Subcellular compartments interplay for carbon and nitrogen allocation in *Chromera velia* and *Vitrella brassicaformis*

Zoltán Füssy[1,2,*], Tereza Faitová[1,2,3], and Miroslav Oborník[1,2,*]

[1]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

[2]Institute of Parasitology, Biology Centre CAS, České Budějovice, Czech Republic

[3]Faculty of Engineering and Natural Sciences, Johannes Kepler University, Linz, Austria

* corresponding authors: zoltan.fussy@gmail.com, obornik@paru.cas.cz

ABSTRACT

Endosymbioses necessitate functional cooperation of cellular compartments to avoid pathway redundancy and streamline the control of biological processes. To gain insight into the metabolic compartmentation in chromerids, phototrophic relatives to apicomplexan parasites, we prepared a reference set of proteins probably localized to mitochondria, cytosol and the plastid, taking advantage of available genomic and transcriptomic data. Training of prediction algorithms with the reference set now allows a genome-wide analysis of protein localization in *C. velia* and *V. brassicaformis*. We confirm that the chromerid plastids house enzymatic pathways needed for their maintenance and photosynthetic activity, but for carbon and nitrogen allocation, metabolite exchange is necessary with the cytosol and mitochondria. This indeed suggests that the regulatory mechanisms operate in the cytosol to control carbon metabolism based on the availability of both light and nutrients. We discuss that this arrangement is largely shared with apicomplexans and dinoflagellates, possibly

stemming from a common ancestral metabolic architecture, and supports the mixotrophy of the chromerid algae.

KEYWORDS: chromerid, protein localization, prediction algorithm, endosymbiosis, plastid integration, mixotrophy

INTRODUCTION

Endosymbiotic organelles play crucial roles in cellular biochemistry. Mitochondria, alpha-proteobacterial endosymbionts of eukaryotes (Oborník 2019; Gruber 2019), represent an energetic hub, where balancing of catabolic and anabolic processes takes place tightly regulated with the speed of respiration (e.g. Searcy 2003; Pagliarini and Rutter 2013; Gray 2015). In plastids, domesticated cyanobacteria (Gruber 2019; Oborník 2019), inorganic carbon is fixed into sugars and several essential compounds are synthesized, such as fatty acids, isoprenoid units, tetrapyrroles, and amino acids (e.g. Tetlow et al. 2017). These semiautonomous endosymbiont-derived organelles contain genetic information and their own translation apparatuses, but by far do not encode all the proteins required for their function. Due to endosymbiotic gene transfer, most essential genes were transferred from their genomes to the nuclear genome and the organelles are greatly dependent on the import of proteins synthesized in the cytosol (e.g. Mallo et al. 2018). Sorting of proteins to subcellular locations specifically is therefore crucial for the correct function of both the proteins and the organelles, and thus targeting signals and protein translocation are key to our understanding of organellar biology (Kunze & Berger 2015).

Proteins destined to plastids and mitochondria typically encode an N-terminal motif, i.e. a targeting presequence. The targeting presequence of mitochondrial proteins is termed the mitochondrial transit peptide (mTP) and has the physicochemical properties of an amphiphilic helix. Similarly, the chloroplast transit peptides of primary algae (rhodophytes, chlorophytes and glaucophytes) and plants are amphiphilic helices, though they are typically enriched in hydroxylated amino acids and less positively charged than mTPs (Kunze & Berger 2015; Garg & Gould 2016). In comparison, complex algae (those that maintain eukaryotic endosymbionts) including the chromerids, target

proteins to the plastid via the endomembrane (secretory) pathway, using chloroplast transit peptides directly preceded by an ER signal peptide, which are referred to as bipartite targeting sequences (BTS) (reviewed in Patron & Waller 2007).

Reconstruction of ancestral traits (Joy et al. 2016) allows us to unveil changes in lifestyle and genome organization in an evolutionary perspective and to compare functionalities among the organisms of interest. The discovery and genome characterization of chromerids *Chromera velia* and *Vitrella brassicaformis*, the closest known photosynthetic relatives of apicomplexan parasites, have provided an excellent framework to study the transition from free-living phototrophs to obligate parasites (Moore et al. 2008; Oborník et al. 2009; Janouškovec et al. 2010; Burki et al. 2012; Janouškovec et al. 2015; Woo et al. 2015; Füssy & Oborník 2017b). Much knowledge has accumulated about the function of the apicomplexan remnant plastid, the apicoplast (reviewed in Boucher et al. 2018), which given their shared origin structurally and molecularly resembles the photosynthetic plastid of chromerids (Moore et al. 2008; Janouškovec et al. 2010; Woo et al. 2015). Nevertheless, the protein composition of the chromerid plastid is largely unknown, except for a recent work that focused on *C. velia* photosystems (Sobotka et al. 2017), and therefore a pre-transition model of the apicoplast could not be studied in detail. The mitochondrial genome of apicomplexans is massively reduced in gene content and found to contain only three protein-coding genes, *cox1*, *cox3* and *cyb*, with the majority of mitochondrial proteins requiring import from cytosol (Nash et al. 2008; Janouškovec et al. 2013; Flegontov et al. 2015). Strikingly, the *C. velia* mitochondrion was found to contain only two of these genes, *cox1* and *cox3* (Flegontov et al. 2015). It has been hypothesized that the reduction of the apicomplexan (and chromerid) mitochondrial genome could be linked with the change in lifestyle strategy, particularly a change to facultative anaerobiosis (Dorrell et al.

2013), which is consistent with the observed reduction of the respiratory chain in all myzozoans (dinoflagellates, chromerids and apicomplexans) (Flegontov et al. 2015; Oborník & Lukeš 2015).

The relatively small nuclear genome size (up to 193 Mb) and largely complete sequence data of chromerids (Woo et al. 2015) make them ideal for large-scale targeting signal recognition and, by extension, organellar proteome prediction. Up to now, plastid proteomes have been determined in only a handful of organisms, mainly plants and green algae (van Wijk & Baginsky 2011; Terashima et al. 2011; Dorrell et al. 2017), but also a handful of complex algae and protist parasites (Hopkins et al. 2012; Boucher et al. 2018). Similarly, mitochondrial proteomic data are rather scarce and focused on model organisms, such as humans (Calvo & Mootha 2010; Palmfeldt & Bross 2017), yeast (Gonczarowska-Jorge et al. 2017), plants (Huang et al. 2013), and protists (Smith et al. 2007; Atteia et al. 2009; Panigrahi et al. 2009; Danne et al. 2013; Gawryluk et al. 2014).

The aim of the work is to define and characterize the subcellular proteomes of chromerids by bioinformatic tools with an emphasis on plastid- and mitochondrion-destined proteins. For the analysis, we compiled sets of compartment-specific proteins of *C. velia* and *V. brassicaformis* and optimized the performance of the ASAFind (Gruber et al. 2015) prediction tool with the reference amino acid frequency matrices. Our analyses bring first implications on carbon and nitrogen allocation among the plastid, cytosol and mitochondria in chromerids, suggesting interplay of these compartments is in place for efficient carbon metabolism under changing light and nutrient conditions. This work also confirms biochemical peculiarities ancestrally shared with apicomplexans and dinoflagellates, such as the lack of the canonical mitochondrial pyruvate decarboxylase and cytosolic amino acid synthesis.

DATA SOURCES AND METHODS

The sequence data of the chromerid algae *Chromera velia* CCMP2878 and *Vitrella brassicaformis* CCMP3155 were retrieved from CryptoDB (www.cryptodb.org, version 34). Additional transcriptomic data were retrieved from NCBI GenBank (Dorrell et al. 2014; Woehle et al. 2011) and MMETSP sequence databases (MMETSP0290 and MMETSP1451, (Keeling et al. 2014; Cohen et al. 2016). The sequence data were annotated using the information available at KEGG servers (Kanehisa et al. 2017) and using the InterProScan annotation tool of Geneious (Jones et al. 2014; Kearse et al. 2012).

Plastid-targeted reference sequences were identified based on several lines of evidence; a) the protein had a clear role in the plastid metabolism (in synthesis of pigments and cofactors, or being a subunit of the photosynthetic machinery, etc.), with an emphasis on filling the gaps between well-defined enzymatic steps; b) reassuring sequence completeness, an N-terminal extension (40-80 amino acids) that could possibly encode a bipartite targeting sequence preceded the mature protein (as determined by InterProScan), though essentially no targeting sequence prediction was employed to avoid circular reasoning (including predictor-positive proteins and using them to evaluate this predictor); c) there was a phylogenetic relationship to another plastid-targeted protein among chromerids or Apicomplexa (in case of ribosomal proteins; Gupta et al. 2014). Mitochondrial references were compiled similarly, only the N-terminal extension was found shorter. Cytosolic references lacked an extension and secretory proteins had an identifiable role in the endomembrane system or at the cytoplasmic membrane. Metabolic gaps were filled by targeted BLAST searches in the genomic (CryptoDB) as well as transcriptomic data (CryptoDB, GenBank, MMETSP) using known-function apicomplexan sequences and KEGG orthologs as queries.

To define the best tool for the subcellular localization of proteins, the sets of reference sequences of *Chromera* and *Vitrella* were analyzed by prediction algorithms. The tools tested were selected to be suitable for large-scale analyses and included: TargetP (Emanuelsson et al. 2000), SignalP (v. 4.1) (Petersen et al. 2011), ASAFind (Gruber et al. 2015), HECTAR v1.3 (Gschloessl et al. 2008), MultiLoc2 (Blum et al. 2009), PrediSi (Hiller et al. 2004) and PredSL (Petsalaki et al. 2006). All the prediction algorithms except HECTAR were run locally with default parameters; SignalP was run with sensitive cutoff values (-u 0.3 -U 0.3). The sensitivity (proportion of recognized true positives) and precision (proportion of positive results, also termed the positive predictive value) of the prediction algorithms were compared, setting a certain threshold specified for each of the predictors. Sensitivity was computed as:

(Eq.1) $sensitivity = (true\ positives) / (true\ positives + false\ negatives)$

and positive predictive value was computed as:

(Eq.2) $precision = (true\ positives) / (true\ positives + false\ positives).$

An optimal threshold would cover maximum positive proteins while including a low number of false positives (proteins falsely predicted to the organelle in question).

Bit score-weighted matrices of amino acid positions surrounding the signal cleavage sites were calculated separately for *C. velia* and *V. brassicaformis*, as described by Crooks et al. (2004). Only cleavage sites which were consensually (by majority agreement) predicted by PrediSi, PredSL, SignalP and ASAFind were considered. The transit peptide sequence logos and frequency plots of plastid-targeted proteins from *C. velia* and *V. brassicaformis* were created with WebLogo (Crooks et al. 2004; http://weblogo.berkeley.edu/; version 2.8.2).

Where applicable, closest hits for proteins were found using BLAST against nr or RefSeq databases, and using DIAMOND (Buchfink et al. 2014) against an in-house made

159 database consisting from sequences collected from NCBI, MMETSP (Keeling et al. 2014;
160 Cohen et al. 2016) and Ensembl Genomes (release 37; Kersey et al. 2016). Sequences
161 were aligned using MAFFT v.7 (Katoh & Standley 2013) and automatically trimmed by
162 trimAL (Capella-Gutiérrez et al. 2009). Maximum likelihood trees were inferred from the
163 trimmed alignments using the best-fitting substitution model as determined by the IQ-
164 TREE –m TEST option limited to LG matrix by -mset (Nguyen et al. 2015). Branch
165 supports were determined by rapid bootstrapping followed by 1,000 ultra-fast
166 bootstrap replicates (-bb 1000).
167
168
169 RESULTS
170 **Sequence completeness and reference compilation**
171 During the annotation and alignment of reference proteins, we noticed that some
172 contigs retrieved from CryptoDB (Woo et al. 2015) are apparently truncated at their 5'-
173 ends. The sequence data appear gene-rich but are still highly fragmented, with 5,966
174 and 1,064 genomic scaffolds assembled for *C. velia* and *V. brassicaformis*, respectively,
175 although the number of chromosomes in *C. velia* is estimated to be much smaller (Vazač
176 et al. 2017). Since it is essential to identify a protein's complete N-terminus to predict its
177 localization to the plastid or to mitochondria (see Introduction), we pursued an
178 independent assessment of N-termini completeness. We searched for (almost) identical
179 transcripts in chromerid transcriptomes available in GenBank and MMETSP and used
180 these contigs to extend those from CryptoDB towards their 5'-end, where possible.
181 While these fused contigs are indeed chimeras of orthologs from different strains (the
182 transcripts were not completely identical), we assume they code for *bona fide* N-termini
183 in all these sequenced strains of *C. velia* and *V. brassicaformis*. These chimeric contigs are

184 for clear distinction marked in the reference sequence list (Supplementary Table S1).
185 Other, apparently truncated contigs were omitted from the reference sets.

186 We performed a systematic search of housekeeping and metabolically active
187 proteins to obtain the reference sets from *C. velia* and *V. brassicaformis* data. To find
188 plastid references, we searched for pathways intimately linked with their biogenesis and
189 photosynthesis, including the biosynthesis of fatty acids (type II FAS) and lipids, iron-
190 sulfur clusters, terpenoids and terpenoid derivatives (photosynthetic pigments,
191 vitamins), tetrapyrrole cofactors, protein translocator subunits and components of the
192 photosynthetic electron transport chain (recently characterized by Sobotka et al. 2017)
193 (Fig. 1), and the enzymes of carbon fixation (Fig. 2). Among mitochondrial pathways, we
194 looked for the enzymes of the tricarboxylic acid (TCA) cycle and the components of the
195 respiratory chain (Flegontov et al. 2015) (Fig. 2). From pathways localized in the
196 cytosol, we identified the enzymes of glycolysis, storage amylopectin biosynthesis and
197 breakdown (Coppin et al. 2005), and the enzymes of the pentose phosphate cycle that
198 we distinguished from the enzymes of the plastid carbon fixation by their N-terminal
199 extensions (Fig. 2). Using phylogenetic analyses (not shown), we also identified
200 ribosomal proteins from all three translationally active compartments (Gupta et al.
201 2014). We also included biosynthesis pathways of several amino acids and other
202 compounds, where enzymatic steps showed consistent localization (Fig. 2). Organellar
203 genome-encoded proteins were dropped from the reference sets (Supplementary
204 Table 1). In a scarcity of experimental data, we believe this is the most reliable approach
205 to compile reference sequences.

206 [insert Fig. 1]

207 Though most of the searched pathways are near-complete, we failed to identify
208 representatives of some enzymatic steps in the available transcriptomic and genomic

209 data. These include the acyl-[acyl carrier protein (ACP)] ligase (in both chromerids) and
210 two lipid phosphatases of the glycero(phospho)lipid biosynthesis (phosphatidate
211 phosphatase in *C. velia*, phosphatidylglycerophosphatase in *V. brassicaformis*), several
212 enzymes of the carotenoid (neoxanthin synthase), terpenoid-quinone (tocopherol
213 cyclase), and amino acid biosynthesis pathways (undescribed steps of Lys and Pro
214 synthesis), and one enzyme of the pentose phosphate cycle (cytosolic ribose 5-
215 phosphate isomerase; Fig. 1, Fig. 2). Some of these steps are possibly catalyzed by
216 distantly related enzymes that were not recognized by our searches, but other absences
217 may be of biological relevance. Ribose 5-phosphate isomerase is necessary to recycle
218 ribulose 5-phosphate, but a cytosolic isoform was not recovered, and maybe
219 alternatively spliced transcripts encode for a truncated, non-plastid-targeted protein.
220 Similarly, the final steps of fucoxanthin biosynthesis are missing. Although there are
221 biochemical data that *Chromera* and *Vitrella*, respectively, accumulate (iso)fucoxanthin
222 and vaucheriaxanthin – both derivates of neoxanthin – as accessory photosynthetic
223 pigments (Moore et al. 2008; Oborník et al. 2012), neoxanthin synthase was not found in
224 our data, while the last enzymes of the respective pathways are unknown (Mikami &
225 Hosokawa 2013). Plant neoxanthin synthase is a neofunctionalized lycopene cyclase
226 (Bouvier et al. 2000), opening the possibility that a promiscuous activity of the latter
227 enzyme is responsible for neoxanthin synthesis in chromerids. In comparison, the
228 absence of acyl-ACP synthase might be compensated if fatty acids are not released from
229 the bond with the ACP but rather directly used for lipid synthesis (Bisanz et al. 2006).
230 Alternatively, intermediate glycerolipids (diacylglycerol esterified with fatty acids of
231 various length and saturation) could be imported from other cell compartments (Jouhet
232 et al. 2007).

233 [insert Fig. 2]

234

235 **Prediction performance**

236 ASAFind and HECTAR (Gruber et al. 2015; Gschloessl et al. 2008) were designed
237 to predict the BTS of plastid proteins in complex algae, particularly in heterokonts.
238 ASAFind identifies plastid proteins based on the output of SignalP (Petersen et al. 2011)
239 and a sliding-window scan for the highly conserved Phe residue around the predicted
240 cleavage site. HECTAR uses a combination of predictors in three decision modules and
241 aims to classify proteins based on presence of four types of targeting modules: signal
242 peptides, type II signal anchors, chloroplast transit peptides, and mitochondrial transit
243 peptides. To predict the mTP, several predictors are available. These include TargetP,
244 MultiLoc2, and HECTAR (Emanuelsson et al. 2000; Blum et al. 2009; Gschloessl et al.
245 2008).

246 Both ASAFind and HECTAR were specifically trained on stramenopile sequences,
247 though ASAFind offers the possibility to use an alternative bit score matrix derived from
248 a training set. TargetP and MultiLoc2 were trained using plant and animal sequences.
249 Having reference dataset at hand, we could test the performance of these algorithms on
250 chromerid datasets, which has been unknown. Our results are summarized in Table 1
251 (full results in Supplementary Table S2). Mitochondrial predictors offered only
252 moderate sensitivity and precision, with *V. brassicaformis* sequences being better
253 resolved. We defined two thresholds for each predictor, one having higher sensitivity
254 (around 75 %), the other more selective (with around 85 % precision). HECTAR offers
255 higher precision with the 75%-level sensitivity compared to other mitochondrial
256 predictors (Supplementary Fig. S3). With plastid sequences, ASAFind and HECTAR
257 performed comparably, having both high sensitivity and precision. However, it must be
258 noted that the threshold for HECTAR is very low and most of our plastid controls were

marked as "secretory proteins" by this predictor. Our datasets might be over-fit by lacking enough endomembrane system proteins as negative controls, therefore precision is expected to drop with broader reference sets. We also used our plastid references to derive bit score-weighted matrices of amino acids flanking the cleavage site independently for *C. velia* and *V. brassicaformis* (as in Gruber et al. 2015). Using this matrix, we could further improve the performance of ASAFind (Table 1). This suggests that indeed chromerid plastid-targeting presequences differ from those of stramenopiles, but possibly also from each other.

[insert Table 1]

To visualize the amino acid enrichment around the signal cleavage site, we created logo-plots of amino acid frequencies in this motif in chromerids. We found that both chromerids have a conserved Phe residue just after the cleavage site, though in *C. velia* the Phe motif is more frequently found (Fig. 3).

[insert Fig. 3]

Plastid housekeeping

Most of the enzymes we expected to localize in chromerid plastids are directly or indirectly related to photosynthesis. These include the photosystems core subunits, the proteins of light-harvesting antennae and the photosynthetic electron transport chain, but also insertase proteins TAT, SEC and SRP/Alb3 that embed these factors in the thylakoid membrane (Fig. 2, discussed in further detail by Sobotka et al. 2017). Furthermore, we detected the many enzymatic steps that produce photosystems cofactors and accessory pigments (Fig. 1). Tetrapyrrole synthesis is one of the prime biochemical pathways of plastids, since heme is a vital component to the electron transport chain and retrograde plastid signaling, while chlorophyll is a fundamental

284 cofactor in light harvesting complexes (de Souza et al., 2017). In *C. velia*, the tetrapyrrole
285 pathway starts curiously with the delta-aminolevulinic acid (ALA) synthesis from
286 glycine and succinyl-CoA in the mitochondrion, but the rest of the pathway is predicted
287 to reside in the plastid (Kořený et al. 2011). This feature is shared by *V. brassicaformis*
288 and indeed supported by our results (Fig. 1). Heme is also required for the function of
289 cytosolic and mitochondrial hemoproteins (e.g., respiratory chain components), so it is
290 exported from the chromerid plastid, as there is apparently no tetrapyrrole synthesis
291 activity in other compartments besides ALA synthesis.

292 Chlorophyll *a* (Chl *a*) is the only chlorophyll species employed by chromerids
293 (Moore et al. 2008). During the last step of Chl *a* synthesis, a terpenoid (phytyl) moiety is
294 attached to the Mg^{2+}-coordinated tetrapyrrole, chlorophyllide *a*. Phytyl diphosphate is
295 synthesized from methyl-D-erythritol 4-phosphate (MEP) and the pathway also resides
296 in the chromerid plastid. The alternative cytosolic (mevalonate) pathway for terpenoid
297 biosynthesis is missing, but terpenoid diphosphates are apparently exported from the
298 plastid for further elongation by cytosolic and mitochondrial polyprenyl diphosphate
299 synthases to produce, among others, ubiquinone for the mitochondrial respiratory chain
300 (Supplementary Table S1, Imlay & Odom 2014).

301 Terpenoid derivatives include carotenoids and terpenoid quinones. Carotenoids
302 violaxanthin and (iso)fucoxanthin are accessory photosynthetic pigments involved in
303 light-harvesting, and the de-epoxidation of violaxanthin was suggested to be
304 fundamental for photoprotection in *C. velia* (Kotabová et al. 2011). Plastoquinol is an
305 electron transport chain electron mediator, while tocopherol has a protective role in
306 oxidative stress (Müller & Kappes 2007). The plastid localization of these pathways
307 therefore conditions the photosynthetic abilities of chromerids.

Fatty acid and lipid synthesis are important for the generation of plastid membranes and the modulation of their physicochemical properties. Chromerid plastids house the type II fatty acid synthesis pathway that is fed with acetyl-CoA by the plastid pyruvate dehydrogenase complex (Foth et al. 2005). The produced acyl-ACP molecules may then be directly esterified with glycerol 3-phosphate to become intermediates of glycerolipid synthesis. Galactolipids of the mono- and digalactosyl-diacylglycerolipid groups (MGDG and DGDG) are major lipids of plastids in *C. velia* (Botté et al. 2011), probably required for proper embedding of photosystems into the thylakoid membranes but possibly also involved in plastid protein translocation (Schleiff et al. 2003). Their biosynthesis is limited to plastids in both *C. velia* and *V. brassicaformis*, as the respective protein contain plastid-targeting presequences (Fig. 1).

Flavin cofactors are critical for multiple enzymes of the TCA cycle, fatty acid oxidation, photosynthesis, respiratory chain, and *de novo* pyrimidine biosynthesis. In chromerids, riboflavin is synthesized in the plastid from a Calvin cycle intermediate, ribulose 5-phosphate, and flavin nucleotides are then supposedly exported from the plastid (Fig. 2).

Iron-sulfur clusters are components of redox proteins and in plastids, they function as the prosthetic group of the cytochrome b_6f and the ferredoxin redox relay that signals the plastid redox state to downstream enzymes. Plastid iron-sulfur clusters are typically synthesized by the SUF system, using ferrous ions and cysteine-derived sulfur as substrates. Both Cys synthesis and SUF system are found in chromerid plastids (Fig. 1), suggesting that plastid client proteins have regular access to Fe-S clusters.

Unexpectedly, only few enzymes of amino acid biosynthesis localize to plastids in *C. velia* and *V. brassicaformis*. We could reliably predict that only three enzymes of Cys synthesis reside in the plastid, but this amino acid is directly metabolized by the SUF

system to assemble Fe-S clusters. Strikingly, plant chloroplasts synthesize several amino acids (Lys, Arg, Ala, Trp, Tyr, Phe; Van Dingenen et al. 2016) and participate on nitrogen assimilation with the glutamine synthase / glutamine oxoglutarate aminotransferase (GS/GOGAT) cycle. Apparently, chromerid plastids are much less involved in nitrogen metabolism than previously studied systems (see below).

In summary, chromerid plastids are well-equipped for the biogenesis and maintenance of the photosynthetic machinery. In addition, these organelles prime fatty acid and terpenoid biosynthesis by the formation of short-chain intermediates (short-chain saturated fatty acids and monoterpenoid diphosphates), that are however exported for further processing. Heme is another crucial compound produced in this compartment. All these processes seem closely coupled to photosynthesis and, importantly, to carbon fixation by the Calvin cycle, which will be overviewed in the next section.

Compartment interplay

Calvin cycle is the very center of the second phase of photosynthesis; it utilizes ATP and NADPH produced by the light-dependent reactions to fix carbon dioxide into carbohydrates. We identified in chromerids all the enzymes of the cycle, discriminating them from the cytosolic carbohydrate metabolism variants by their N-terminal extensions having BTS characteristics. For three molecules of CO_2, the Calvin cycle produces one triose phosphate molecule (dihydroxyacetone phosphate – DHAP, or 3-phosphoglycerate) that can be metabolized by other pathways (see above), exported to the cytosol or reintroduced to the cycle to produce sugar phosphates with more carbons.

Triose phosphates enter the cytosol via the triose phosphate transporters, and those identified in *Plasmodium falciparum* prefer DHAP and phospho*enol*pyruvate (PEP)

as substrates, while also accepting 3-phosphoglycerate (Lim et al. 2010). Upon the entry of triose phosphates to the cytosol, they can take two major paths, depending on the metabolic state. Gluconeogenetic pathway leads to the accumulation of storage saccharides (amylopectin), while glycolysis yields pyruvate, a hub compound for both anabolic and catabolic reactions. Generally, favorable and illuminated growth of algae promote storage and anabolic pathways, while dark growth and starvation promote spending of sugar phosphates in the TCA cycle and respiration.

We found enzymes involved in all processes of polysaccharide accumulation and degradation (Coppin et al. 2005). Some of the storage sugar enzymes encode signal peptides or transmembrane domains, suggestive of their extracellular or membrane-associated activity. This is consistent with chromerids forming a thick cellulose cell-wall (Moore et al. 2008; Oborník et al. 2012; Füssy & Oborník 2017a).

The Calvin cycle is not only a source of triose phosphates for cytosolic glycolysis/gluconeogenesis, but also of pentose phosphates that are the starting substrates for nucleotide (ribose 5-phosphate), amino acid (ribose 5-phosphate and erythrose 4-phosphate), and vitamin synthesis (ribulose 5-phosphate). In this respect it complements the cytosolic pentose phosphate cycle that also provides these sugar phosphates. Based on our predictions, the plastid ribulose 5-phosphate pool serves as the substrate for flavin nucleotide synthesis, while cytosolic ribulose 5-phosphate is used for pyridoxin synthesis. Notably, we could not find any cytosolic 5-phosphoribosyl-1-pyrophosphate (PRPP) synthases that would provide PRPP for the synthesis of His, Trp and nucleotides in this compartment. Two pairs of orthologs are found in chromerids. Each of these sequences has a presequence, though only two orthologs have BTS (the other pair have mTP-like N-termini and were discarded from predictions due to uncertain localization; Supplementary Table S1). Comparison with PRPP synthase

383 sequences from apicomplexans revealed that some of those too have predicted
384 presequences (including both *Toxoplasma gondii* ME49 paralogs), but proteins having
385 these presequences do not cluster together (Supplementary Table S1). The localization
386 of PRPP synthases is therefore questionable and might be achieved by alternative
387 splicing.

388 Other anabolic reactions also stem from the cytosolic pool of sugar phosphates.
389 Biosynthesis of most amino acids in chromerids depends on glycolysis and pentose
390 phosphate cycle intermediates. Cys, Met and Ser derive from 3-phosphoglycerate, while
391 Phe, Trp and Tyr are synthesized from erythrose 4-phosphate via chorismate, and His is
392 a derivative of PRPP (but see above). Ala, Arg, Asp, Asn, Lys, Pro and Thr derive from 2-
393 oxoglutarate (2-OG) transaminated to form Glu by the GS/GOGAT cycle (Fig. 2). Notably,
394 the cytosolic 2-OG pool for GS/GOGAT cycle exists by the action of cytosolic copies of
395 TCA cycle enzymes (Fig. 2, Supplementary Table S1). Chromerid plastids do not
396 principally participate in the amino acid biosynthesis, with a few exceptions. A part of
397 the Arg pathway (ornithine synthesis) is predicted to localize in the plastid, which might
398 be a rate-limiting, regulatory measure on ornithine production via dependence on
399 stromal ATP levels. This is not unexpected, as in plants Arg synthesis is regulated at the
400 level of N-acetylglutamine phosphorylation (Ferrario-Méry et al. 2006). Notably,
401 chromerids utilize argJ (Supplementary Table S1), which is a glutamate transacetylating
402 enzyme that allows to recycle N-acetylglutamate after the production of ornithine. This
403 reaction does not produce free acetate and thus does not require additional ATP for
404 acetyl-CoA recycling. In *Plasmodium*, *Cryptosporidium* and *Eimeria*, Arg (via ornithine) is
405 used for polyamine synthesis (Cook et al. 2007) but this pathway shows cytosolic
406 localization congruently in apicomplexans and chromerids (Supplementary Table S1;
407 Shanmugasundram et al. 2013). Gly may be produced by glycine

hydroxymethyltransferase and threonine aldolase in mitochondria, cytosol, and plastids (Supplementary Table S1), reflecting its involvement in multiple pathways as a reaction intermediate (for plastid formylmethionine and mitochondrial ALA synthesis, for instance). There is some incongruence about the localization of Ile, Leu and Val synthesis among chromerids, as *V. brassicaformis* appears to localize Val and Ile biosynthesis to both mitochondria and plastid while Leu biosynthesis is cytosolic. In comparison, all these pathway steps are cytosolic in *C. velia*. Chromerid plastids thus appear to have minor roles in amino acid biosynthesis.

Both catabolic and anabolic reactions were found to take place in the chromerid mitochondria, though they are missing some common components. Both apicomplexans and chromerids ancestrally lack the respiratory complex I (and *C. velia* also lacks complex III) and utilize alternative NADH dehydrogenases that pass electrons to ubiquinone but do not contribute to the proton gradient (Fig. 2, Flegontov et al. 2015). Like Apicomplexa, chromerids also lack a canonical mitochondrial pyruvate dehydrogenase and instead take advantage of the promiscuous activity of the branched-chain amino acid dehydrogenase (BKCDH) (Foth et al. 2005; Danne et al. 2013; Oppenheim et al. 2014). The NAD^+-dependent isocitrate dehydrogenase is missing and replaced by an $NADP^+$-dependent isozyme, which could be linked to the loss of the canonical NADH dehydrogenase (respiratory complex I). $NADP^+$-dependent isocitrate dehydrogenase might in turn support the activity of NADPH-dependent enzymes (Danne et al. 2013). This is notable because in mammals both NAD^+ and $NADP^+$ isocitrate dehydrogenases are operational, the latter typically in reverse (reductive) direction (Sazanov & Jackson 1994; Yoo et al. 2008). In contrast with apicomplexans, both fumarate hydratase types are present in chromerids (apicomplexans express only type I; Bulusu et al. 2011), while an ortholog of the conserved apicomplexan malate:quinone

433 oxidoreductase is missing (Danne et al. 2013). Chromerids localize to mitochondria
434 several enzymes of amino acid decomposition (Supplementary Table S1), suggesting
435 that at least some catabolic pathways feed into the mitochondrial metabolism. A set of
436 lactate dehydrogenases in chromerids allow fermentation of pyruvate to lactate under
437 temporary dark anaerobic conditions (Flegontov et al. 2015; Oborník & Lukeš 2015) and
438 participate in methylglyoxal detoxication (Cordeiro et al. 2012). Furthermore, the
439 mitochondrial metabolism is equipped with malic enzyme (decarboxylating malate
440 dehydrogenase), which allows the regeneration of pyruvate for anabolic reactions when
441 the cycle is fed by fatty acid beta-oxidation. Therefore, mitochondria are well-integrated
442 in the chromerid carbon metabolism in both catabolic and anabolic directions.

443 Carbon metabolism also affects the synthesis rates of nitrogen pathways. The
444 GS/GOGAT cycle is responsible for nitrogen (ammonia) assimilation into amino acids in
445 most phototrophs. In chromerids, GOGAT is an NADH-dependent cytosolic enzyme, and
446 the restriction of amino acid pathways to the cytosol suggests that they are decoupled
447 from the redox state of the plastid and rather reflect the redox state of the cytosol
448 (Fig. 2). This also suggests that the plastids of chromerids are not as deeply involved in
449 primary metabolism as the plastids of primary algae and plants by lacking the ability to
450 synthesize amino acids and polysaccharides. Instead, plastid activity is crucial in lower
451 carbohydrate and fatty acid metabolism and appears to be sensed indirectly, through
452 the supply of triose phosphates. Under favorable conditions, the cytosol is fed with
453 photosynthetic sugar phosphates, and reducing agents and ATP are produced by
454 glycolysis. This promotes anabolic reactions that allow the accumulation of
455 polysaccharides and production of amino acids and lipids. In the dark or under nutrient
456 scarcity, the shortage of energy must be compensated by the reactions of the TCA cycle
457 and respiration in mitochondria. Unfortunately, there are no published large-scale

458 quantitative data suitable for tracing in more detail the metabolic flows through the
459 described pathways.
460
461
462 DISCUSSION
463 The physicochemical character of mitochondrial and plastid targeting
464 presequences remains quite similar across long evolutionary distances among
465 eukaryotes. Still, the predicting power of localization algorithms decreases with
466 divergence between the reference and the analyzed sets. Prediction algorithms perform
467 best if trained using lineage-specific datasets, usually based on available experimental
468 data (e.g. (Emanuelsson et al. 2007; Kaundal et al. 2013; Gruber et al. 2015).
469 Consistently, plastid-localization signals show specific variability among algal clades
470 (Patron & Waller 2007). Although many tools have been implemented to determine
471 mitochondrion- and plastid-localized proteins in *C. velia* and *V. brassicaformis* (Kořený et
472 al. 2011; Woehle et al. 2011; Petersen et al. 2014; Flegontov et al. 2015; Woo et al. 2015;
473 Sobotka et al. 2017), none of them have been tested on reference proteins in terms of
474 predictive power. To find a suitable tool to predict protein localization in chromerids,
475 we prepared a manually curated inventory of references that included proteins from
476 plastid, mitochondrion, and several other compartments, as negative controls. To date,
477 two works (Flegontov et al. 2015; Sobotka et al. 2017) have investigated the metabolism
478 of chromerids on organellar level, and only the latter work supports the localization of
479 analyzed (plastid) proteins with proteomic data. Our dataset included sequences of
480 typical plastid-targeted proteins as well as proteins with unambiguous localization to
481 mitochondria and other compartments, conserved in other eukaryotic lineages. To
482 ensure that our sequences are largely complete at their N-termini, we used protein

483 models generated by two independent sequencing initiatives, EuPathDB (deposited at
484 CryptoDB; (Woo et al. 2015) and MMETSP (Keeling et al. 2014).

485 We analyzed the performance of several algorithms based on their sensitivity
486 (percentage of true positive sequences passing a threshold) and precision (percentage
487 of false positives in the set passing the same threshold). For plastid proteomes, ASAFind
488 using a species-specific bit score matrix was found to be the most efficient for each
489 chromerid species. The sensitivity of ASAFind with *Phaeodactylum tricornutum*
490 sequences was comparable to our results (80%; Gruber et al. 2015). With mitochondrial
491 references, we could not observe predictive differences for *V. brassicaformis* sequences.
492 With *C. velia* datasets, HECTAR was more precise at higher sensitivity levels, and
493 MultiLoc2 could not reach an 85% precision (Table 1). TargetP performed similarly to
494 HECTAR (TargetP is indeed part of HECTAR's mitochondrial module) and is widely used
495 for finding mitochondria-targeted genes with sensitivity around 60-75% (depending on
496 the model and reliability class) in various organisms (Emanuelsson et al. 2007),
497 including the chromerids (Kořený et al. 2011; Flegontov et al. 2015; Woo et al. 2015).
498 This accuracy is also relatively lower because a portion of mitochondrial proteins use
499 alternative routes or signals for translocation to this organelle (Sun & Habermann
500 2017). Several of our plastid and mitochondrial reference sequences were recovered as
501 false negatives not passing the probability threshold. Indeed, 136 of the 1,141 reference
502 sequences had alternative open reading frames or were unannotated and had to be
503 manually adjusted using homology annotation, alignments and phylogenetic analyses
504 (Supplementary Table S1). For instance, we could not obtain consistent localizations for
505 the enzymes of the MEP pathway that is thought to be exclusively plastid-localized
506 (Fig. 1). With untreated data, the false negative discovery rate would be much higher,
507 leading to orphan enzymes predicted to unexpected compartments. This points out the

problems with automated analyses – there is an essential need for highly complete sequence data, which worsens large-scale predictions in the chromerids.

Although targeting presequences are generally not generally not conserved on sequence level, the conservation of amino acids flanking the BTS cleavage sites was found to be crucial for proper plastid targeting (Gruber et al. 2007). Proteins targeted to the rhodophyte-derived complex plastids generally expose an invariant Phe at their N-terminus after signal peptide cleavage (Gould et al. 2006; Patron & Waller 2007). Based on our frequency matrices, in *Vitrella* plastid-targeted proteins the Phe motif is not strictly conserved, while most plastid-targeted proteins of *Chromera* do possess this Phe (Fig. 3). This suggests there is some versatility of the translocation machineries in chromerids. Our observations are similar to the results of Woehle et al. (2011), although the frequency of Phe in their *C. velia* dataset appears higher. AT richness above 57 % was shown to correlate with a shift in amino acid composition of transit peptides towards AT-rich codons (Ralph et al. 2004) but the difference in the GC percentage of the plastid reference transcripts in chromerids appears unlikely to be the cause for a diminished Phe (53.0 % and 59.6 % GC in *C. velia* and *V. brassicaformis*, respectively). Lastly, we cannot exclude the possibility that mis-identified cleavage sites or mis-assembled transcripts in our data affected the amino acid frequencies. Nevertheless, the Phe motif is less strongly retained in *Toxoplasma* and apparently absent in *Plasmodium* apicoplast-targeted proteins (Patron & Waller 2007), consistent with our results. In addition, not all rhodophyte-derived lineages retain a high percentage of plastid-targeted proteins with the Phe; despite Phe occurs predominantly in cryptophytes (*Guillardia theta*) and heterokonts (*Thalassiosira pseudonana* and *Phaeodactylum tricornutum*), haptophytes apparently do not rely on Phe in their transit peptide presequences (Kilian & Kroth 2005; Patron & Waller 2007; Gruber et al. 2015). Phe

533 motif is also absent from the transit peptides of chlorophyte-derived algae (Patron &
534 Waller 2007).

535 The pathways that localize into plastids are typically associated with
536 photosynthesis. We identified enzymes responsible for the synthesis of tetrapyrroles,
537 terpenoids, carotenoids, lipids, iron-sulfur clusters and carbohydrates, hence
538 compounds required for the proper assembly and function of the photosynthetic
539 machinery. Our findings are consistent with previous biochemical analyses, showing
540 that chromerids have a limited set of photosynthetic pigments (Moore et al. 2008;
541 Kotabová et al. 2011) and that they exhibit unique structural changes to the
542 photosystems (Sobotka et al. 2017). We show that terpenoid and lipid pathways are
543 primed with substrates produced by the plastid carbohydrate metabolism (DHAP,
544 pyruvate, and PEP) and linked with the rate of carbon fixation by substrate availability
545 (Fig. 1). This is not unprecedented, as plant chloroplasts show a similar arrangement in
546 photosynthetically active and inactive plastids (e.g. Neuhaus and Emes 2000).

547 Nevertheless, plastid products might be essential beyond photosynthesis, which
548 is best illustrated by comparison of algae with relatives that lost photosynthetic abilities
549 (e.g., Hadariová et al. 2018). Indeed, the biology of non-photosynthetic plastids has been
550 best studied in apicomplexans to find suitable weak-spots of these infamous parasites.
551 The apicoplast produces fatty acids and terpenoids and participates in heme synthesis
552 to sustain parasite growth in hosts where salvage of these compounds is not possible
553 (reviewed by van Dooren and Striepen 2013). This dependence on a remnant plastid can
554 be regarded as an evolutionary constraint that cannot be easily overcome once parallel
555 pathways in the cytosol or mitochondria are lost (though losses of plastids occasionally
556 happen, see Füssy & Oborník 2018; Oborník 2018). Consistent with this view,
557 apicomplexans lack apicoplast-independent pathways for terpenoid (via mevalonate),

558 fatty acid (using type I fatty acid synthase, or FASI) and tetrapyrrole synthesis. Some
559 apicomplexans appear to possess FASI-like enzymes, though functional analyses suggest
560 that *Cryptosporidum* FASI is not involved in *de novo* fatty acid synthesis and rather
561 accepts long-chain fatty acyl thioesters as substrates for elongation (Zhu et al. 2004).
562 The importance of *Toxoplasma* FASI remains unclear, while other apicomplexan FASI-
563 like enzymes might be involved in polyketide synthesis (Kohli et al. 2016).

564 The photosynthetic relatives to apicomplexans, chromerids are also likely to lack
565 cytosolic pathways for fatty acid, terpenoid and tetrapyrrole synthesis and thus rely
566 entirely on the plastid synthesis. In addition, chromerid plastids synthesize FAD
567 cofactors and a nitrogen metabolism intermediate, ornithine (see below). Conversely,
568 chromerid plastids do not host any fatty acid elongases or desaturases, therefore short-
569 chain fatty acids need to be exported, processed and reimported for incorporation into
570 plastid lipids (Botté et al. 2011). This is analogous to the fatty acid synthesis architecture
571 in apicomplexans (Mazumdar & Striepen 2007). Similarly, chromerid tetrapyrrole
572 biosynthesis relies on the import of the starting substrate, ALA, from the cytosol
573 (Kořený et al. 2011). Therefore, chromerid plastids pathways are interdependent with
574 the cytosol, pointing out possible feedback regulatory mechanisms to limit their
575 biosynthetic activity.

576 In the dark, the reduced triose phosphate supply from the chromerid plastid must
577 be counterbalanced by catabolic reactions of the mitochondrial TCA cycle and
578 respiration. The chromerid TCA cycle and respiratory chain represent modifications to
579 the canonical mitochondrial pathways (see Results). The list of chromerid mitochondrial
580 enzymes is largely shared with parasitic apicomplexans (Flegontov et al. 2015; Jacot et
581 al. 2016). As yet, it is unclear what impact this arrangement has on the metabolism of
582 photoautotrophic organisms, but it stands out that through the TCA cycle chromerid

583 mitochondria are metabolically versatile and integrate catabolic pathways with enzymes
584 priming anabolic reactions and enzymes typically associated with anaerobiosis
585 (Flegontov et al. 2015). Intracellular stages of *Toxoplasma gondii* actively catabolize host
586 glucose via the oxidative TCA cycle to produce energy efficiently (MacRae et al. 2012). In
587 comparison, for asexual stages of *Plasmodium*, purine salvage from oxaloacetate is vital,
588 while the importance of the TCA cycle is diminished (Bulusu et al. 2011; Ke et al. 2015).
589 Similarly, cytosolic ATP citrate lyase contributes to acetyl-CoA production in *T. gondii*
590 (Tymoshenko et al. 2015). As such, chromerid mitochondria are likely to have major
591 influence on metabolic fluxes in the cell.

592 We observed that amino acid biosynthesis pathways consistently showed
593 cytosolic localizations, with minor exceptions. This is probably not due to mis-
594 annotation as we could not extend the sequences towards an alternative N-terminus in
595 most cases. Chromerid plastids thus appear to host only parts of Cys and Arg synthesis,
596 suggesting that algae do not necessarily synthesize amino acids in plastids as plants do
597 (Van Dingenen et al. 2016). While plastid Cys synthesis is required by the Fe-S cluster
598 assembly system SUF, the plastid production of ornithine for Arg biosynthesis might
599 have a regulatory role in chromerids. Indeed, nitrogen metabolism is energy-demanding
600 and subject to intense cross-talk with carbon metabolism, and plastids play a crucial role
601 in balancing the fluxes (Stitt 2002; Németh et al. 2018). As yet, the primarily cytosolic
602 localization of nitrogen metabolism in chromerids appears extraordinary among
603 phototrophs (Allen et al. 2011; Bromke 2013; de la Torre et al. 2013; Dorrell et al. 2017).
604 Of studied apicomplexans, *Toxoplasma* has the broadest capacity to produce amino acids
605 *de novo* or secondarily from specific precursors, being auxotrophic only for Arg, His and
606 Trp (Chaudhary et al. 2014; Tymoshenko et al. 2015). Although the middle steps of Pro
607 and Lys synthesis are currently unknown (Shanmugasundram et al. 2013), none of these

608 amino acid synthesis pathways appears to be placed exclusively in the apicoplast, in line
609 with our results. Arg scavenging from outer sources might have therefore triggered the
610 loss of the ornithine cycle in the ancestor of apicomplexans. Reliance on amino acid
611 import from the host resulted in even greater reduction of amino acid synthesis
612 capabilities. Thus, *Plasmodium* synthesizes six amino acids (Gly, Glu, Gln, Pro, Asp, Asn),
613 while *Cryptosporidium,* which entirely lacks an apicoplast (Keithly et al. 2000;
614 Chaudhary et al. 2014), synthesizes only Gly, Glu and Pro. It is tempting to speculate that
615 the major role of the cytosol in amino acid synthesis facilitated the plastid loss in a
616 distantly related non-photosynthetic dinoflagellate, *Hematodinium* (Gornik et al. 2015).
617 In apicomplexans, indeed, the main source of energy is glycolysis of host-drawn
618 glucose, but the parasites appear to use the same carbohydrate subpathways as
619 chromerids. These include the compartment exchange of triose phosphates via
620 phosphate transporters (Lim et al. 2010), parallel cytosolic and plastid glycolysis to
621 pyruvate (Fleige et al. 2007), cytosolic accumulation of storage polysaccharides (Coppin
622 et al. 2005) and a notable relocation of the canonical pyruvate dehydrogenase in the
623 apicoplast to feed the fatty acid synthesis (Foth et al. 2005; Fleige et al. 2007). There are
624 also obvious commonalities in the arrangement of mitochondrial metabolism, including
625 the modifications of the canonical TCA cycle, the respiratory chain, and the anabolic
626 reactions of mitochondria. As a previous comparative analysis of mitochondrial
627 metabolism of apicomplexans and dinoflagellates showed, this arrangement is largely
628 shared among all myzozoans, and likely represents an ancestral state employed by the
629 dinoflagellate-apicomplexan progenitor (Danne et al. 2013).
630 The overall architecture of biosynthetic pathways in chromerids suggests that
631 their cytosol represents the compartment which integrates the cellular energy status.
632 Through metabolite flow it can directly regulate anabolic and catabolic reactions based

633 on photosynthate supply from the plastid and nutrient availability. Hypothetically, such
634 a central role for the cytosol could have been employed by a free-living unicellular
635 eukaryovore/algivore that ingests its prey into a food vacuole and gradually digests it,
636 much like the one we picture was the ancestor of Myzozoa (Tikhonenkov et al. 2014).
637 Indeed, also dinoflagellates employ a remarkable spectrum of trophic strategies
638 (reviewed in Waller & Kořený 2017). Despite this diversity, there are commonalities in
639 the arrangement of mitochondrial and plastid pathways among myzozoans (Danne et al.
640 2013; Gornik et al. 2015; Waller et al. 2016), pointing out that the cytosolic pyruvate
641 hub may be persisting through plastid endosymbiotic events.

642 The apparently ancestral potential to exploit external resources and the richness
643 of apicomplexan cell-surface transporters raises the question whether chromerids also
644 take considerable advantage of extracellular nutrients. Both chromerids were isolated as
645 coral-associated organisms (Moore et al. 2008; Oborník et al. 2012) and intriguing data
646 have been presented on mixotrophy of *C. velia* and its association with corals (Cumbo et
647 al. 2013; Foster et al. 2014; Mohamed et al. 2018). This question therefore needs to be
648 addressed in more detail, as it might present implications on the early evolution of
649 Apicomplexa. However, experimental data are currently unavailable to present an in-
650 depth metabolomic model for chromerids and we can merely compare our results with
651 those from other alveolates. Here we present a benchmark set of biochemical pathways
652 that can be investigated by quantitative or phylogenetic approaches for a deeper
653 understanding of one of the "alveolate ways" of trophic transition from photosynthetic
654 algae to obligatory parasites.

655

656

657 CONCLUSIONS

Chromerids occupy a pivotal position in the tree of alveolates and hold the key to our understanding of transition to parasitism; they are free-living phototrophic algae with relatively canonical chromosomes and branch sister to the apicomplexans. To sketch a few more pathways on the metabolic map of chromerids, we prepared reference datasets of plastid, cytosolic and mitochondrial enzymes. We have manually curated these protein sequences so that most of them are complete, which is crucial for correct predictions of their subcellular localization. We unveiled that chromerid plastids vividly exchange compounds with the cytosol and mitochondria in order to produce terpenoids, lipids and tetrapyrroles. In contrast, chromerid plastids appear to have only a minor role in amino acid synthesis, as most of these pathways reside in the cytosol. Uniquely, chromerids were found to use a plastid ornithine cycle combined with a cytosolic Arg cycle for synthesis and decomposition of this amino acid. We outlined a major hub represented by lower glycolysis and gluconeogenesis enzymes that appears to regulate carbon and nitrogen metabolite flow depending on the photoactivity of the plastid. When compared to apicomplexan metabolic pathways, our model confirms a conserved architecture of carbon and nitrogen metabolism in these groups. Further analyses are though needed to gain insight into the regulation of these pathways in response to various cues. Using the suggested prediction tools, it is now also possible to introduce more enzymatic steps to the picture.

ABBREVIATIONS: 2-OG – 2-oxoglutarate, ACP – acyl-carrier protein, ALA – delta-aminolevulinic acid, BTS – bipartite (plastid) targeting signal, Chl *a* – chlorophyll *a*, DGDG/MGDG – di-/monogalactosyl-diacylglycerolipid, DHAP – dihydroxyacetone phosphate, Ery-4P – erythrose 4-phosphate, FAD – flavin adenine dinucleotide,

683 GS/GOGAT – glutamine synthase / glutamine oxoglutarate aminotransferase, IMS –
684 inter-membrane space, MEP – methyl-D-erythritol 4-phosphate, mTP – mitochondrial
685 transit peptide, OAA – oxaloacetate, orn – ornithine, PEP – phospho*enol*pyruvate, PQ –
686 plastoquinol, PRPP – 5-phosphoribosyl-1-pyrophosphate, pyr – pyruvate, Ru-5P –
687 ribulose 5-phosphate, TCA – tricarboxylic acid, THF - tetrahydrofolate, UQ - ubiquinone.
688
689 ETHICS APPROVAL AND CONSENT TO PARTICIPATE
690 Not applicable.
691
692 CONSENT FOR PUBLICATION
693 Not applicable.
694
695 COMPETING INTEREST
696 The authors declare that they have no competing interests.
697
698 AUTHORS' CONTRIBUTIONS
699 MO and ZF conceived the study. ZF and TF performed all the bioinformatic
700 analyses and prepared the figures and tables. ZF drafted the manuscript. MO acquired
701 funding. All authors edited the manuscript and approved its final form.
702
703
704 ACKNOWLEDGEMENTS
705 The authors would like to thank the Czech Science Foundation (project 16-
706 24027S granted to MO) and the European Regional Development Fund (ERDF) (No.
707 CZ.02.1.01/0.0/0.0/16_019/0000759) for funding. Computational resources from

MetaCentrum and CERIT-SC, Brno, Czech Republic, are greatly appreciated. We are grateful to Ansgar Gruber for helpful discussions.

REFERENCES

Allen AE et al. 2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. Nature. 473:203–207. doi: 10.1038/nature10074.

Atteia A et al. 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the α-proteobacterial mitochondrial ancestor. Mol. Biol. Evol. 26:1533–1548. doi: 10.1093/molbev/msp068.

Bisanz C et al. 2006. *Toxoplasma gondii* acyl-lipid metabolism: *de novo* synthesis from apicoplast-generated fatty acids versus scavenging of host cell precursors. Biochem. J. 394:197–205. doi: 10.1042/BJ20050609.

Blum T, Briesemeister S, Kohlbacher O. 2009. MultiLoc2: Integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. BMC Bioinformatics. 10:274. doi: 10.1186/1471-2105-10-274.

Botté CY et al. 2011. Identification of plant-like galactolipids in *Chromera velia*, a photosynthetic relative of malaria parasites. J. Biol. Chem. 286:29893–29903. doi: 10.1074/jbc.M111.254979.

Boucher MJ et al. 2018. Integrative proteomics and bioinformatic prediction enable a high-confidence apicoplast proteome in malaria parasites. PLoS Biol. 16:1–29. doi: 10.1371/journal.pbio.2005895.

Bouvier F, Harlingue AD, Backhaus RA, Kumagai MH, Camara B. 2000. Identification of neoxanthin synthase as a carotenoid cyclase paralog. Eur J Biochem. 267:6346–6352.

papers2://publication/uuid/4B30A843-FB08-4B02-AD19-FC30B447D8E4.

Bromke MA. 2013. Amino acid biosynthesis pathways in diatoms. Metabolites. 3:294–311. doi: 10.3390/metabo3020294.

Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. Nat. Methods. 12:59. doi: 10.1038/nmeth.3176.

Bulusu V, Jayaraman V, Balaram H. 2011. Metabolic fate of fumarate, a side product of the purine salvage pathway in the intraerythrocytic stages of *Plasmodium falciparum*. J. Biol. Chem. 286:9236–9245. doi: 10.1074/jbc.M110.173328.

Burki F et al. 2012. Re-evaluating the green versus red signal in eukaryotes with secondary plastid of red algal origin. Genome Biol. Evol. 4:626–635. doi: 10.1093/gbe/evs049.

Calvo SE, Mootha VK. 2010. The mitochondrial proteome and human disease. Annu. Rev. Genomics Hum. Genet. 11:25–44. doi: 10.1146/annurev-genom-082509-141720.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973. doi: 10.1093/bioinformatics/btp348.

Chaudhary K, Fox BA, Bzik DJ. 2014. Comparative aspects of nucleotide and amino acid metabolism in *Toxoplasma gondii* and other Apicomplexa. In: Toxoplasma gondii (2nd Edition). Weiss, LM & Kim, K, editors. Elsevier pp. 663–706. doi: 10.1016/B978-0-12-396481-6.00020-9.

Cohen L, Alexander H, Brown CT. 2016. Marine Microbial Eukaryotic Transcriptome Sequencing Project, re-assemblies. figshare. 3840153.

Cook T et al. 2007. Divergent polyamine metabolism in the Apicomplexa. Microbiology. 153:1123–1130. doi: 10.1099/mic.0.2006/001768-0.

Coppin A et al. 2005. Evolution of plant-like crystalline storage polysaccharide in the

protozoan parasite *Toxoplasma gondii* argues for a red alga ancestry. J. Mol. Evol. 60:257–267. doi: 10.1007/s00239-004-0185-6.

Cordeiro C et al. 2012. The glyoxalase pathway in protozoan parasites. Int. J. Med. Microbiol. 302:225–229. doi: 10.1016/j.ijmm.2012.07.005.

Crooks G, Hon G, Chandonia J, Brenner S. 2004. WebLogo: A sequence logo generator. Genome Res. 14:1188–1190. doi: 10.1101/gr.849004.1.

Cumbo VR et al. 2013. *Chromera velia* is endosymbiotic in larvae of the reef corals *Acropora digitifera* and *A. tenuis*. Protist. 164:237–244. doi: 10.1016/j.protis.2012.08.003.

Danne JC, Gornik SG, MacRae JI, McConville MJ, Waller RF. 2013. Alveolate mitochondrial metabolic evolution: Dinoflagellates force reassessment of the role of parasitism as a driver of change in apicomplexans. Mol. Biol. Evol. 30:123–139. doi: 10.1093/molbev/mss205.

Van Dingenen J, Blomme J, Gonzalez N, Inzé D. 2016. Plants grow with a little help from their organelle friends. J. Exp. Bot. 67:6267–6281. doi: 10.1093/jxb/erw399.

van Dooren GG, Striepen B. 2013. The algal past and parasite present of the apicoplast. Annu. Rev. Microbiol. 67:271–289. doi: 10.1146/annurev-micro-092412-155741.

Dorrell RG et al. 2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. eLife. 6:e23717. doi: 10.7554/eLife.23717.

Dorrell RG, Butterfield ER, Nisbet RER, Howe CJ. 2013. Evolution: Unveiling early alveolates. Curr. Biol. 23:R1093–R1096. doi: 10.1016/j.cub.2013.10.055.

Dorrell RG, Drew J, Nisbet RER, Howe CJ. 2014. Evolution of chloroplast transcript processing in *Plasmodium* and its chromerid algal relatives Dutcher, SK, editor. PLoS Genet. 10:e1004008. doi: 10.1371/journal.pgen.1004008.

Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell

783 using TargetP, SignalP and related tools. Nat. Protoc. 2:953–971. doi:
784 10.1038/nprot.2007.131.
785 Emanuelsson O, Nielsen H, Brunak S, Von Heijne G. 2000. Predicting subcellular
786 localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol.
787 300:1005–1016. doi: 10.1006/jmbi.2000.3903.
788 Ferrario-Méry S, Besin E, Pichon O, Meyer C, Hodges M. 2006. The regulatory PII protein
789 controls arginine biosynthesis in *Arabidopsis*. FEBS Lett. 580:2015–2020. doi:
790 10.1016/j.febslet.2006.02.075.
791 Flegontov P et al. 2015. Divergent mitochondrial respiratory chains in phototrophic
792 relatives of apicomplexan parasites. Mol. Biol. Evol. 32:1115–1131. doi:
793 10.1093/molbev/msv021.
794 Fleige T, Fischer K, Ferguson DJP, Gross U, Bohne W. 2007. Carbohydrate metabolism in
795 the *Toxoplasma gondii* apicoplast: Localization of three glycolytic isoenzymes, the
796 single pyruvate dehydrogenase complex, and a plastid phosphate translocator.
797 Eukaryot. Cell. 6:984–996. doi: 10.1128/EC.00061-07.
798 Foster C, Portman N, Chen M, Šlapeta J. 2014. Increased growth and pigment content of
799 *Chromera velia* in mixotrophic culture. FEMS Microbiol. Ecol. 88:121–128. doi:
800 10.1111/1574-6941.12275.
801 Foth BJ et al. 2005. The malaria parasite *Plasmodium falciparum* has only one pyruvate
802 dehydrogenase complex, which is located in the apicoplast. Mol. Microbiol. 55:39–53.
803 doi: 10.1111/j.1365-2958.2004.04407.x.
804 Füssy Z, Oborník M. 2017a. Chromerids and their plastids. In: Advances in Botanical
805 Research. Hirakawa, Y, editor. Vol. 84 pp. 187–218. doi: 10.1016/bs.abr.2017.07.001.
806 Füssy Z, Oborník M. 2018. Complex endosymbioses I: From primary to complex plastids,
807 multiple independent events. Methods Mol. Biol. 1829:17–35. doi: 10.1007/978-1-

4939-8654-5_2.

Füssy Z, Oborník M. 2017b. Reductive evolution of apicomplexan parasites from phototrophic ancestors. In: Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts. Pontarotti, P, editor. Springer International Publishing: Cham pp. 217–236. doi: 10.1007/978-3-319-61569-1_12.

Garg SG, Gould SB. 2016. The role of charge in protein targeting evolution. Trends Cell Biol. 26:894–905. doi: 10.1016/j.tcb.2016.07.001.

Gawryluk RMR, Chisholm KA, Pinto DM, Gray MW. 2014. Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J. Proteomics. 109:400–416. doi: 10.1016/j.jprot.2014.07.005.

Gonczarowska-Jorge H, Zahedi RP, Sickmann A. 2017. The proteome of baker's yeast mitochondria. Mitochondrion. 33:15–21. doi: 10.1016/j.mito.2016.08.007.

Gornik SG et al. 2015. Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. Proc. Natl. Acad. Sci. U. S. A. 112:5767–72. doi: 10.1073/pnas.1423400112.

Gould SB et al. 2006. Protein targeting into the complex plastid of cryptophytes. J. Mol. Evol. 62:674–681. doi: 10.1007/s00239-005-0099-y.

Gray MW. 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. Proc. Natl. Acad. Sci. U. S. A. 112:10133–8. doi: 10.1073/pnas.1421379112.

Gruber A et al. 2007. Protein targeting into complex diatom plastids: Functional characterisation of a specific targeting motif. Plant Mol. Biol. 64:519–530. doi: 10.1007/s11103-007-9171-x.

Gruber A. 2019. What's in a name? How organelles of endosymbiotic origin can be

833 distinguished from endosymbionts. Microb. Cell. 6:123–133. doi:
834 10.15698/mic2019.02.668.
835 Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. 2015. Plastid proteome prediction
836 for diatoms and other algae with secondary plastids of the red lineage. Plant J.
837 81:519–528. doi: 10.1111/tpj.12734.
838 Gschloessl B, Guermeur Y, Cock JM. 2008. HECTAR: A method to predict subcellular
839 targeting in heterokonts. BMC Bioinformatics. 9:393. doi: 10.1186/1471-2105-9-393.
840 Gupta A et al. 2014. Reduced ribosomes of the apicoplast and mitochondrion of
841 *Plasmodium* spp. and predicted interactions with antibiotics. Open Biol. 4:140045.
842 doi: 10.1098/rsob.140045.
843 Hadariová L, Vesteg M, Hampl V, Krajčovič J. 2018. Reductive evolution of chloroplasts in
844 non-photosynthetic plants, algae and protists. Curr. Genet. 64:365–387. doi:
845 10.1007/s00294-017-0761-0.
846 Hiller K, Grote A, Scheer M, Münch R, Jahn D. 2004. PrediSi: Prediction of signal peptides
847 and their cleavage positions. Nucleic Acids Res. 32:375–379. doi:
848 10.1093/nar/gkh378.
849 Hopkins JF et al. 2012. Proteomics reveals plastid- and periplastid-targeted proteins in
850 the chlorarachniophyte alga *Bigelowiella natans*. Genome Biol. Evol. 4:1391–1406.
851 doi: 10.1093/gbe/evs115.
852 Huang S, Shingaki-Wells RN, Taylor NL, Millar AH. 2013. The rice mitochondria
853 proteome and its response during development and to the environment. Front. Plant
854 Sci. 4:16. doi: 10.3389/fpls.2013.00016.
855 Imlay L, Odom AR. 2014. Isoprenoid metabolism in apicomplexan parasites. Curr. Clin.
856 Microbiol. Reports. 1:37–50. doi: 10.1007/s40588-014-0006-7.
857 Jacot D, Waller RF, Soldati-Favre D, MacPherson DA, MacRae JI. 2016. Apicomplexan

858 energy metabolism: Carbon source promiscuity and the quiescence hyperbole.
859 Trends Parasitol. 32:56–70. doi: 10.1016/j.pt.2015.09.001.
860 Janouškovec J et al. 2013. Colponemids represent multiple ancient alveolate lineages.
861 Curr. Biol. 23:2546–2552. doi: 10.1016/j.cub.2013.10.062.
862 Janouškovec J et al. 2015. Factors mediating plastid dependency and the origins of
863 parasitism in apicomplexans and their close relatives. Proc. Natl. Acad. Sci.
864 112:10200–10207. doi: 10.1073/pnas.1423790112.
865 Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin
866 of the apicomplexan, dinoflagellate, and heterokont plastids. Proc. Natl. Acad. Sci.
867 107:10949–10954. doi: 10.1073/pnas.1003335107.
868 Jones P et al. 2014. InterProScan 5: genome-scale protein function classification.
869 Bioinformatics. 30:1236–1240. doi: 10.1093/bioinformatics/btu031.
870 Jouhet J, Maréchal E, Block MA. 2007. Glycerolipid transfer for the building of
871 membranes in plant cells. Prog. Lipid Res. 46:37–55. doi:
872 10.1016/j.plipres.2006.06.002.
873 Joy JB, Liang RH, McCloskey RM, Nguyen T, Poon AFY. 2016. Ancestral reconstruction.
874 PLoS Comput. Biol. 12:e1004763. doi: 10.1371/journal.pcbi.1004763.
875 Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new
876 perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 45:D353–
877 D361. doi: 10.1093/nar/gkw1092.
878 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
879 Improvements in performance and usability. Mol. Biol. Evol. 30:772–780. doi:
880 10.1093/molbev/mst010.
881 Kaundal R, Sahu SS, Verma R, Weirick T. 2013. Identification and characterization of
882 plastid-type proteins from sequence-attributed features using machine learning. BMC

883 Bioinformatics. 14:S7. doi: 10.1186/1471-2105-14-S14-S7.

884 Ke H et al. 2015. Genetic investigation of tricarboxylic acid metabolism during the

885 *Plasmodium falciparum* life cycle. Cell Rep. 11:164–174. doi:

886 10.1016/j.celrep.2015.03.011.

887 Kearse M et al. 2012. Geneious Basic: An integrated and extendable desktop software

888 platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–

889 1649. doi: 10.1093/bioinformatics/bts199.

890 Keeling PJ et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing

891 Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the

892 oceans through transcriptome sequencing. PLoS Biol. 12:e1001889. doi:

893 10.1371/journal.pbio.1001889.

894 Keithly JS, Marchewka MJ, Zhu G, Marchewka MJ, Keithly JS. 2000. *Cryptosporidium*

895 *parvum* appears to lack a plastid genome. Microbiology. 146:315–321. doi:

896 10.1055/s-2001-18357.

897 Kersey PJ et al. 2016. Ensembl Genomes 2016: more genomes, more complexity. Nucleic

898 Acids Res. 44:D574–D580. doi: 10.1093/nar/gkv1209.

899 Kilian O, Kroth PG. 2005. Identification and characterization of a new conserved motif

900 within the presequence of proteins targeted into complex diatom plastids. Plant J.

901 41:175–183. doi: 10.1111/j.1365-313X.2004.02294.x.

902 Kohli GS, John U, Van Dolah FM, Murray SA. 2016. Evolutionary distinctiveness of fatty

903 acid and polyketide synthesis in eukaryotes. ISME J. 10:1877–1890. doi:

904 10.1038/ismej.2015.263.

905 Kořený L, Sobotka R, Janouškovec J, Keeling PJ, Oborník M. 2011. Tetrapyrrole synthesis

906 of photosynthetic chromerids is likely homologous to the unusual pathway of

907 apicomplexan parasites. Plant Cell. 23:3454–3462. doi: 10.1105/tpc.111.089102.

908 Kotabová E, Kaňa R, Jarešová J, Prášil O. 2011. Non-photochemical fluorescence
909 quenching in *Chromera velia* is enabled by fast violaxanthin de-epoxidation. FEBS
910 Lett. 585:1941–1945. doi: 10.1016/j.febslet.2011.05.015.
911 Kunze M, Berger J. 2015. The similarity between N-terminal targeting signals for protein
912 import into different organelles and its evolutionary relevance. Front. Physiol. 6:1–27.
913 doi: 10.3389/fphys.2015.00259.
914 de la Torre F, El-Azaz J, Ávila C, Cánovas FM. 2013. Deciphering the role of aspartate and
915 prephenate aminotransferase activities in plastid nitrogen metabolism. Plant Physiol.
916 164:92–104. doi: 10.1104/pp.113.232462.
917 Lim L, Linka M, Mullin KA, Weber APM, Mcfadden GI. 2010. The carbon and energy
918 sources of the non-photosynthetic plastid in the malaria parasite. FEBS Lett.
919 584:549–554. doi: 10.1016/j.febslet.2009.11.097.
920 MacRae JI et al. 2012. Mitochondrial metabolism of glucose and glutamine is required for
921 intracellular growth of *Toxoplasma gondii*. Cell Host Microbe. 12:682–692. doi:
922 10.1016/j.chom.2012.09.013.
923 Mallo N, Fellows J, Johnson C, Sheiner L. 2018. Protein import into the endosymbiotic
924 organelles of apicomplexan parasites. Genes (Basel). 9:412. doi:
925 10.3390/genes9080412.
926 Mazumdar J, Striepen B. 2007. Make it or take it: Fatty acid metabolism of apicomplexan
927 parasites. Eukaryot. Cell. 6:1727–1735. doi: 10.1128/ec.00255-07.
928 Mikami K, Hosokawa M. 2013. Biosynthetic pathway and health benefits of fucoxanthin,
929 an algae-specific xanthophyll in brown seaweeds. Int. J. Mol. Sci. 14:13763–13781.
930 doi: 10.3390/ijms140713763.
931 Mohamed AR et al. 2018. Deciphering the nature of the coral-*Chromera* association.
932 ISME J. 12:776–790. doi: 10.1038/s41396-017-0005-9.

933 Moore RB et al. 2008. A photosynthetic alveolate closely related to apicomplexan
934 parasites. Nature. 451:959–963. doi: 10.1038/nature06635.
935 Müller S, Kappes B. 2007. Vitamin and cofactor biosynthesis pathways in *Plasmodium*
936 and other apicomplexan parasites. Trends Parasitol. 23:112–121. doi:
937 10.1016/j.pt.2007.01.009.
938 Nash EA, Nisbet RER, Barbrook AC, Howe CJ. 2008. Dinoflagellates: a mitochondrial
939 genome all at sea. Trends Genet. 24:328–335. doi: 10.1016/j.tig.2008.04.001.
940 Németh E, Nagy Z, Pécsváradi A. 2018. Chloroplast glutamine synthetase, the key
941 regulator of nitrogen metabolism in wheat, performs its role by fine regulation of
942 enzyme activity via negative cooperativity of Its subunits. Front. Plant Sci. 9:191. doi:
943 10.3389/fpls.2018.00466.
944 Neuhaus HE, Emes MJ. 2000. Nonphotosynthetic metabolism in plastids. Annu. Rev.
945 Plant Physiol. Plant Mol. Biol. 51:111–140. doi: 10.1146/annurev.arplant.51.1.111.
946 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective
947 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol.
948 32:268–274. doi: 10.1093/molbev/msu300.
949 Oborník M. 2019. In the beginning was the word: How terminology drives our
950 understanding of endosymbiotic organelles. Microb. Cell. 6:134–141. doi:
951 10.15698/mic2019.02.669.
952 Oborník M et al. 2012. Morphology, ultrastructure and life cycle of *Vitrella brassicaformis*
953 n. sp., n. gen., a novel chromerid from the Great Barrier Reef. Protist. 163:306–323.
954 doi: 10.1016/j.protis.2011.09.001.
955 Oborník M. 2018. The birth of red complex plastids: One, three, or four times? Trends
956 Parasitol. 34:923–925. doi: 10.1016/j.pt.2018.09.001.
957 Oborník M, Janouškovec J, Chrudimský T, Lukeš J. 2009. Evolution of the apicoplast and

its hosts: From heterotrophy to autotrophy and back again. Int. J. Parasitol. 39:1–12. doi: 10.1016/j.ijpara.2008.07.010.

Oborník M, Lukeš J. 2015. The organellar genomes of *Chromera* and *Vitrella*, the phototrophic relatives of apicomplexan parasites. Annu. Rev. Microbiol. 69:129–144. doi: 10.1146/annurev-micro-091014-104449.

Oppenheim RD et al. 2014. BCKDH: The missing link in apicomplexan mitochondrial metabolism is required for full virulence of *Toxoplasma gondii* and *Plasmodium berghei*. PLoS Pathog. 10:e1004263. doi: 10.1371/journal.ppat.1004263.

Pagliarini DJ, Rutter J. 2013. Hallmarks of a new era in mitochondrial biochemistry. Genes Dev. 27:2615–2627. doi: 10.1101/gad.229724.113.

Palmfeldt J, Bross P. 2017. Proteomics of human mitochondria. Mitochondrion. 33:2–14. doi: 10.1016/j.mito.2016.07.006.

Panigrahi AK et al. 2009. A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. Proteomics. 9:434–450. doi: 10.1002/pmic.200800477.

Patron NJ, Waller RF. 2007. Transit peptide diversity and divergence: A global analysis of plastid targeting signal. BioEssays. 29:1048–1058. doi: 10.1002/bies.20638.

Petersen J et al. 2014. *Chromera velia*, endosymbioses and the rhodoplex hypothesis - Plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). Genome Biol. Evol. 6:666–684. doi: 10.1093/gbe/evu043.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods. 8:785–786. doi: 10.1038/nmeth.1701.

Petsalaki EI, Bagos PG, Litou ZI, Hamodrakas SJ. 2006. PredSL: A tool for the N-terminal sequence-based prediction of protein subcellular localization. Genomics, Proteomics Bioinforma. 4:48–55. doi: 10.1016/S1672-0229(06)60016-8.

983 Ralph SA, Foth BJ, Hall N, McFadden GI. 2004. Evolutionary pressures on apicoplast
984 transit peptides. Mol. Biol. Evol. 21:2183–2194. doi: 10.1093/molbev/msh233.
985 Sazanov LA, Jackson JB. 1994. Proton-translocating transhydrogenase and NAD- and
986 NADP-linked isocitrate dehydrogenases operate in a substrate cycle which
987 contributes to fine regulation of the tricarboxylic acid cycle activity in mitochondria.
988 FEBS Lett. 344:109–116. doi: 10.1016/0014-5793(94)00370-X.
989 Schleiff E, Soll J, Küchler M, Kühlbrandt W, Harrer R. 2003. Characterization of the
990 translocon of the outer envelope of chloroplasts. J. Cell Biol. 160:541–551. doi:
991 10.1083/jcb.200210060.
992 Searcy DG. 2003. Metabolic integration during the evolutionary origin of mitochondria.
993 Cell Res. 13:229–238. doi: 10.1038/sj.cr.7290168.
994 Shanmugasundram A, Gonzalez-Galarza FF, Wastling JM, Vasieva O, Jones AR. 2013.
995 Library of Apicomplexan Metabolic Pathways: A manually curated database for
996 metabolic pathways of apicomplexan parasites. Nucleic Acids Res. 41:706–713. doi:
997 10.1093/nar/gks1139.
998 Smith DGS et al. 2007. Exploring the mitochondrial proteome of the ciliate protozoon
999 *Tetrahymena thermophila*: Direct analysis by tandem mass spectrometry. J. Mol. Biol.
1000 374:837–863. doi: 10.1016/j.jmb.2007.09.051.
1001 Sobotka R et al. 2017. Extensive gain and loss of photosystem I subunits in chromerid
1002 algae, photosynthetic relatives of apicomplexans. Sci. Rep. 7:13214. doi:
1003 10.1038/s41598-017-13575-x.
1004 Stitt M. 2002. Steps towards an integrated view of nitrogen metabolism. J. Exp. Bot.
1005 53:959–970. doi: 10.1093/jexbot/53.370.959.
1006 Sun S, Habermann BH. 2017. A guide to computational methods for predicting
1007 mitochondrial localization. In: Mitochondria: Practical protocols. Mokranjac, D &

1008 Perocchi, F, editors. Springer New York: New York, NY pp. 1–14. doi: 10.1007/978-1-
1009 4939-6824-4_1.
1010 Terashima M, Specht M, Hippler M. 2011. The chloroplast proteome: A survey from the
1011 *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. Curr.
1012 Genet. 57:151–168. doi: 10.1007/s00294-011-0339-1.
1013 Tetlow IJ, Rawsthorne S, Raines C, Emes MJ. 2005. Plastid metabolic pathways. In:
1014 Annual Plant Reviews. Møller, SG, editor. Vol. 13 Blackwell Publishing Ltd pp. 60–125.
1015 doi: 10.1002/9781119312994.apr0123.
1016 Tikhonenkov D V. et al. 2014. Description of *Colponema vietnamica* sp.n. and
1017 *Acavomonas peruviana* n. gen. n. sp., two new alveolate phyla (Colponemidia nom.
1018 nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the
1019 ancestral state of alveolates and eukaryotes Gribaldo, S, editor. PLoS One. 9:e95467.
1020 doi: 10.1371/journal.pone.0095467.
1021 Tymoshenko S et al. 2015. Metabolic needs and capabilities of *Toxoplasma gondii*
1022 through combined computational and experimental analysis. PLoS Comput. Biol.
1023 11:e1004261. doi: 10.1371/journal.pcbi.1004261.
1024 Vazač J, Füssy Z, Hladová I, Killi S, Oborník M. 2017. Ploidy and number of chromosomes
1025 in the alveolate alga *Chromera velia*. Protist. 169:53–63. doi:
1026 10.1016/j.protis.2017.12.001.
1027 Waller RF, Gornik SG, Kořený L, Pain A. 2016. Metabolic pathway redundancy within the
1028 apicomplexan-dinoflagellate radiation argues against an ancient chromalveolate
1029 plastid. Commun. Integr. Biol. 9:e1116653. doi: 10.1080/19420889.2015.1116653.
1030 Waller RF, Kořený L. 2017. Plastid complexity in dinoflagellates: A picture of gains,
1031 losses, replacements and revisions. In: Advances in Botanical Research. Hirakawa, Y,
1032 editor. Vol. 84 pp. 105–143. doi: 10.1016/bs.abr.2017.07.004.

van Wijk KJ, Baginsky S. 2011. Plastid proteomics in higher plants: Current state and future goals. Plant Physiol. 155:1578–1588. doi: 10.1104/pp.111.172932.

Woehle C, Dagan T, Martin WF, Gould SB. 2011. Red and problematic green phylogenetic signals among thousands of nuclear genes from the photosynthetic and apicomplexa-related Chromera velia. Genome Biol. Evol. 3:1220–1230. doi: 10.1093/gbe/evr100.

Woo YH et al. 2015. Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. eLife. 4:e06974. doi: 10.7554/eLife.06974.

Yoo H, Antoniewicz MR, Stephanopoulos G, Kelleher JK. 2008. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. J. Biol. Chem. 283:20621–20627. doi: 10.1074/jbc.M706494200.

Zhu G et al. 2004. Expression and functional characterization of a giant Type I fatty acid synthase (CpFAS1) gene from *Cryptosporidium parvum*. Mol. Biochem. Parasitol. 134:127–135. doi: 10.1016/j.molbiopara.2003.11.011.

FIGURES:

Figure 1: Overview of reference pathways, focused on the inter-connected reactions of terpenoid, lipid and tetrapyrrole biosynthesis, mitochondrial sulfur assimilation and glycine cleavage, and ribosomal proteins. Pathways are color-coded and enzymes/proteins are numbered according to the Supplementary Table S1. Abbreviations: ACP – acyl-carrier protein, ALA – delta-aminolevulinic acid, Chl *a* – chlorophyll *a*, DGDG – digalactosyldiacylglycerolipids, DHAP – dihydroxyacetone phosphate, GGPP – geranylgeranyl pyrophosphate, IMS – inter-membrane space, MEP –

methyl-D-erythritol 4-phosphate, PEP – phospho*enol*pyruvate, pyr – pyruvate, SQDG – sulfoquinovosyldiacylglycerolipids, THF – tetrahydrofolate.

Figure 2: Overview of reference pathways, continued, focused on the inter-connected reactions of carbohydrate and amino acid biosynthesis, mitochondrial respiration and plastid photosynthesis. Pathways are color-coded and enzymes/proteins are numbered or coded according to the Supplementary Table S1. 2-OG – 2-oxoglutarate, DHAP – dihydroxyacetone phosphate, Ery-4P – erythrose 4-phosphate, FAD – flavin adenine dinucleotide, IMS – inter-membrane space, OAA – oxaloacetate, orn – ornithine, PEP – phospho*enol*pyruvate, PQ – plastoquinol, PRPP – 5-phosphoribosyl-1-pyrophosphate, pyr – pyruvate, Ru-5P – ribulose 5-phosphate, TCA – tricarboxylic acid, THF – tetrahydrofolate, UQ – ubiquinone.

Figure 3: Sequence motifs surrounding the signal cleavage site differ in chromerids. While *C. velia* retains a highly conserved Phe residue that follows after the cleavage site (marked by +1), in *V. brassicaformis* Phe seems less conserved. Logoplot created by WebLogo (Crooks et al. 2004).

Table 1: Performance of various localization predictors as employed in this study. ASAFind and HECTAR-plastid were used for plastid reference assessment. Note that ASAFind has a cumulative score, with our data reaching up to values of 6. ASAFind+ denotes ASAFind with species-specific bit score matrices designed with *C. velia* and *V. brassicaformis* plastid references. TargetP, HECTAR-mitochondrion and MultiLoc2 were used for mitochondrial predictions and two thresholds are presented for these

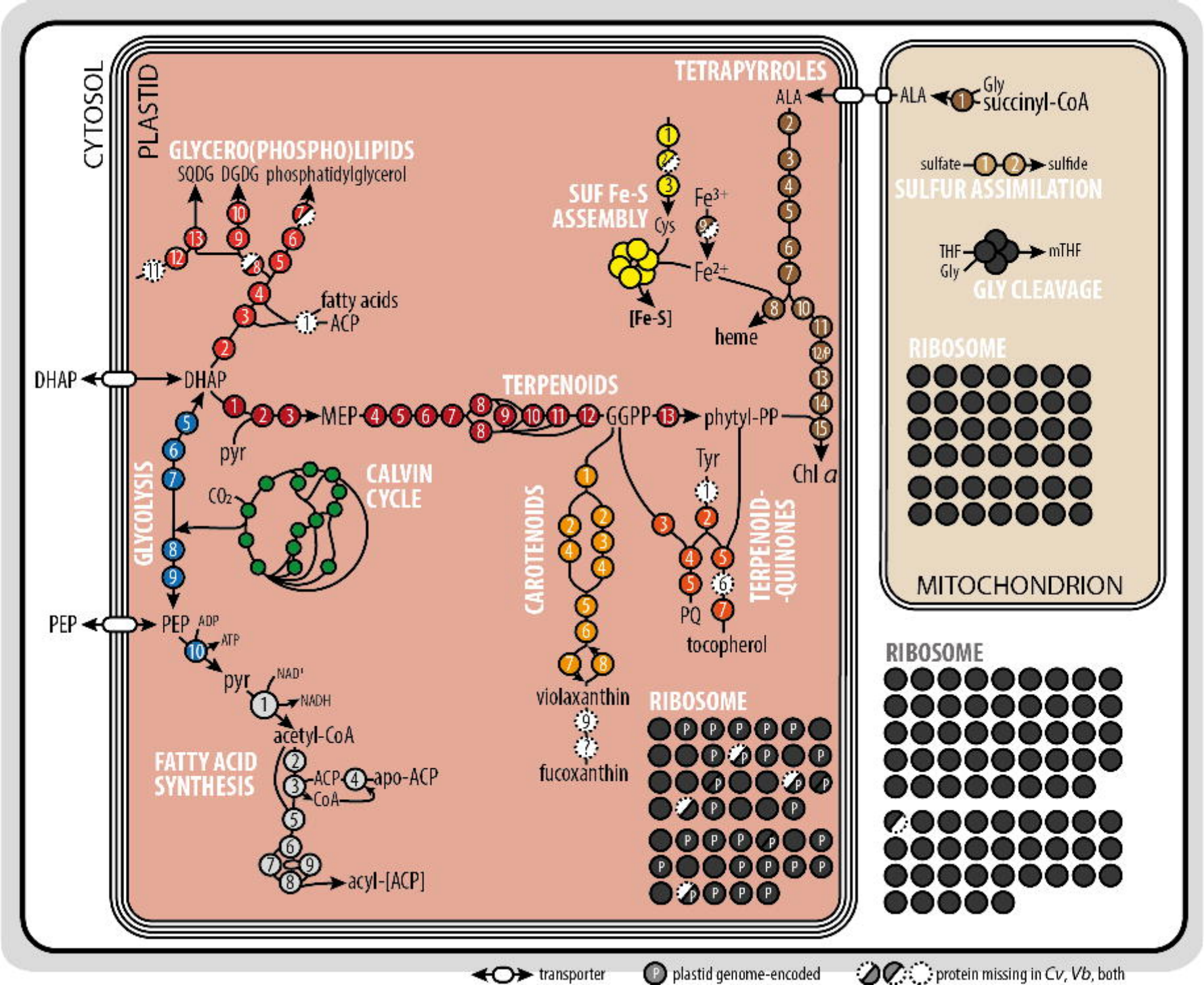
predictors, one aimed at higher sensitivity (s – 75%), the other aimed at high precision (p – 85%).

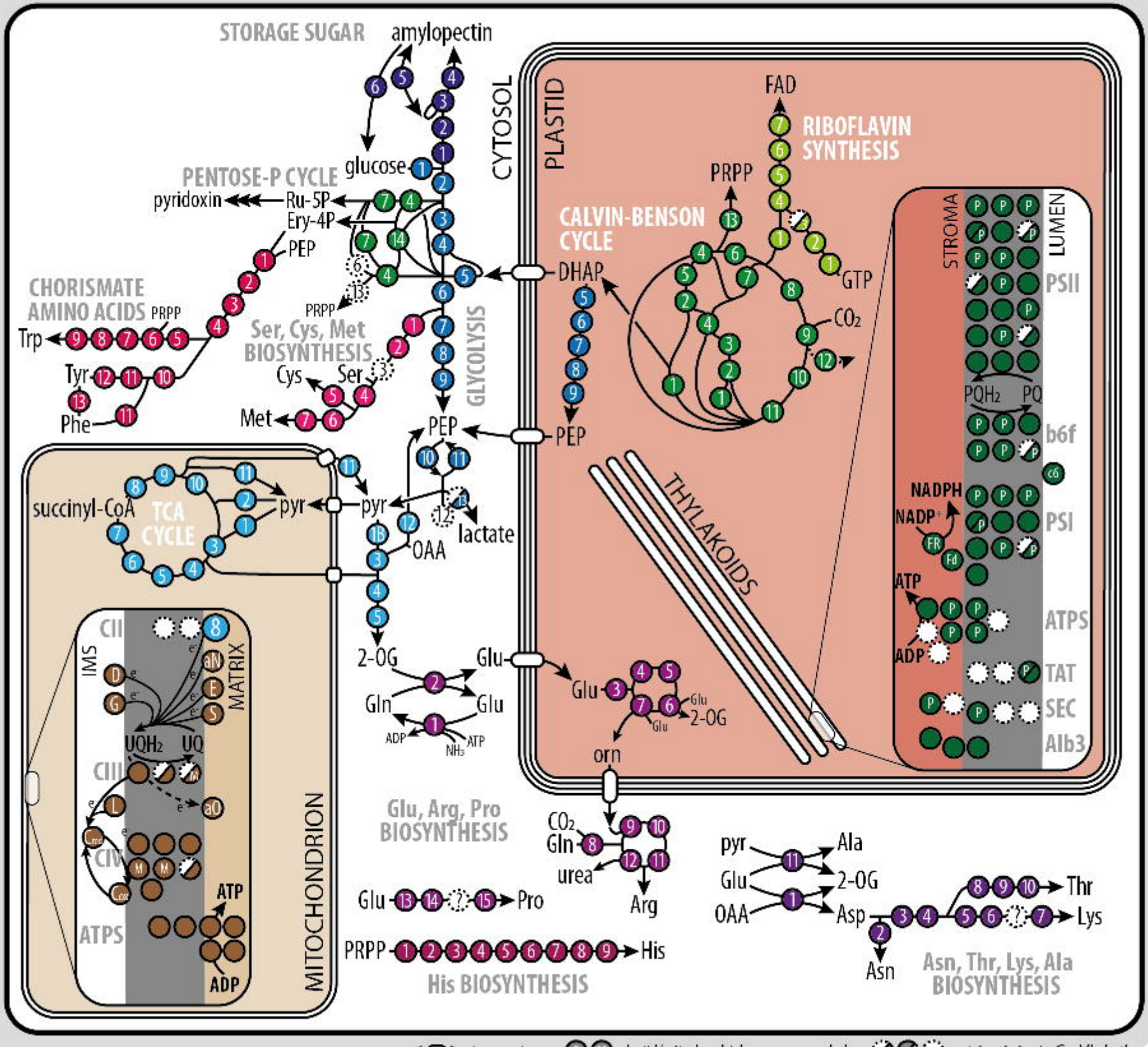
SUPPLEMENTARY MATERIAL:

Supplementary Table S1: Reference protein lists grouped by metabolic function; inferred phylogenies of argJ, polyprenyl-PP synthases, and PRPP synthase; set of updated reference protein sequences. Trees in Newick format, accessions are marked on leaves. The maximum-likelihood trees were inferred by the IQ-TREE software (see main text Methods). Sequence updates are based on homology searches within alternative transcriptomic datasets (elongated sequences) and comparison of orthologous sequences in *C. velia* and *V. brassicaformis* (merged contigs).

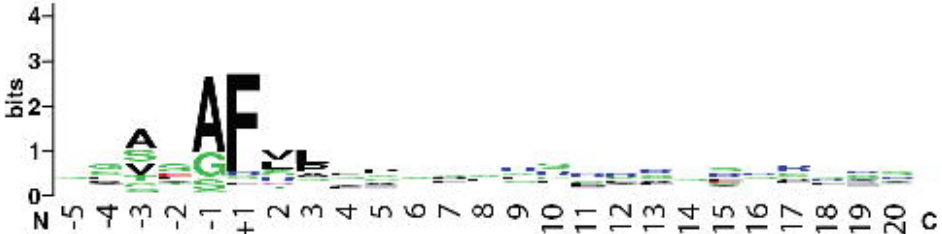
Supplementary Table S2: Prediction scores for reference proteins and the species-specific ASAFind scoring matrices of *C. velia* and *V. brassicaformis*.

Supplementary Figure S3: Targeting peptide score distributions for reference sequences. Reference colors: green – plastid, blue – mitochondrial, grey – other (cytosolic or non-plastid endomembrane route).

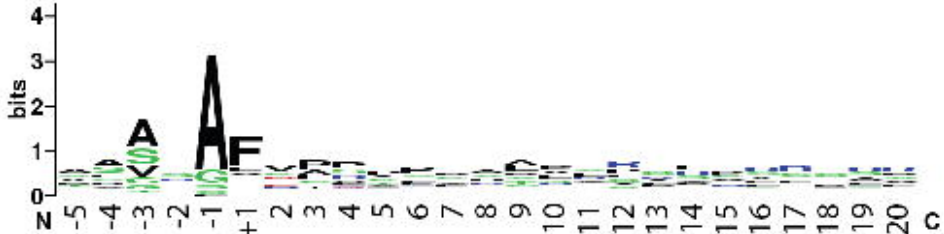

CYTOSOL
PLASTID
TETRAPYRROLES
ALA
Gly
succinyl-CoA
GLYCERO(PHOSPHO)LIPIDS
SQDG DGDG phosphatidylglycerol
fatty acids
ACP
SUF Fe-S ASSEMBLY
Cys
Fe3+
Fe2+
[Fe-S]
heme
sulfate
sulfide
SULFUR ASSIMILATION
THF
Gly
mTHF
GLY CLEAVAGE
RIBOSOME
DHAP
DHAP
TERPENOIDS
MEP
GGPP
phytyl-PP
Chl a
pyr
CALVIN CYCLE
CO2
GLYCOLYSIS
CAROTENOIDS
Tyr
TERPENOID-QUINONES
PQ
tocopherol
violaxanthin
fucoxanthin
MITOCHONDRION
PEP
PEP
ADP
ATP
pyr
NAD+
NADH
acetyl-CoA
FATTY ACID SYNTHESIS
ACP
apo-ACP
CoA
acyl-[ACP]
RIBOSOME
RIBOSOME
transporter
plastid genome-encoded
protein missing in Cv, Vb, both


STORAGE SUGAR
amylopectin
glucose
CYTOSOL
PLASTID
PENTOSE-P CYCLE
pyridoxin
Ru-5P
Ery-4P
PEP
CHORISMATE
AMINO ACIDS
PRPP
Trp
Tyr
Phe
Ser, Cys, Met
BIOSYNTHESIS
Cys
Ser
Met
GLYCOLYSIS
DHAP
PEP
CALVIN-BENSON
CYCLE
PRPP
GTP
CO2
FAD
RIBOFLAVIN
SYNTHESIS
STROMA
LUMEN
PSII
PQH2
PQ
b6f
cc6
NADPH
NADP+
FR
Fd
PSI
ATP
ADP
ATPS
TAT
SEC
Alb3
THYLAKOIDS
succinyl-CoA
TCA
CYCLE
pyr
lactate
OAA
2-OG
Glu
Gln
ADP
ATP
NH3
orn
Glu, Arg, Pro
BIOSYNTHESIS
CO2
Gln
urea
Arg
Glu
Pro
PRPP
His
His BIOSYNTHESIS
pyr
Glu
OAA
Ala
2-OG
Asp
Asn
Thr
Lys
Asn, Thr, Lys, Ala
BIOSYNTHESIS
MITOCHONDRION
IMS
MATRIX
CII
CIII
CIV
ATPS
UQH2
UQ
ATP
ADP
transporter
plastid/mitochondrial genome-encoded
protein missing in Cv, Vb, both

Chromera velia



Vitrella brassicaformis



	ASAFind	ASAFind+	HECTAR	TargetP - mTP		HECTAR - mTP		MultiLoc2 - mTP	
C. velia				(s)	(p)	(s)	(p)	(s)	(p)
THRESHOLD:	1.4	1.4	0.05	0.35	0.8	0.14	0.41	0.25	0.95
SENSITIVITY:	0.774	0.830	0.769	0.757	0.586	0.748	0.586	0.748	0.361
PRECISION:	0.921	0.921	0.872	0.375	0.844	0.529	0.855	0.485	0.736
V. brassicaformis									
THRESHOLD:	1	1	0.16	0.56	0.81	0.18	0.37	0.64	0.89
SENSITIVITY:	0.759	0.897	0.744	0.757	0.586	0.748	0.559	0.755	0.582
PRECISION:	0.886	0.916	0.967	0.712	0.855	0.664	0.861	0.697	0.842