *Trichomonas*
1317 hydrogenosomes and *Giardia* mitosomes. *Nat Ecol Evol* 1:0092.
1318 Li Jie, Ahat E, Wang Yanzhuang, Li J, Ahat · E, Wang Y. 2019. Golgi Structure and Function in
1319 Health, Stress, and Diseases. *Results Probl Cell Differ* 67:441–485.
1320 Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic
1321 genomes. *Genome Res* 13:2178–2189.
1322 Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein
1323 or nucleotide sequences. *Bioinformatics* 22:1658–1659.

1324 Liapounova NA, Hampl V, Gordon PMK, Sensen CW, Gedamu L, Dacks JB. 2006. Reconstructing
1325 the mosaic glycolytic pathway of the anaerobic eukaryote Monocercomonoides. *Eukaryot*
1326 *Cell* 5:2138–2146.
1327 Lindmark DG, Müller M. 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic
1328 flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem* 248:7724–
1329 7728.
1330 Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using
1331 only nanopore sequencing data. *Nat Methods* 12:733–735.
1332 Lucattini R, Likić VA, Lithgow T. 2004. Bacterial Proteins Predisposed for Targeting to
1333 Mitochondria. *Mol Biol Evol* 21:652–658.
1334 Major P, Embley TM, Williams TA. 2017. Phylogenetic Diversity of NTT Nucleotide Transport
1335 Proteins in Free-Living and Parasitic Bacteria and Eukaryotes. *Genome Biol Evol* 9:480.
1336 Martin W. 2010. Evolutionary origins of metabolic compartmentalization in eukaryotes.
1337 *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:847–855.
1338 Martin W, Müller M. 1998. The hydrogen hypothesis for the first eukaryote. *Nature* 392:37–41.
1339 Maxwell Burroughs A, Aravind L. 2014. A highly conserved family of domains related to the
1340 DNA-glycosylase fold helps predict multiple novel pathways for RNA modifications. *RNA*
1341 *Biol* 11:360–372.
1342 Maxwell Burroughs A, Glasner ME, Barry KP, Taylor EA, Aravind L. 2019. Oxidative opening of
1343 the aromatic ring: Tracing the natural history of a large superfamily of dioxygenase domains
1344 and their relatives. *Journal of Biological Chemistry* 294:10211–10235.
1345 McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server.
1346 *Bioinformatics* 16:404–405.

1347 Mertens E. 1993. ATP versus pyrophosphate: glycolysis revisited in parasitic protists. *Parasitology*
1348 *Today* 9:122–126.
1349 Mi-ichi F, Yousuf MA, Nakada-Tsukui K, Nozaki T, Abu Yousuf M, Nakada-Tsukui K, Nozaki
1350 T. 2009. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway.
1351 *Proceedings of the National Academy of Sciences* 106:21731–21736.
1352 Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of
1353 large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE).
1354 IEEE. p. 1–8.
1355 Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S,
1356 O'Leary MA. 2015. A RESTful API for Access to Phylogenetic Tools via the CIPRES
1357 Science Gateway. *Evolutionary Bioinformatics* 11:EBO.S21501.
1358 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome
1359 annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185.
1360 Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der Giezen
1361 M, Tielens AGM, Martin WF. 2012. Biochemistry and evolution of anaerobic energy
1362 metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76:444–495.
1363 Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, Heaps SE, Hirt RP,
1364 Embley TM. 2013. Reduction and expansion in microsporidian genome evolution: New
1365 insights from comparative genomics. *Genome Biol Evol* 5:2285–2303.
1366 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
1367 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*
1368 32:268–274.

1369 Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura
1370 K, Takishita K. 2015. Metabolic Capacity of Mitochondrion-related Organelles in the Free-
1371 living Anaerobic Stramenopile Cantina marsupialis. *Protist* 166:534–550.
1372 Noodleman L, Case DA. 1992. Density-functional theory of spin polarization and spin coupling in
1373 iron—sulfur clusters. *Adv Inorg Chem* 38:423-458.
1374 Novák L, Zubáčová Z, Karnkowska A, Kolisko M, Hroudová M, Stairs CW, Simpson AGB,
1375 Keeling PJ, Roger AJ, Čepička I, et al. 2016. Arginine deiminase pathway enzymes:
1376 evolutionary history in metamonads and other eukaryotes. *BMC Evol Biol* 16:197.
1377 Nývltová E, Stairs CW, Hrdý I, Rídl J, Mach J, Paqes J, Roger AJ, Tachezy J. 2015. Lateral gene
1378 transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi*
1379 hydrogenosomes. *Mol Biol Evol* 32:1039–1055.
1380 Nývltová E, Šuťák R, Žárský V, Harant K, Hrdý I, Tachezy J. 2017. Lateral gene transfer of p-
1381 cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba*
1382 *balamuthi. Environ Microbiol* 19:1091–1102.
1383 O'Kelly CJ, Farmer MA, Nerad TA. 1999. Ultrastructure of *Trimastix pyriformis* (Klebs) Bernard
1384 et al.: similarities of *Trimastix* species with retortamonad and jakobid flagellates. *Protist*
1385 150:149–162.
1386 Panigrahi AK, Ogata Y, Zíková A, Anupama A, Dalley RA, Acestor N, Myler PJ, Stuart KD. 2009.
1387 A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. *Proteomics*
1388 9:434–450.
1389 Park K, Ju S, Kim N, Park SY. 2021. The Golgi complex: a hub of the secretory pathway. *BMB*
1390 *Rep* 54:246.

1391 Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022. GTDB: an
1392 ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank
1393 normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794.
1394 Payne MJ, Chapman A, Cammack R. 1993. Evidence for an [Fe]-type hydrogenase in the parasitic
1395 protozoan *Trichomonas vaginalis*. *FEBS Lett* 317:101–104.
1396 Peikert CD, Mani J, Morgenstern M, Käser S, Knapp B, Wenger C, Harsman A, Oeljeklaus S,
1397 Schneider A, Warscheid B. 2017. Charting organellar importomes by quantitative mass
1398 spectrometry. *Nat Commun* 8:15272.
1399 Peña-Diaz J, Montalvetti A, Flores C-L, Constán A, Hurtado-Guerrero R, De Souza W, Gancedo
1400 C, Ruiz-Perez LM, Gonzalez-Pacanowska D. 2004. Mitochondrial localization of the
1401 mevalonate pathway enzyme 3-Hydroxy-3-methyl-glutaryl-CoA reductase in the
1402 Trypanosomatidae. *Mol Biol Cell* 15:1356–1363.
1403 Petrat F, Paluch S, Dogruöz E, Dörfler P, Kirsch M, Korth HG, Sustmann R, de Groot H. 2003.
1404 Reduction of Fe(III) Ions Complexed to Physiological Ligands by Lipoyl Dehydrogenase and
1405 Other Flavoenzymes in Vitro: implications for an enzymatic reduction of Fe(III) ions of the
1406 labile iron pool. *Journal of Biological Chemistry* 278:46403–46413.
1407 Pfanner N, Warscheid B, Wiedemann N. 2019. Mitochondrial proteins: from biogenesis to
1408 functional networks. *Nature Reviews Molecular Cell Biology 2018 20:5* 20:267–284.
1409 Pittis AA, Gabaldón T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic
1410 ancestry. *Nature 2016 531:7592* 531:101–104.
1411 Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo A v., Somenahally
1412 AC, Elias DA. 2015. Global prevalence and distribution of genes and microorganisms
1413 involved in mercury methylation. *Sci Adv* 1(9):e1500675.

1414 Poinar GO. 2009. Description of an early Cretaceous termite (Isoptera: Kalotermitidae) and its
1415 associated intestinal protozoa, with comments on their co-evolution. *Parasit Vectors* 2.
1416 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees
1417 for Large Alignments. *PLoS One* 5:e9490.
1418 Pyrih J, Žárský V, Fellows JD, Grosche C, Wloga D, Striepen B, Maier UG, Tachezy J. 2021. The
1419 iron-sulfur scaffold protein HCF101 unveils the complexity of organellar evolution in SAR,
1420 Haptista and Cryptista. *BMC Ecol Evol* 21:1–19.
1421 Pyrihová E, Motyčková A, Voleman L, Wandyszewska N, Fišer R, Seydlová G, Roger A, Kolísko
1422 M, Doležal P. 2018. A single tim translocase in the mitosomes of *Giardia intestinalis*
1423 illustrates convergence of protein import machines in anaerobic eukaryotes. *Genome Biol Evol*
1424 10:2813–2822.
1425 Rada P, Makki A, Žárský V, Tachezy J. 2019. Targeting of tail-anchored proteins to *Trichomonas*
1426 *vaginalis* hydrogenosomes. *Mol Microbiol* 111:588–603.
1427 Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG,
1428 Burki F, Vargas C de. 2022. EukProt: a database of genome-scale predicted proteins across
1429 the diversity of eukaryotes. *bioRxiv* 2020.06.30.180687.
1430 Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria.
1431 *Current Biology* 27:R1177–R1192.
1432 Saas J, Ziegelbauer K, von Haeseler A, Fast B, Boshart M. 2000. A developmentally regulated
1433 aconitase related to iron-regulatory protein-1 is localized in the cytoplasm and in the
1434 mitochondrion of *Trypanosoma brucei*. *J Biol Chem* 275:2745–2755.

1435 Saga Y, Kawashima M, Sakai S, Yamazaki K, Kaneko M, Takahashi M, Sato N, Toyoda Y, Takase
1436 S, Nakano T, et al. 2020. Plant-specific domains and fragmented sequences imply non-
1437 canonical functions in plant aminoacyl-trna synthetases. *Genes (Basel)* 11:1–13.
1438 Sakuraba H, Satomura T, Kawakami R, Yamamoto S, Kawarabayasi Y, Kikuchi H, Ohshima T.
1439 2002. L-aspartate oxidase is present in the anaerobic hyperthermophilic archaeon *Pyrococcus*
1440 *horikoshii* OT-3: Characteristics and role in the de novo biosynthesis of nicotinamide adenine
1441 dinucleotide proposed by genome sequencing. *Extremophiles* 6:275–281.
1442 Salas-Leiva DE, Tromer EC, Curtis BA, Jerlström-Hultqvist J, Kolisko M, Yi Z, Salas-Leiva JS,
1443 Gallot-Lavallée L, Williams SK, Kops GJPL, et al. 2021. Genomic analysis finds no evidence
1444 of canonical eukaryotic DNA processing complexes in a free-living protist. *Nat Commun*
1445 12:6003.
1446 Schut GJ, Adams MWW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin
1447 and NADH synergistically: A new perspective on anaerobic hydrogen production. *J Bacteriol*
1448 191(13):4451-4457.
1449 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E v., Zdobnov EM. 2015. BUSCO: assessing
1450 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*
1451 31:3210–3212.
1452 Slamovits CH, Keeling PJ. 2006. Pyruvate-phosphate dikinase of oxymonads and Parabasalia and
1453 the evolution of pyrophosphate-dependent glycolysis in anaerobic eukaryotes. *Eukaryot Cell*
1454 5:148–154.
1455 Smith AC, Blackshaw JA, Robinson AJ. 2012. MitoMiner: A data warehouse for mitochondrial
1456 proteomics data. *Nucleic Acids Res* 40:D1160-7.

1457 Smith AC, Robinson AJ. 2018. MitoMiner v4.0: an updated database of mitochondrial localization
1458 evidence, phenotypes and diseases. *Nucleic Acids Res* 47: D1225–D1228.
1459 Song L, Shankar DS, Florea L. 2016. Rascaf: Improving Genome Assembly with RNA Sequencing
1460 Data. *Plant Genome* 9:0.
1461 Spalding MD, Prigge ST. 2010. Lipoic acid metabolism in microbial pathogens. *Microbiol Mol*
1462 *Biol Rev* 74:200–228.
1463 Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen M,
1464 Roger AJ. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related
1465 organelles of the anaerobic protist *Pygsuia*. *Current Biology* 24:1176–1186.
1466 Stairs CW, Leger MM, Roger AJ. 2015. Diversity and origins of anaerobic metabolism in
1467 mitochondria and related organelles. *Philos Trans R Soc Lond B Biol Sci* 370:20140326.
1468 Stairs CW, Táborský P, Salomaki ED, Kolisko M, Pánek T, Eme L, Hradilová M, Vlček Č,
1469 Jerlström-Hultqvist J, Roger AJ, et al. 2021. Anaeramoebae are a divergent lineage of
1470 eukaryotes that shed light on the transition from anaerobic mitochondria to hydrogenosomes.
1471 *Curr Biol* 31:5605-5612.e5.
1472 Stamatakis A. 2006. The RAxML 7.0.4 Manual. *Bioinformatics* 22(21):2688–2690.
1473 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1474 phylogenies. *Bioinformatics* 30:1312–1313.
1475 Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron
1476 submodel. *Bioinformatics* 19:ii215–ii225.
1477 Stechmann A, Baumgartner M, Silberman JD, Roger AJ. 2006. The glycolytic pathway of
1478 *Trimastix pyriformis* is an evolutionary mosaic. *BMC Evol Biol* 6:101.

1479 Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GSS, Van Der Giezen M, Clark
1480 CGG, Roger AJ, Giezen M Van Der, Clark CGG, et al. 2008. Report Organelles in
1481 *Blastocystis* that Blur the Distinction between Mitochondria and Hydrogenosomes. *Current*
1482 *Biology* 18:580–585.
1483 Stephens TG, Bhattacharya D, Ragan MA, Chan CX. 2016. PhySortR: a fast, flexible tool for
1484 sorting phylogenetic trees in R. *PeerJ* 12;4:e2038.
1485 Stipanuk MH. 1986. Metabolism of sulfur-containing amino acids. *Annu Rev Nutr* 6:179-209.
1486 Swearingen JW, Fuentes DE, Araya MA, Plishker MF, Saavedra CP, Chasteen TG, Vásquez CC.
1487 2006. Expression of the ubiE gene of Geobacillus stearothermophilus V in *Escherichia coli*
1488 K-12 mediates the evolution of selenium compounds into the headspace of selenite- and
1489 selenate-amended cultures. *Appl Environ Microbiol* 72:963–967.
1490 Tachezy J, Doležal P. 2007. Iron–Sulfur Proteins and Iron–Sulfur Cluster Assembly in Organisms
1491 with Hydrogenosomes and Mitosomes. In: Origin of Mitochondria and Hydrogenosomes.
1492 Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105–133.
1493 Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA
1494 transcripts. *Nucleic Acids Res* 43:1–10.
1495 Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to Identify Repetitive Elements in
1496 Genomic Sequences. In: Current Protocols in Bioinformatics. Vol. 25. Hoboken, NJ, USA:
1497 John Wiley & Sons, Inc. p. 4.10.1-4.10.14.
1498 Thiergart T, Landan G, Schenk M, Dagan T, Martin WF. 2012. An evolutionary network of genes
1499 present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial
1500 origin. *Genome Biol Evol* 4:466–485.

1501 Treitli SC., Kolisko M, Husník F, Keeling PJ, Hampl V. 2019. Revealing the metabolic capacity
1502 of Streblomastix strix and its bacterial symbionts using single-cell metagenomics.
1503 *Proceedings of the National Academy of Sciences* 116:19675–19684.
1504 Treitli SC, Kotyk M, Yubuki N, Jirounková E, Vlasáková J, Smejkalová P, Šípek P, Čepička I,
1505 Hampl V. 2018. Molecular and Morphological Diversity of the Oxymonad Genera
1506 *Monocercomonoides* and *Blattamonas* gen. nov. *Protist* 169:744–783.
1507 Treitli SC, Peña-Diaz P, Hałakuc P, Karnkowska A, Hampl V. 2021. High quality genome
1508 assembly of the amitochondriate eukaryote Monocercomonoides exilis. *Microb Genom*
1509 7(12):000745.
1510 Tsaousis AD, Kunji ERS, Goldberg A v, Lucocq JM, Hirt RP, Embley TM. 2008. A novel route
1511 for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature*
1512 453:553–556.
1513 Vacek V, Novák LVF, Treitli SC, Táborský P, Čepička I, Kolísko M, Keeling PJ, Hampl V, Ruiz-
1514 Trillo I. 2018. Fe-S Cluster Assembly in Oxymonads and Related Protists. *Mol Biol Evol*
1515 35(11):2712-2718.
1516 Valasatava Y, Rosato A, Banci L, Andreini C. 2016. MetalPredator: A web server to predict iron-
1517 sulfur cluster binding proteomes. *Bioinformatics* 32(18): 2850–2852.
1518 Vargová R, Hanousková P, Salamonová J, Žihala D, Silberman JD, Eliáš M, Čepička I. 2022.
1519 Evidence for an Independent Hydrogenosome-to-Mitosome Transition in the CL3 Lineage of
1520 Fornicates. *Front Microbiol* 13:1229.
1521 Vargová R, Wideman JG, Derelle R, Klimeš V, Kahn RA, Dacks JB, Eliáš M. 2021. A Eukaryote-
1522 Wide Perspective on the Diversity and Evolution of the ARF GTPase Protein Family. *Genome*
1523 *Biol Evol* 13(8):evab157.

1524 Vlahou G, Eliáš M, von Kleist-Retzow JC, Wiesner RJ, Rivero F. 2011. The Ras related GTPase
1525 Miro is not required for mitochondrial transport in *Dictyostelium discoideum*. *Eur J Cell Biol*
1526 90:342–355.
1527 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
1528 Wortman J, Young SK, et al. 2014. Pilon: An integrated tool for comprehensive microbial
1529 variant detection and genome assembly improvement. *PLoS One* 9:e112963.
1530 Williams K, Lowe PN, Leadlay PF. 1987. Purification and characterization of pyruvate: ferredoxin
1531 oxidoreductase from the anaerobic protozoon *Trichomonas vaginalis*. *Biochem J* 246:529–
1532 536.
1533 Xia L, Björnstedt M, Nordman T, Eriksson LC, Olsson JM. 2001. Reduction of ubiquinone by
1534 lipoamide dehydrogenase. An antioxidant regenerating pathway. *Eur J Biochem* 268:1486–
1535 1490.
1536 Xu F, Jerlström-Hultqvist J, Einarsson E, Ástvaldsson Á, Svärd SG, Andersson JO. 2014. The
1537 Genome of Spironucleus salmonicida Highlights a Fish Pathogen Adapted to Fluctuating
1538 Environments. *PLoS Genet* 10:e1004053.
1539 Yagi T, Shounaka M, Yamamoto S. 1990. Distribution of aspartate aminotrasnsferase activity in
1540 yeasts, and purification and characterization of mitochondrial and cytosolic isoenzymes from
1541 *Rhodotorula marina*. *J Biochem* 107:151–159.
1542 Yazaki E, Kume K, Shiratori T, Eglit Y, Tanifuji G, Harada R, Simpson AGB, Ishida KI,
1543 Hashimoto T, Inagaki Y. 2020. Barthelonids represent a deep-branching metamonad clade
1544 with mitochondrion-related organelles predicted to generate no ATP. *Proceedings of the*
1545 *Royal Society B* 287:20201538.

1546 Zhang Q, Táborský P, Silberman JD, Pánek T, Čepička I, Simpson AGB. 2015. Marine Isolates of
1547 *Trimastix marina* Form a Plesiomorphic Deep-branching Lineage within Preaxostyla,
1548 Separate from Other Known Trimastigids (*Paratrimastix* n. gen.). *Protist* 166:468–491.
1549 Zhou X, Yuan Y, Yang Y, Rutzke M, Thannhauser TW, Kochian L v., Li L. 2009. Involvement of
1550 a broccoli COQ5 methyltransferase in the production of volatile selenium compounds. *Plant*
1551 *Physiol* 151:528–540.
1552 Zítek J, Füssy Z, Treitli SC, Peña-Diaz P, Vaitová Z, Zavadska D, Harant K, Hampl V. 2022.
1553 Reduced mitochondria provide an essential function for the cytosolic methionine cycle.
1554 *bioRxiv* 2022.04.01.486701.
1555 Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček Č,
1556 Hampl V. 2013. The Mitochondrion-Like Organelle of *Trimastix pyriformis* Contains the
1557 Complete Glycine Cleavage System. *PLoS One* 8:e55417.
1558

1559 **FIGURE LEGENDS**

1560 **Figure 1.** Distribution of orthologous groups (OGs) and other discussed features among
1561 metamonads. **A)** Schematic phylogenetic tree of Preaxostyla within Metamonada showing the
1562 distribution of categories of OGs, lateral gene transfer events, and key biological features mapped
1563 to nodes. Core set: estimated relaxed core gene set for Metamonada; increment: clade-specific
1564 increment of the relaxed core gene set; specific: unique OGs for the external leaves (species); new:
1565 novel or extremely divergent genes detected in the species or lineage; LGT: lateral gene transfer
1566 events; SUF: iron-sulfur cluster assembly SUF system; L2: glycine cleavage system protein L2;
1567 FF: flavodoxin-ferredoxin fusion protein; FH: fused hydrogenase; FC: folate cycle; MC:
1568 methionine cycle; PotE: b(0,+)-type amino acid transporter; SPNS: sphingolipid transporter; RFC:
1569 Reduced Folate Carrier; TauE: sulfite exporter TauE/SafE family protein; SP: sugar porters; Hg:
1570 mercury methylation; Se: selenium volatilization; pC: *p*-cresol synthesis; ADI: arginine deiminase;
1571 ADIp: arginine deiminase pathway; CK: carbamate kinase. **B)** Venn diagram showing the
1572 distribution of OGs among the three main analyzed lineages of Metamonada. **C)** Venn diagram
1573 showing the distribution of OGs among the three lineages of Preaxostyla. **D)** Venn diagram
1574 showing the distribution of OGs among the three studied species of Oxymonadida. The IDs and
1575 functional annotation of OGs is given in Supplementary file 2. The IDs and functional annotation
1576 of LGT events is given in Supplementary file 3.

1577 **Figure 2.** Schematic representation of the structure of HgcAB protein in *P. pyriformis* and its
1578 comparison to the prokaryotic HgcA-HgcB operon. The upper panel gives the overview of the full
1579 alignment, the middle panel zooms on the primary structure of the functionally most important
1580 regions, and the lower panel displays the complete domain structure.

1581 **Figure 3.** Complement of Golgi-associated proteins in Preaxostyla. This Coulson plot shows the

1582 set of proteins present in the Preaxostyla predicted proteomes. Empty segments denote failure to
1583 identity a candidate orthologue, while filled segments denote success, with paralogue numbers inset
1584 as relevant. Candidate proteins are identified by homology-searching and verified by phylogenetics
1585 as relevant. Details are given in Supplementary file 6.

1586 **Figure 4.** Hypothetical energy metabolism in Preaxostyla. The glycolysis reactions are simplified.
1587 Bold outline indicates alternative glycolytic enzymes. Abbreviations and Enzyme Commission
1588 numbers are given in Supplementary file 6. Presence of the enzymes in Preaxostyla data sets is
1589 indicated by a color code.

1590 **Figure 5.** Phylogenetic relationship among hydrogenases of Preaxostyla. A) Schematic
1591 representation of domain architectures of hydrogenases among Preaxostyla species; domain
1592 architecture is indicated. HydA: [FeFe] hydrogenase; CysJ: NADPH-dependent sulfite reductase;
1593 NuoG: NADH-quinone oxidoreductase. B) Detailed view on the part of the tree comprising the
1594 clade of long *P. pyriformis* hydrogenases. The domain architecture of the proteins is indicated by
1595 colour bars. The full tree is given in Supplementary file 4.

1596 **Figure 6.** Summary of the searches for proteins physically associated with MROs. In the upper
1597 panel, the numbers of candidates recovered by the four strategies for the five Preaxostyla species
1598 (colour-coded) are summarized. Detailed information on the searches and candidates is given in
1599 Supplementary file 9. In the lower panel, the mitochondrion hallmark proteins detected in the data
1600 sets of *P. pyriformis* and *T. marina* are summarized. No such candidate was recovered for any
1601 oxymonad representative.

1602 **SUPPLEMENTARY FIGURE LEGENDS**

1603 **Figure S1.** Hypothetical map of amino acid metabolism in *P. pyriformis*. Brown color indicates
1604 enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes
1605 possibly involved in ATP production. Note that some of the connections between metabolites
1606 correspond to the mere transfer of the amino group rather than conversion of the carbon backbone
1607 of the molecule. Abbreviations and Enzyme Commission numbers are given in Supplementary file
1608 6.

1609 **Figure S2.** Hypothetical map of amino acid metabolism in *T. marina*. Brown color indicates
1610 enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes
1611 possibly involved in ATP production. Abbreviations and Enzyme Commission numbers are given
1612 in Supplementary file 6.

1613 **Figure S3.** Hypothetical map of amino acid metabolism in *M. exilis*, *B. nauphoetae*, and *S. strix*.
1614 Brown color indicates enzymes possibly involved in amino acid biosynthesis pathways. Red color
1615 indicates enzymes possibly involved in ATP production. Abbreviations and Enzyme Commission
1616 numbers are given in Supplementary file 6.

1617 **Figure S4.** Venn diagram showing the distribution of orthologous groups (OGs) of Fe-S cluster-
1618 containing proteins among the species of Preaxostyla. The identity of the OGs and of the
1619 component proteins are provided in Supplementary file 8.

1620 **TABLES**

1621 **Table 1. General features of the Preaxostyla genomes discussed in this study**

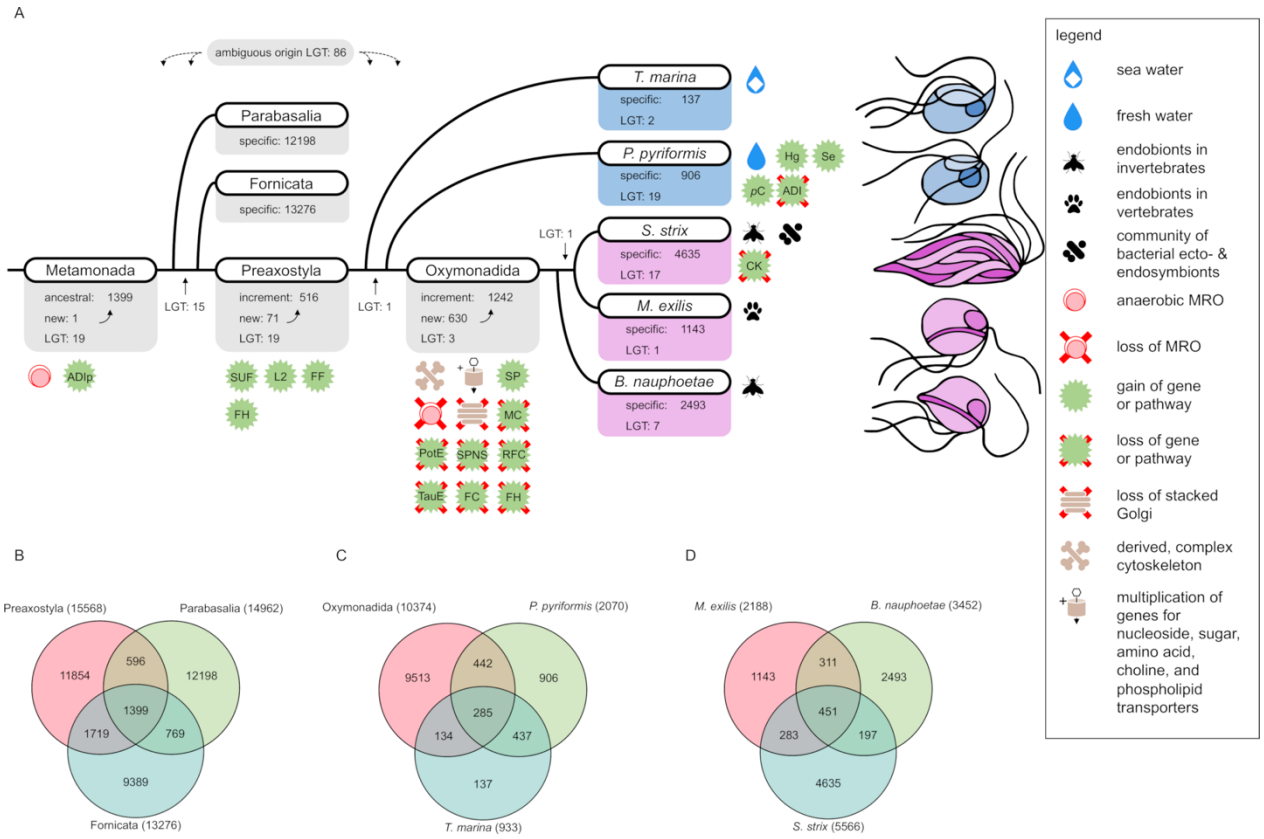
Sample	Scaffolds	Total length (bp)	N50 (kbp)	Completeness (%; BUSCO odb9)	G+C content (%)	Protein-Coding Loci	NCBI BioProject	NCBI BioSample
P. pyriformis	633	56,627,582	276.6	82.1	60.94	13,466	PRJNA903905	SAMN31819772
B. nauphoetae	879	88,537,989	199.5	76.6	44.96	25,221	PRJNA887011	SAMN31149887
M. exilis	2,092	74,712,536	71.44	75.3	36.8	16,768	PRJNA304271	SAMN04297179
S. strix	50,889	152,152,197	5.18	69.6	26.6	56,706	PRJNA524138	SAMN10998475

1622

1623

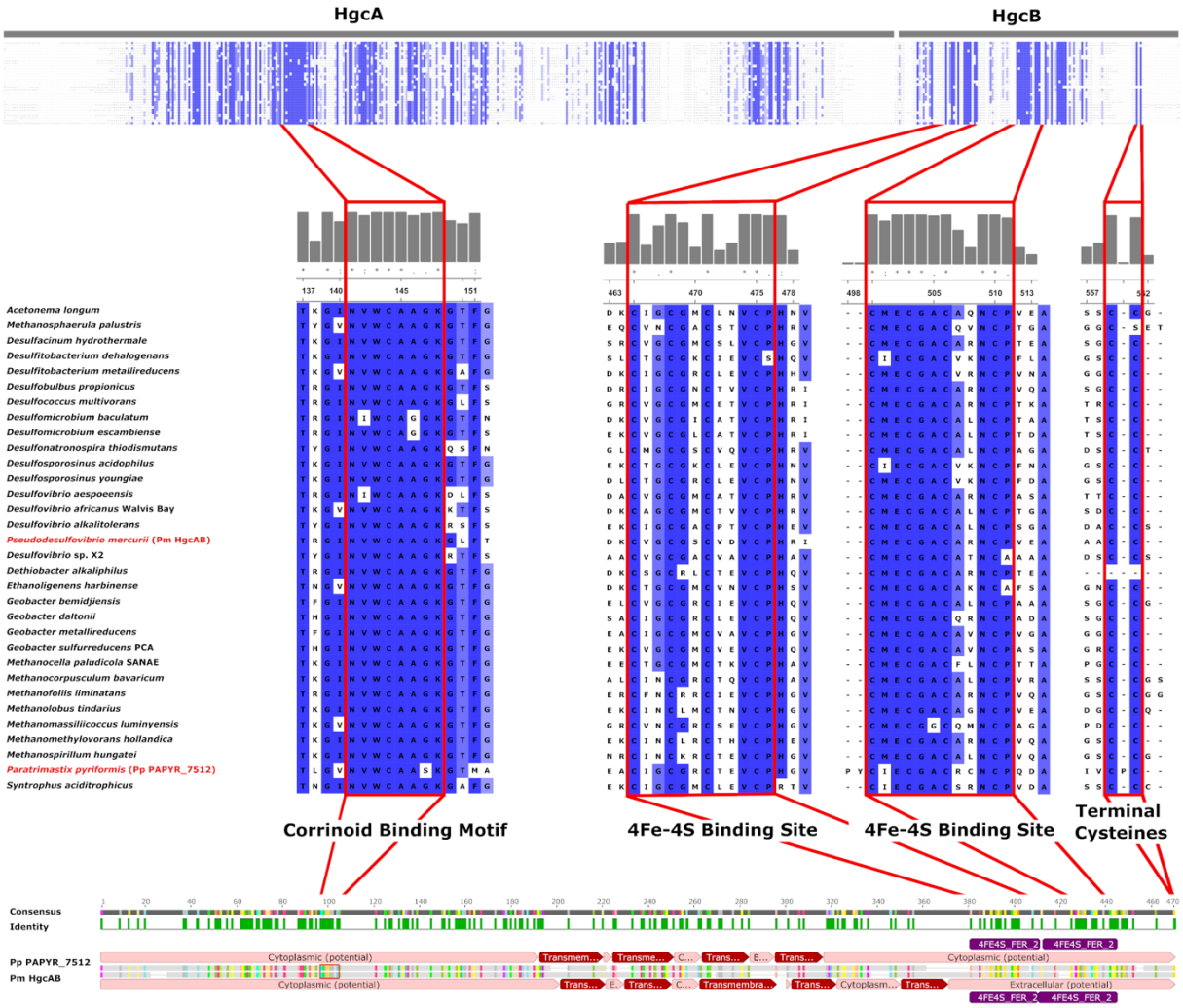
1624 **FIGURES**

1625 **Figure 1**

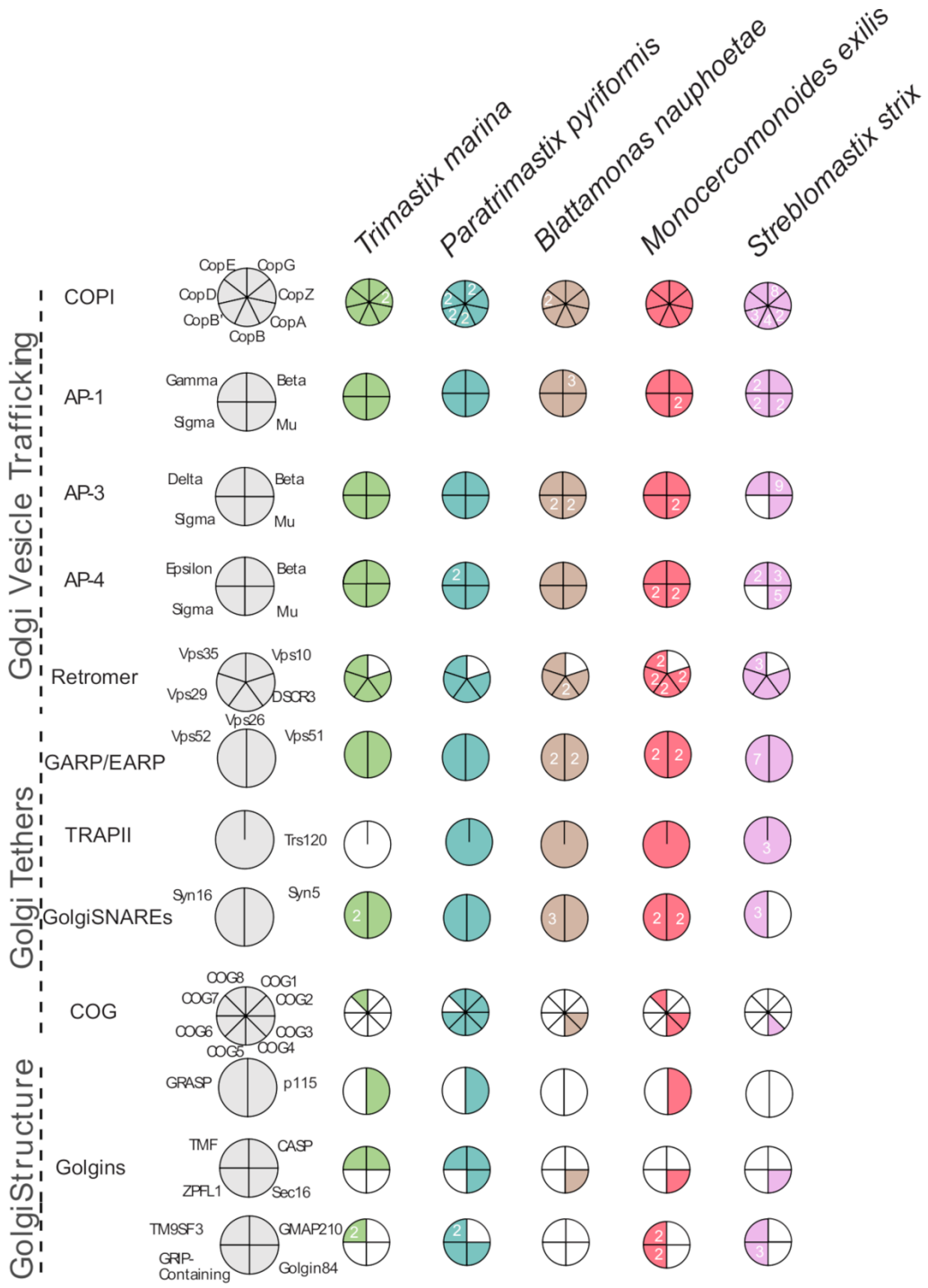


1626

Figure 2

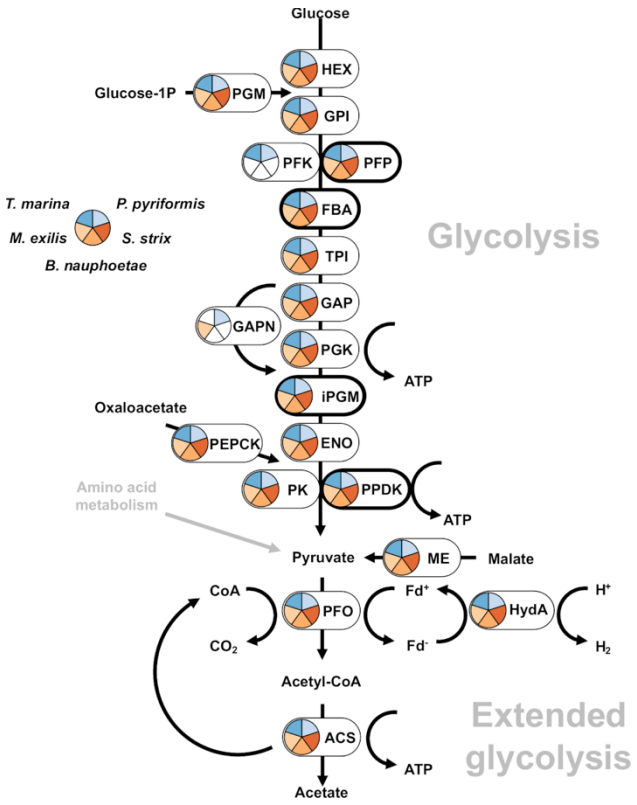

HgcA
HgcB
Acetonema longum
Methanosphaerula palustris
Desulfacinum hydrothermale
Desulfitobacterium dehalogenans
Desulfitobacterium metallireducens
Desulfobulbus propionicus
Desulfococcus multivorans
Desulfomicrobium baculatum
Desulfomicrobium escambiense
Desulfonatronospira thiodismutans
Desulfosporosinus acidophilus
Desulfosporosinus youngiae
Desulfovibrio aespoeensis
Desulfovibrio africanus Walvis Bay
Desulfovibrio alkalitolerans
Pseudodesulfovibrio mercurii (Pm HgcAB)
Desulfovibrio sp. X2
Dethiobacter alkaliphilus
Ethanoligenens harbinense
Geobacter bemidjiensis
Geobacter daltonii
Geobacter metallireducens
Geobacter sulfurreducens PCA
Methanocella paludicola SANAE
Methanocorpusculum bavaricum
Methanofollis liminatans
Methanolobus tindarius
Methanomassiliicoccus luminyensis
Methanomethylovorans hollandica
Methanospirillum hungatei
Paratrimastix pyriformis (Pp PAPYR_7512)
Syntrophus aciditrophicus
Corrinoid Binding Motif
4Fe-4S Binding Site
4Fe-4S Binding Site
Terminal Cysteines
Consensus
Identity
Pp PAPYR_7512
Pm HgcAB
Cytoplasmic (potential)
Extracellular (potential)
4FE4S_FER_2

1629 **Figure 3**



1630

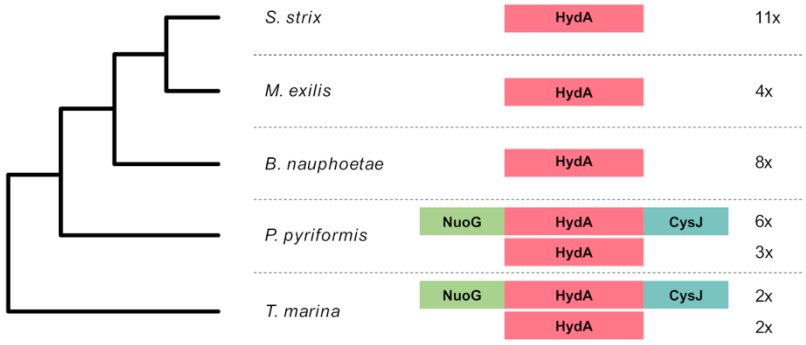
1631 **Figure 4**



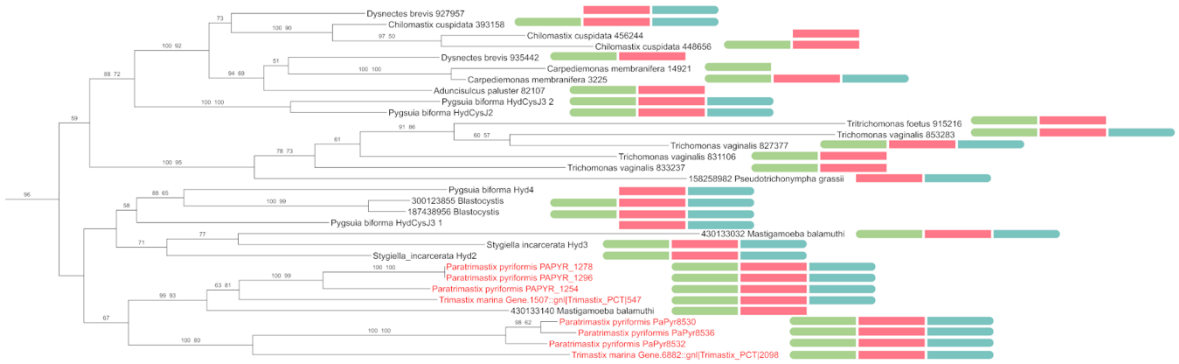
1632

1633 **Figure 5**

A



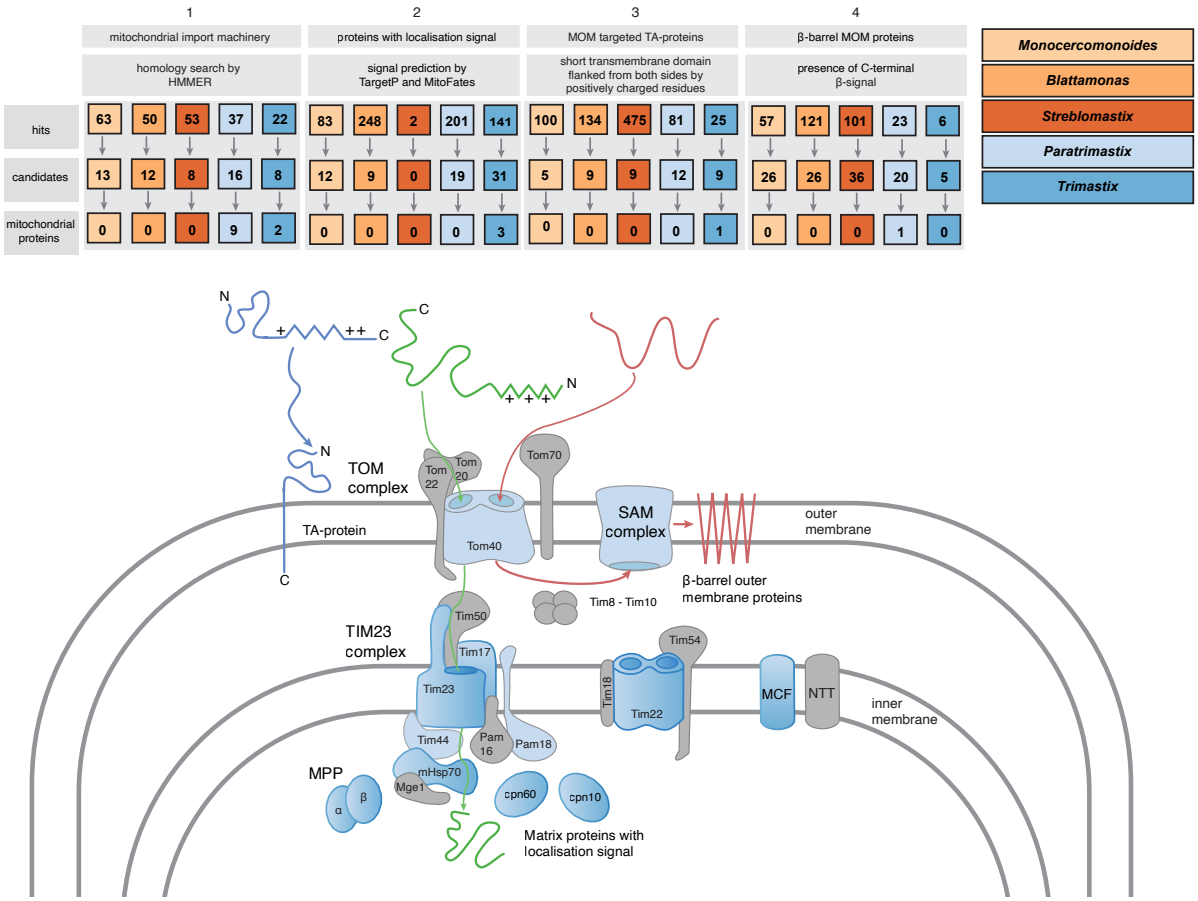
B



1634

1635 **Figure 6**

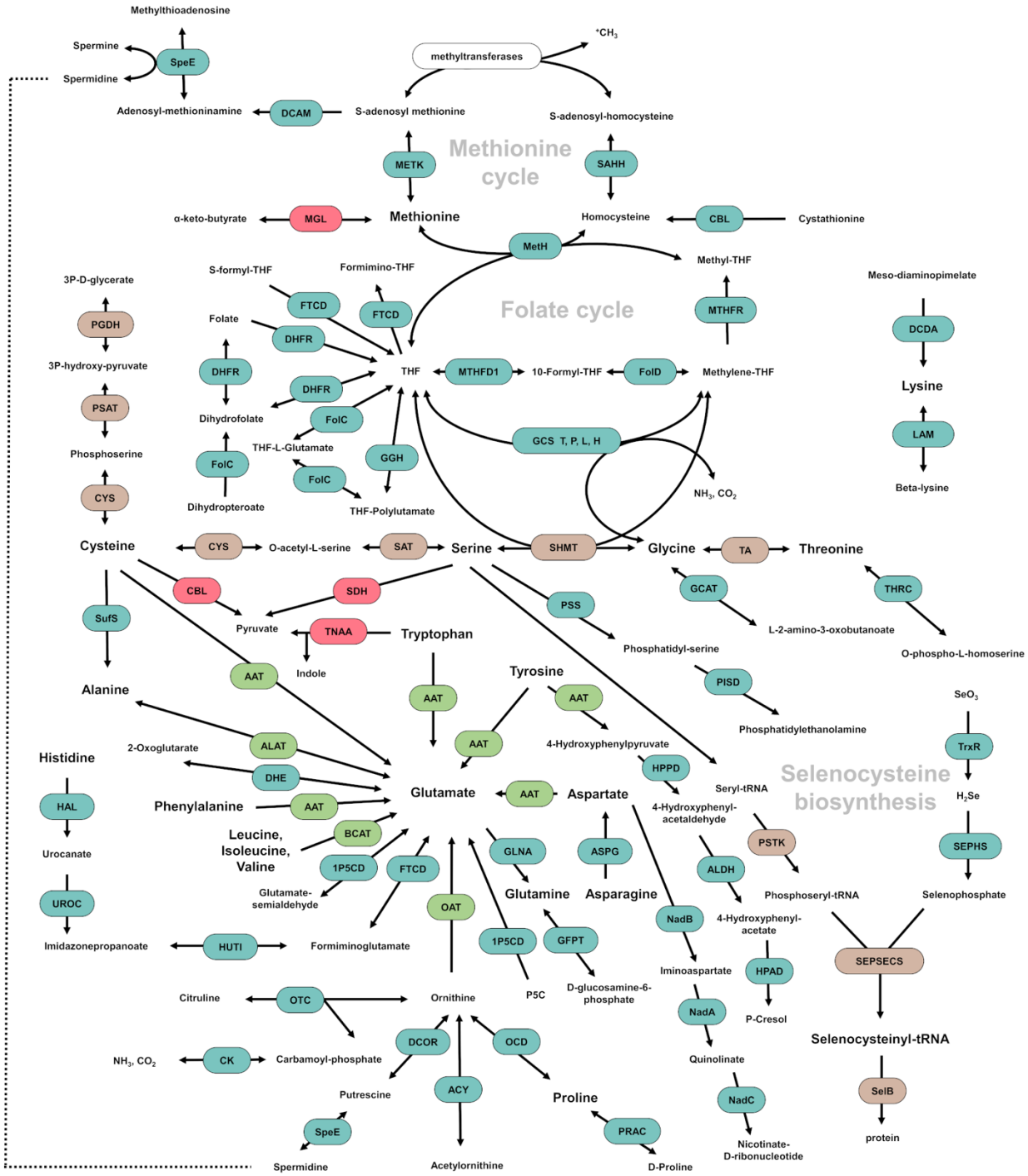
1636



1637

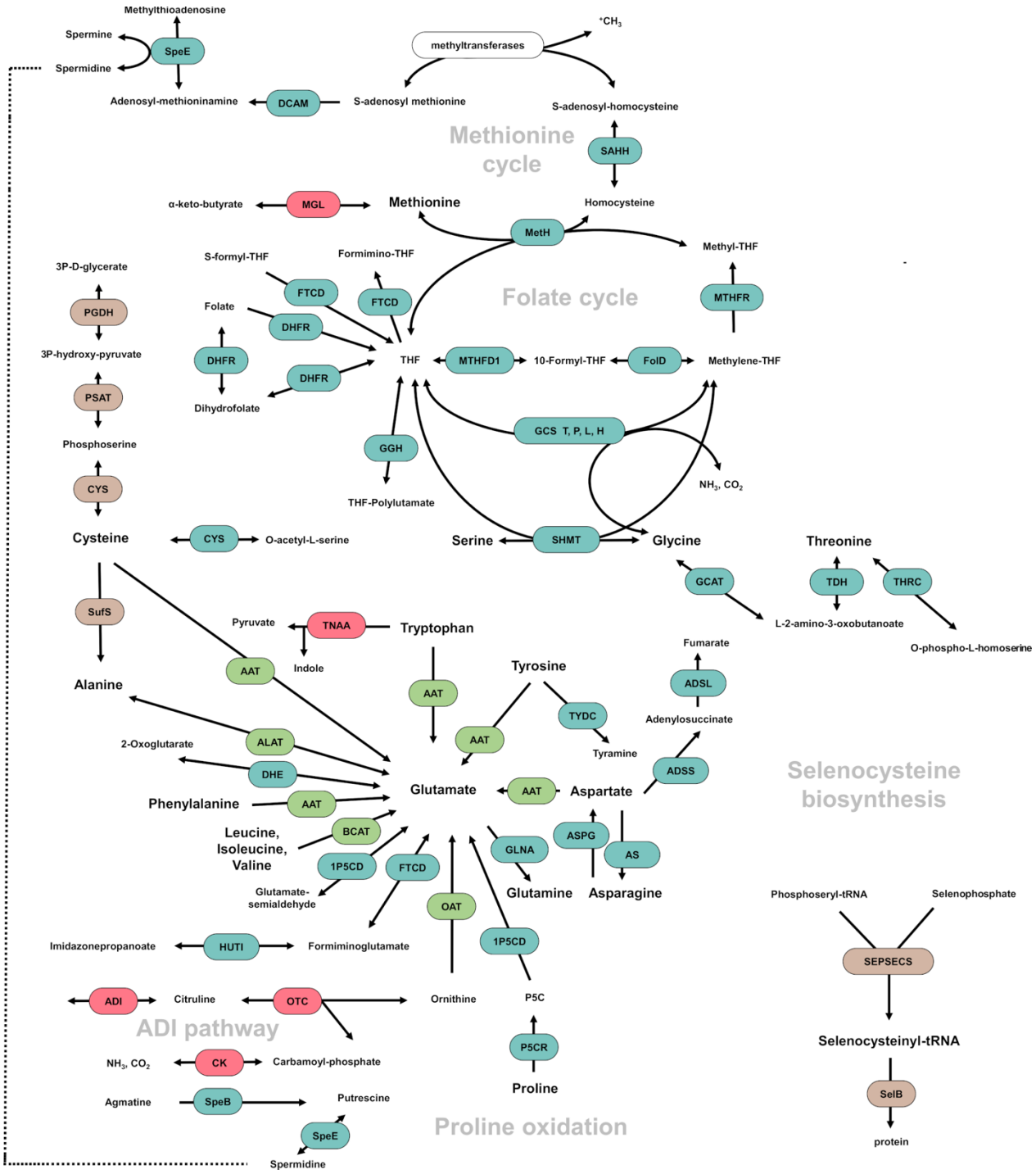
1638 **SUPPLEMENTARY FIGURES**

1639 **Figure S1**



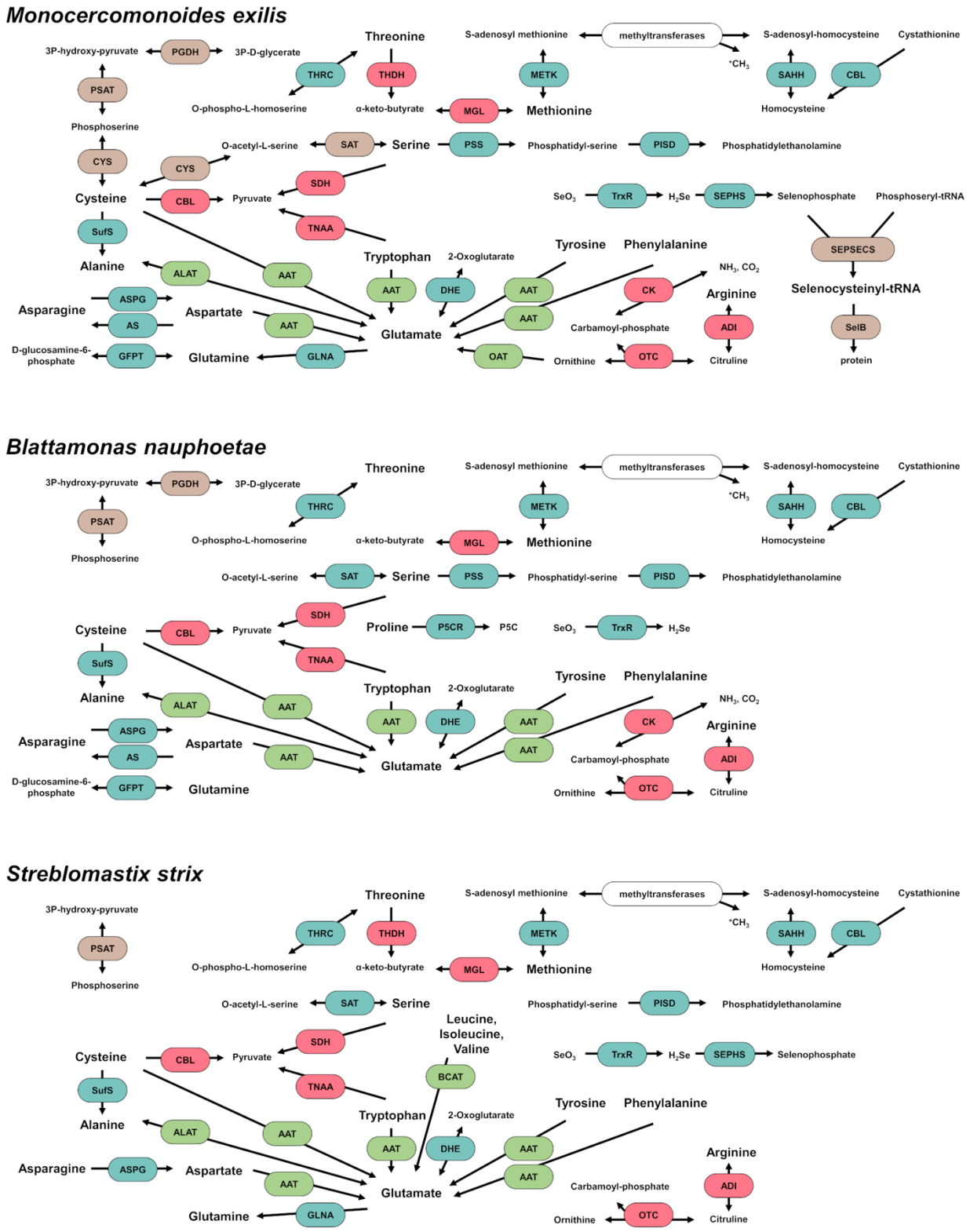
1640

1641 **Figure S2**



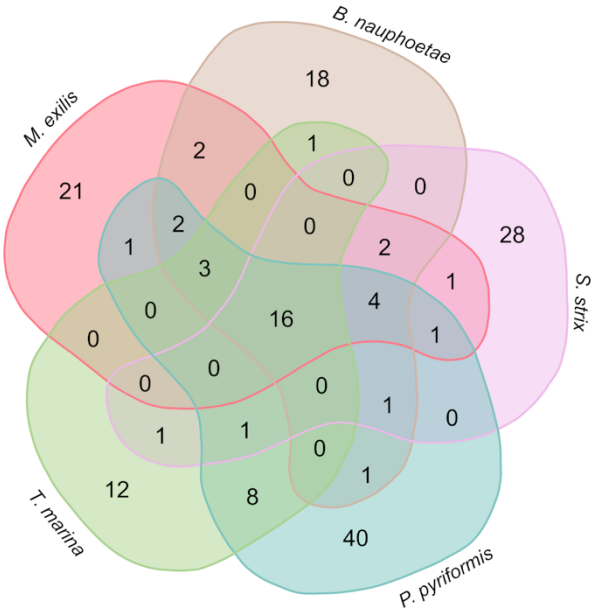
1642

1643 **Figure S3**



1644

Figure S4


B. nauphoetae
M. exilis
S. sthx
T. marina
P. pyriformis
18
21
28
12
40
2
1
0
0
0
2
0
2
1
3
1
4
0
16
1
0
0
0
0
1
0
1
1
0
1
8