PHOTOSYNTHESIS DRIVES RETENTION OF BACTERIAL-LIKE TRNA METABOLISM IN PLANT ORGANELLES

Rachael Ann DeTar[1,], Joanna Chustecki[2], Anna Martinez-Hottovy[2], Luis Federico Cerrioti[3,4], Amanda K. Broz[1], M. Virginia Sanchez-Puerta,[3,4], Christian Elowsky[5], Alan C. Christensen[2], Daniel B. Sloan[1]*

[1] Department of Biology, Colorado State University, Fort Collins, CO, USA
[2] School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA
[3] IBAM, Universidad Nacional de Cuyo, CONICET, Facultad de Ciencias Agrarias, Almirante Brown 500, M5528AHB Chacras de Coria, Argentina
[4] Facultad de Ciencias Exactas y Naturales, Padre Jorge Contreras 1300, Universidad Nacional de Cuyo, M5502JMA Mendoza, Argentina
[5] Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA
* Corresponding Author

Keywords: organelle gene expression, plastome, mitogenome, heme biosynthesis, tRNA, photosynthesis, aminoacyl-tRNA synthetase

Abstract

Eukaryotic nuclear genomes often encode distinct sets of protein translation machinery for function in the cytosol vs. organelles (mitochondria and plastids). This phenomenon raises questions about why multiple translation systems are maintained even though they are capable of comparable functions, and whether they evolve differently depending on the compartment where they operate. These questions are particularly interesting in land plants because translation machinery, including aminoacyl-tRNA synthetases (aaRS), is often dual-targeted to both the plastids and mitochondria. These two organelles have quite different metabolisms, with much higher rates of translation in plastids to supply the abundant, rapid-turnover proteins required for photosynthesis. Previous studies have indicated that plant organellar aaRS evolve more slowly compared to mitochondrial aaRS in other eukaryotes that lack plastids. Thus, we investigated the evolution of nuclear-encoded organellar and cytosolic translation machinery across a broad sampling of angiosperms, including non-photosynthetic (heterotrophic) plant species with reduced rates of plastid gene expression to test the hypothesis that translational demands associated with photosynthesis constrain the evolution of bacterial-like enzymes involved in organellar tRNA metabolism. Remarkably, heterotrophic plants exhibited wholesale loss of many organelle-targeted aaRS and other enzymes, even though translation still occurs in their mitochondria and plastids. These losses were often accompanied by apparent retargeting of cytosolic enzymes and tRNAs to the organelles, sometimes preserving aaRS-tRNA charging relationships but other times creating surprising mismatches between cytosolic aaRS and mitochondrial tRNA substrates. Our findings indicate that the presence of a photosynthetic plastid drives the retention of specialized systems for organellar tRNA metabolism.

Significance

The process by which endosymbionts are integrated into a host and become organelles results in a combination of gene loss, transfer to the nucleus, and retention in the organellar genome. It is not well understood why some endosymbiont-derived genes may be retained when a functional host counterpart exists whose gene product could simply be retargeted to the organelles. This study revealed that the photosynthetic activity in plant plastids may be responsible for retention of functionally redundant tRNA processing machinery, while mitochondria are more flexible regarding substitution with cytosolic type enzymes. Therefore, functional constraint in the plastid is likely more important than in the mitochondria for shaping the evolution and retention of translation machinery that is dual-targeted to both organelles.

Introduction

The evolutionary transition from a free-living bacterium to an endosymbiont and eventually to an organelle requires the physical, biochemical, and genomic integration of two organisms. One common consequence of this process is a reduction in the genome of the endosymbiont, driven by relocation of

5 essential genes to the nucleus via horizontal transfer or outright loss of some genes that are no longer
6 useful after transition to an intracellular lifestyle. This trajectory has been observed repeatedly for
7 various endosymbionts and organelles, including mitochondria, plastids, and more recently acquired
8 intracellular bacteria such as the cyanobacterial-derived chromatophores of *Paulinella* amoeba or the
9 many heritable endosymbionts in insects (1–6). However, there are some endosymbiotically derived
10 genes that are recalcitrant to loss despite the presence of host counterparts with comparable functions.
11 Redundant genes may be maintained in the organelle genomes or in the nuclear genome alongside
12 genes encoding proteins for the same role in other cellular compartments. These observations elicit
13 some interesting questions; 1) Why are some functionally redundant genes retained? 2) Do these
14 redundant genes evolve differently depending on the cellular compartment where they operate?
15 The genes that encode the machinery required for organelle translation are widely retained, as
16 eukaryotes maintain distinct translation systems for nuclear and mitochondrial transcripts. Translation
17 becomes even more complicated in eukaryotes bearing three genomes. For example, plant cells contain
18 nuclear, mitochondrial and plastid genomes. The translation machinery associated with each of these
19 cellular compartments includes ribosomal RNAs (rRNAs), ribosomal proteins, transfer RNAs (tRNAs), and
20 the accompanying enzymes involved in tRNA processing and aminoacylation. The aminoacyl-tRNA
21 synthetases (aaRS) that catalyze the "charging" of a tRNA with its conjugate amino acid tend to be highly
22 conserved, with distinct versions for function in the organelles and cytosol (7). The degree of
23 redundancy of tRNAs and their interactors between cellular compartments also varies significantly
24 within eukaryotes. These characteristics make tRNAs and their interactors, particularly aaRS, an ideal
25 case study for understanding cytonuclear gene redundancy.
26 The concerted activity of the aaRS along with other enzymes involved in post-transcriptional
27 modification such as tRNases P and Z (end processing in tRNA maturation) and CCAases (addition of CCA
28 tail) ensures the tRNA has the proper structure to fulfill its role in translation. Generally, organisms

29 would need a complement of at least 20 aaRS, one for each amino acid. However, eukaryotes often have
30 more than this minimum requirement, with some overlapping sets of enzymes targeted to different
31 compartments (8, 9). It is not fully understood why some eukaryotes retain distinct, redundant sets of
32 tRNA processing machinery, while others do not. However, retention of organelle-targeted aaRS
33 correlates with the retention of tRNA genes within organelle genomes. For example, *Homo sapiens*
34 retains 38 aaRS with nearly distinct sets for the cytosol and mitochondria (9–11). Correspondingly, all
35 necessary tRNA genes are present in the mitochondrial genome (mitogenome) (8). In contrast,
36 trypanosomatids do not encode any tRNAs in their mitogenomes but rather import nuclear-encoded
37 tRNAs into the mitochondria and have a reduced set of aaRS which colocalize to the cytosol and the
38 mitochondria (9, 12, 13). In plants, closely related species may have very different numbers of tRNAs
39 encoded in their mitogenomes, with import of cytosolic tRNAs to compensate for losses (8, 14, 15).
40 Notably, plants also dual-target and import an organellar set of enzymes into both plastids and
41 mitochondria (9, 16). The co-occurrence of redundant tRNAs and aaRS could be explained by the
42 observation that organelle-encoded tRNAs are often structurally divergent from cytosolic tRNAs and
43 require specific enzymatic partners for recognition (8, 17).
44 For most eukaryotes, organellar and cytosolic aaRS are exclusively nuclear-encoded. The shared
45 genomic location of these genes facilitates direct comparison of sequence conservation because any
46 differences in evolutionary rate should be more likely to reflect differences in selection, rather than
47 other factors like compartment-specific mutation rates or effective population size (18–21). Previous
48 studies have quantified rates of evolution for these enzymes in various animal lineages and a small
49 number of plants. Mitochondria-specific aaRS in animals exhibit lower sequence conservation than their
50 cytosolic counterparts (22–24). In contrast, plant organellar aaRS exhibit equivalent or greater sequence
51 conservation compared to their cytosolic counterparts (25). Furthermore, plastid genomes (i.e.,
52 plastomes) tend to retain a set of 30 tRNA genes that is fully sufficient for translation (26, 27). Plastomes

53 also generally have high rates of transcription (28) and translation (29) due to the abundance and rapid
54 turnover of proteins required for photosynthesis. Thus, the disparity in evolutionary rates between
55 animal mitochondrial aaRS and plant dual-organelle targeted aaRS may be due to the specific
56 translational needs of the plastid. We hypothesize that aaRS and other translation-associated enzymes
57 operating in the plastid face great functional constraints, thus promoting unusually high sequence
58 conservation of organellar translation machinery compared to other eukaryotes. A prediction stemming
59 from this hypothesis is that plant lineages with reduced or nonexistent plastid gene expression would
60 exhibit more rapid evolution of organellar aaRS due to relaxed selection.

61 In this work, we test whether photosynthetic activity contributes to sequence conservation of
62 organellar translation machinery by analyzing evolution of this machinery across diverse lineages of
63 autotrophic and heterotrophic plants. It should be noted there is a gradient of heterotrophy amongst
64 plants – for the purposes of this study, we use "heterotrophic" to describe plants that do not
65 photosynthesize at all and exclusively rely on parasitism for acquisition of carbon. They do this by
66 tapping directly into host plants or by exploiting mycorrhizal fungi (mycoheterotrophy) (30, 31). We
67 refer to facultative autotrophs that supplement photosynthetic carbon with parasitism as
68 "hemiparasitic." The hallmark of many heterotrophic/hemiparasitic plants is a reduced or even non-
69 existent plastome (32). Highly expressed genes related to photosynthesis such as the Rubisco large
70 subunit and the Photosystem II D1 reaction center tend to be absent from plastomes in heterotrophs,
71 thereby drastically decreasing demands on plastid gene expression. In addition, many tRNA genes have
72 been lost from the plastomes of these plant species (32), providing a good system to examine the
73 relationship between evolutionary redundancy of aaRS, tRNA content of the organelle genomes, and
74 trophic lifestyle.

75 ***Results***

76 ***Heterotrophic plants exhibit extensive loss of organellar aaRS and tRNA processing enzymes***

77 We searched for organellar and cytosolic aaRS orthologs in a broad sampling of angiosperm species with
78 a range of trophic lifestyles and plastome reduction (**Supp. Dataset 1**). For each
79 hemiparasitic/heterotrophic species in the analysis, we included the closest fully autotrophic relative
80 with a nuclear genome or transcriptome available, excepting the Balanophoraceae where we used
81 hemiparasite *Santalum album* as the reference autotroph. In total, we sampled 7 heterotrophs, 8
82 hemiparasites, and 10 autotrophs, including reference species *Arabidopsis thaliana.* Among
83 heterotrophs, we included the endoparasites *Rafflesia cantleyi* and *Sapria himalayana* from the family
84 Rafflesiaceae, likely the most extreme examples of plant heterotrophy, as several studies have
85 suggested wholesale loss of the plastome in these species (33–36).
86 We found orthologs across these diverse taxa by running Orthofinder (37, 38) on protein models
87 derived from nuclear genome or transcriptome data to find groups of likely orthologous proteins based
88 on similarity and phylogeny, i.e. "orthogroups". Then we extracted sequences that fall into an
89 orthogroup with known *A. thaliana* aaRS (**Supplemental Dataset 1**). Unexpectedly, many lineages of
90 heterotrophs did not have any orthologs for many of the organellar aaRS. This loss of organelle-specific
91 orthologs in heterotrophs was surprising as even species entirely lacking the plastome must still conduct
92 tRNA aminoacylation within mitochondria. Further investigation using reciprocal-best-hit (RBH) searches
93 with BLAST between *A. thaliana* and each species supported that these were true losses. These losses
94 primarily occurred in purely heterotrophic plants, while hemiparasites retained most of their organellar
95 aaRS. (**Fig. 1**). Yet, within the heterotrophs, there was a gradient of loss. The two endophytic species
96 from Rafflesiaceae (*R. cantleyi* and *S. himalayana*) exhibited the most striking loss of organellar aaRS,
97 retaining only PheRS. In contrast, obligate mycoheterotroph *Monotropa hypopitys* retained more aaRS
98 than any other heterotroph, missing only AlaRS, CysRS, ThrRS, and TrpRS (**Fig. 1**).
99 This striking pattern led us to investigate the retention of other enzymes involved in tRNA
100 maturation and processing; see (15) and **Supplemental Dataset 2** for a comprehensive list.

101 Unsurprisingly, the CCAase enzyme responsible for adding CCA tails to tRNAs in all compartments was
102 retained across all species. Two enzymes that maintain distinct cytosolic and organellar versions, tRNase
103 Z (involved in 3' end processing) and Protein Only RNase P (PRORP; required for 5' end processing),
104 exhibited two different trends. We found that plastid tRNase Z (39) was missing from all heterotrophs
105 except *M. hypopitys* and hemiparasites of the *Cuscuta* genus yet was retained in the autotrophic
106 relatives. Conversely, the organellar PRORP1 (40, 41) appeared to be universally retained while cytosolic
107 PRORP2/3 orthologs were missing from the Rafflesiaceae. We also investigated the retention of
108 bacterial-like organelle-specific machinery with no cytosolic counterpart, including the methionyl-tRNA
109 formyltransferase (MTF), tRNA-Ile lysidine synthetase (TilS) and the bacterial-type glutamine
110 amidotransferase (GatCAB) complex. We found orthologs for the putative *A. thaliana* MTF in all species
111 except the heterotrophs *Epigogium roseum* and *M. hypopitys,* and putative TilS orthologs were absent
112 only from the heterotrophs Rafflesiaceae and *Balanophora fungosa*. GatCAB is the key complex used by
113 organelles in land plants to accurately charge $tRNA^{Gln}$ in the absence of a dedicated organellar GlnRS
114 (42–44). This involves a two-step process, whereby $tRNA^{Gln}$ is initially mis-acylated with Glu by GluRS.
115 GatCAB then catalyzes the transamidation of Glu to Gln. Most of our test species, even heterotrophs,
116 retained genes for at least some of the GatCAB subunits. However, Rafflesiaceae species appear to have
117 lost all subunits in this complex.

118 Generally, we observed reductive evolution in heterotrophs via loss of organellar aaRS and
119 bacterial-type enzymes for tRNA modification. This begs the question as to how gene expression is
120 accomplished in the organelles of these species, particularly within mitochondria. One possibility is that
121 cytosolic enzymes are retargeted to the organelles to functionally replace organellar enzymes.

122 ***Heterotrophic plants may compensate for loss of organellar aaRS by retargeting cytosolic***
123 ***counterparts***

124 To test for cytosolic aaRS retargeting, we applied two *in silico* targeting prediction programs to protein
125 models for all cytosolic aaRS in our heterotrophs and their close autotrophic relatives. Our analysis
126 revealed that canonically cytosolic aaRS are disproportionately retargeted to mitochondria in obligate
127 heterotrophs compared to autotrophs (**Fig. 2,** $p = 0.0341$). There was also a non-significant trend
128 towards retargeting to the plastid ($p = 0.190$). Targeting of mitochondrial OXPHOS components and cell
129 wall synthesis enzymes were included as positive and negative controls, respectively. No plastid positive
130 control is provided as the reduced genome and altered metabolism of plastids in heterotrophic plants
131 precludes straightforward selection of a control gene set. For both heterotrophs and autotrophs, an
132 average of about 60% of OXPHOS proteins were predicted to be targeted to the mitochondria, and
133 about 10% to the plastid. About 10% of cell wall synthesis enzymes were predicted to be targeted to the
134 organelles for both groups (**Fig. 2**). Thus, our analysis appears to be detecting a legitimate biological
135 trend toward retargeting of cytosolic aaRS in heterotrophs, albeit with a generally high false negative
136 rate for mitochondrial retargeting. Furthermore, within the heterotrophic species there was a significant
137 correlation between loss of an organellar aaRS and retargeting of the corresponding cytosolic enzyme to
138 one or both compartments ($p = 0.005$, **Fig. S1**).
139 The variable quality of gene models for each species and high false negative rate for retargeting
140 prediction confound interpretation of retargeting for individual enzymes for each species (**Fig. S2**).
141 However, when looking only at *S. himalayana*, a species with relatively high-quality genome-based
142 models, we observed some interesting phenomena. Cytosolic GlnRS, MetRS, ProRS, and TrpRS were
143 predicted to be retargeted to mitochondria. Retargeting of cytosolic GlnRS could explains how this
144 species copes with loss of the organellar GatCAB complex (**Fig. 1**). The cytosolic AspRS and GlyRS were
145 predicted to be dual-targeted to both organelles, while AsnRS and TyrRS were predicted to be targeted
146 to the plastid despite the likely non-existence of a plastid genome. All other cytosolic enzymes were
147 either not predicted to be retargeted or were missing a large portion of the N-terminus.

148 To assess whether our *in silico* predictions hold true *in vivo,* we conducted targeting assays on a
149 subset of aaRS (see **Supplemental Dataset 3** for sequence information) by inducing transient expression
150 of N-terminal fusions of *S. himalayana* cytosolic aaRS transit peptides (TPs) to green fluorescent protein
151 (GFP) in *Nicotiana benthamiana* (**Fig. 3, Fig. S3**). We used GFP linked to the mitochondrial cytochrome C
152 TP from *S. himalayana* as an additional mitochondrial-targeted positive control (**Fig. S3**). We found that
153 *in silico* predictions were generally reflected *in vivo,* but some putative TPs targeted GFP to a different
154 compartment *in vivo*. For example, the AsnRS TP was predicted to be plastid-targeted *in silico*, yet
155 showed mitochondrial targeting during transient expression assays. Conversely, ProRS and TrpRS were
156 predicted to be mitochondrial *in silico*, yet exhibited plastid targeting *in vivo.* Additionally, we gained
157 further evidence that our *in silico* analysis produced false negatives for organellar targeting of some
158 aaRS TPs, as the N-terminal extension of ThrRS was not predicted to be an organellar transit peptide *in*
159 *silico,* yet retargeted GFP to the plastids in *N. benthamiana.* The only TP we cloned expecting organellar
160 retargeting which did not exhibit targeting to either organelle *in vivo* was AspRS (**Fig. S3**). GFP tagged
161 with this TP did not colocalize with mitochondrial FP611 florescence or plastid autofluorescence, but
162 rather formed punctate structures in the cytosol which are likely artefactual aggregates resulting from
163 overexpression. Generally, *in vivo* assays provided evidence for retargeting of cytosolic aaRS to at least
164 one organelle and suggested that *in silico* prediction tools may even be overly conservative.
165 Generally, organellar tRNAs differ from their cytosolic counterparts in evolutionary history,
166 sequence, and posttranscriptional modification (8). Identity elements which govern the specificity of
167 interaction between a tRNA and its conjugate aaRS may also differ between cytosolic and organellar
168 systems (45). However, in many systems (including plants), cytosolic tRNAs can be imported into the
169 organelles (8, 9, 14). Thus, our observation of retargeting opens some interesting questions – are the
170 retargeted aaRS charging organellar tRNAs? Or is there replacement of the entire organellar system of

171 tRNA aminoacylation, including both tRNAs and aaRS? We investigated the tRNA gene content of the
172 organelles of these species to get insight into which tRNA substrates may be in use.

173 ***Organelle-encoded tRNA retention often, but not always, mirrors retention of organelle-type aaRS***

174 To investigate the possibility of enzyme-substrate mismatch, we cataloged the presence/absence of
175 organellar tRNA-encoding genes (*trn*) for all sampled heterotrophic species (**Fig. 4, Supplemental**
176 **Dataset 4**). Generally, there was strong correspondence between retention of organellar aaRS and
177 plastome/mitogenome tRNA genes. As the Rafflesiaceae species are thought to lack plastomes entirely,
178 they likely have no plastome-encoded tRNAs (**Fig. 4**). The Balanophoraceae species also have been
179 reported to exhibit loss of all plastid tRNA genes (46–49). The Rafflesiaceae and Balanophoraceae have
180 additionally lost most of their mitogenome-encoded tRNAs. Correspondingly, these two families also
181 have the most reduced sets of organellar aaRS. At the opposite extreme, *M. hypopitys* retains most of its
182 mitochondrial and plastid tRNA genes and corresponding organellar aaRS.

183 However, there are some incidences where a cytosolic enzyme looks to be charging an
184 organellar tRNA due to loss of the organellar aaRS (**Fig. 4**). For example, the few remaining mitogenome-
185 encoded tRNAs in the Rafflesiaceae and Balanophoraceae species (excepting *S. himalayana* $tRNA^{Phe}$
186 encoded by *trnF*) are missing their organellar aaRS partner and are likely charged by a cytosolic enzyme.
187 For both compartments, there are numerous instances where there is no encoded tRNA despite
188 retention of the organellar aaRS (**Fig. 4**).

189 ***Testing for accelerations in organellar aaRS sequence evolution in heterotrophs***

190 Although wholesale loss of many organellar aaRS in heterotrophs precludes sequence conservation
191 analysis, *M. hypopitys* retained the majority of aaRS sequences while still exhibiting an obligate
192 heterotrophic lifestyle. Thus, we quantified sequence conservation of organellar and cytosolic aaRS in
193 *M. hypopitys* and its autotrophic relative *Rhododendron williamsianum.* We built gene trees and ran
194 relative rate analysis to test for accelerated evolution on the *M. hypopitys* branch (**Supplemental**

195 **Dataset 5**). We also directly compared branch length ratios of *M. hypopitys* to *R. williamsianum* in
196 cytosolic and organellar aaRS (**Fig. S4, Supplemental Dataset 5).** In neither analysis was there a
197 statistically significant trend towards decreased sequence conservation in organellar enzymes of *M.*
198 *hypopitys.* However, the deep rooting of our tree on *A. thaliana* may result in poor resolution of the
199 sequence evolution between *M. hypopitys* and *R. williamsianum.* Additionally, the transition to
200 heterotrophy on the *M. hypopitys* branch may have occurred relatively recently compared to the split
201 from *R. williamsianum*, thus diluting any signals of accelerated evolution.

202 ***Loss of GluRS and trnE in Rafflesiaceae may be facilitated by relocation of tetrapyrrole pathway to the***
203 ***cytosol***

204 In addition to its role in translation, aminoacylated plastid $tRNA^{Glu}$ (encoded by the *trnE* gene) is the first
205 precursor in the plastid tetrapyrrole pathway. This bifurcating pathway begins with a common "stem"
206 set of enzymes providing the 5-aminoolevulinic acid (ALA) precursor for both the heme and chlorophyll
207 biosynthesis branches (50, 51). While heterotrophic plants do not need chlorophyll for photosynthesis,
208 heme is still a necessary cofactor for many enzymes and protein complexes involved in redox chemistry.
209 Accordingly, *trnE* is one of the most-retained genes in the plastomes of heterotrophic plant species (32)
210 and may even drive the retention of plastomes in non-photosynthetic organisms (52, 53). Yet, the
211 heterotrophs from Rafflesiaceae and Balanophoraceae used in this study have lost plastid *trnE* (35, 46).
212 We found that the Rafflesiaceae also are lacking the organellar GluRS typically required to aminoacylate
213 plastid $tRNA^{Glu}$ for use in the tetrapyrrole pathway. Although a previous study of *R. cantleyi* (54)
214 reported the presence of a GluRS gene, which could potentially catalyze the first step in the tetrapyrrole
215 pathway, they did not investigate if it was an organellar or cytosolic enzyme. Intriguingly, the cytosolic
216 GluRS is not predicted to be retargeted to the plastid in the Rafflesiaceae, which prompted us to
217 investigate the presence/absence of tetrapyrrole pathway enzymes in our study species.

218 For both analyses, genes are grouped based on their placement in the tetrapyrrole pathway –
219 the shared "common pathway" or the heme and chlorophyll branches. Unsurprisingly, we found that
220 the heterotrophs (with exception of *Epipogium*) were lacking most genes in the chlorophyll biosynthesis
221 branch of the tetrapyrrole pathway (**Fig. S5**). The Rafflesiaceae also exhibited loss of the heme
222 oxygenase (HO) and phytochromobilin synthase (HY2) enzymes from the tetrapyrrole pathway, which
223 are necessary to produce phytochromobilin for light sensing. This may be a response to their
224 endoparasitic lifestyle where they only emerge from the host to flower and therefore can dispense with
225 light signaling. All heterotrophs retained most of the enzymes in the common pathway and the
226 ferrachelatase (FC), which catalyzes the final step in the production of heme. Therefore, it does not
227 appear that the tetrapyrrole pathway has been lost from any of the species, including those that lack
228 plastid *trnE* and/or organellar GluRS. A previous study on *Lophophytum* suggests that Balanophoraceae
229 species lacking plastid *trnE* likely import the nuclear-encoded $tRNA^{Glu}$ but otherwise utilize the typical
230 plastid heme biosynthesis pathway (46). However, we have found the Rafflesiaceae lack both plastid
231 *trnE* and GluRS, and do not have predicted retargeting of the cytosolic GluRS. Thus, we hypothesized
232 that the Rafflesiaceae may adopt a different strategy to continue heme production by retargeting the
233 entire pathway to the cytosol and using nuclear-encoded $tRNA^{Glu}$ and cytosolic GluRS. We investigated
234 this hypothesis using *in silico* targeting predictions and found that common pathway enzymes are not
235 predicted to be plastid-localized in the Rafflesiaceae, in contrast to autotrophic relative *Manihot*
236 *esculenta*, which retains plastid targeting (**Fig. S6**). Furthermore, plastid targeting was still predicted for
237 most tetrapyrrole enzymes in the other heterotrophic species that retain organellar GluRS, including the
238 Balanophoraceae species. Thus, it is possible that *S. himalayana* and *R. cantleyi* produce heme without a
239 functional plastid gene expression apparatus by relocating some steps in the tetrapyrrole pathway to
240 the cytosol.

241 ***Discussion***

242 ***Loss of photosynthesis relaxes constraints on both plastid and mitochondrial translation machinery***

243 Our primary objective in this study was to test if the high translational demands of the photosynthetic
244 chloroplast create functional constraint on the evolution of shared translation systems in plastids and
245 mitochondria. We accomplished this by investigating relative rates of protein evolution in
246 cytosolic/organellar aaRS in heterotrophic plants compared to their photosynthetic relatives. To our
247 surprise, heterotrophic plants underwent widespread loss of organellar aaRS and many bacterial-like
248 enzymes involved in tRNA processing. These losses suggest an extreme relaxation of selection pressures
249 in the absence of a photosynthesizing plastid and support our hypothesis that the translational demands
250 in chloroplasts are the linchpin responsible for the unusually strong functional constraint on the
251 redundant organelle-specific translation machinery in plants.

252 There was variation in the degree of loss across the heterotrophs, often correlated with the
253 number of genes retained in the plastome, particularly tRNAs (see **Supp. Dataset 1** for
254 plastome/mitogenome content). For example, the most losses occurred in the Rafflesiaceae, all of which
255 are obligate endoparasites and have likely lost their plastomes (33–36, 55). The fewest losses occurred
256 in *M. hypopitys*, which has the highest number of plastome-encoded genes of all the heterotrophic
257 species we sampled – and notably the greatest number of retained plastid tRNAs (56). Thus, *M.*
258 *hypopitys* may not be as far along the path of reductive evolution as the other species, which might
259 provide an additional explanation as to why our comparative rate analyses on the remaining organellar
260 aaRS in *M. hypopitys* did not support accelerated evolution compared to cytosolic enzymes (**Fig. S4**). The
261 Balanophoraceae and *Epipogium* species exhibited intermediate degrees of aaRS and plastome gene
262 loss, although the Balanophoraceae have lost all plastid tRNAs (Schelkunov *et al.*, 2015, 2019; Chen *et*
263 *al.*, 2019; Serna-Sánchez *et al.*, 2021; Yu *et al.*, 2022). An increased plastome mutation rate caused by
264 loss of DNA repair machinery (57, 58) in the Balanophoraceae may explain accelerated plastome (tRNA)

265 gene losses relative to nuclear (aaRS) gene loss. Generally, loss of organellar aaRS correlates well with
266 reduction of the plastome across heterotrophs.

267 ***Loss of organellar aaRS is associated with retargeting of cytosolic aaRS and potential rewiring of***
268 ***aaRS-tRNA interactions***

269 Wholesale loss of organellar aaRS in heterotrophs was unexpected, as most non-photosynthetic plants
270 still contain a minimal plastome to support heme and fatty acid biosynthesis (32, 53) and fully intact
271 mitogenomes (see **Supp. Dataset 1**). This conundrum prompted us to investigate if cytosolic aaRS are
272 retargeted to organelles in heterotrophs to compensate for gene loss, as has been shown for
273 aaRS/tRNAs in mitochondria of other plant species (59). Retargeting a protein to the organelles
274 generally involves obtaining an N-terminal TP sequence that can be recognized by organellar protein
275 import complexes (60–62). We discovered that many cytosolic aaRS in the heterotrophs had N-terminal
276 sequences predicted to target to one or both organelles (**Fig. 2-3, S1-S3**). Therefore, the import of
277 cytosolic aaRS likely compensates for the loss of organellar aaRS and even the loss of the GatCAB
278 complex in Rafflesiaceae. There was not perfect correspondence between loss of organellar aaRS and
279 predicted retargeting of cytosolic aaRS. However, the extent of retargeting may have been
280 underestimated for several reasons: 1) TP sequence composition varies considerably by plant lineage
281 (63), making *in silico* predictions more challenging in uncharacterized species, 2) the low quality of
282 protein models for some of our species may have led to true N-terminal sequences being absent or
283 incomplete, and 3) some organelle-imported cytosolic aaRS may eschew the use of typical TPs that are
284 key for targeting prediction *in silico* (64). Retargeting of cytosolic aaRS may occur through two different
285 mechanisms; 1) alternative splicing or alternative transcription/translation start sites within one gene
286 resulting in cytosolic and organellar isoforms and 2) duplicated genes within the genome encoding
287 cytosolic and organellar paralogs. Our ability to investigate these two mechanisms was limited as we
288 generally only had transcriptome or genome information but not both for each species. However,

289 observations of transcriptomes assembled with consideration for alternative splice isoforms indicate
290 that both mechanisms are potentially at play (**Supplemental Dataset 6**). For example, *R. phalloides* has
291 two separate paralogs for cytosolic ThrRS, one of which is predicted to be mitochondrial-targeted
292 (Rphallo_DN1669_c0_g1) while the other is not (Rphallo_DN1669_c0_g2). Conversely, *R. phalloides*
293 TyrRS is encoded by a single locus (Rphallo_DN204_c0_g1) with a multitude of splice isoforms, some of
294 which are putatively mitochondrial-targeted. Genome sequencing and long-read mRNA sequencing
295 would allow for more thorough analysis of which mechanism is most common in each species.
296 Retargeting of cytosolic aaRS led us to investigate organelle genome tRNA content and whether
297 there was substrate-enzyme mismatch in the organelles, i.e. instances where the cytosolic enzyme was
298 interacting with an organellar tRNA or vice-versa. Our interpretations presume that loss of organellar
299 tRNA genes is compensated by import of the corresponding cytosolic tRNAs (9, 14, 65). We found that
300 the retention of organellar aaRS in heterotrophs generally matched the retention of plastid tRNAs but
301 not necessarily mitochondrial tRNAs (**Fig. 4**). This is further supported by observations in hemiparasite
302 *Viscum album*, which has lost most of its mitochondrial tRNAs, yet retains most organellar aaRS in
303 conjunction with a nearly full set of plastid tRNAs (32, 66, 67). Substrate-enzyme mismatch was rare
304 within the plastid and usually involved loss of a plastid tRNA yet continued presence of the
305 corresponding organellar aaRS (e.g., see *R. phalloides* where all plastid tRNAs are lost but a handful of
306 organellar aaRS remain). There were also instances of organellar aaRS apparently charging imported
307 cytosolic tRNAs in the mitochondria (e.g., *trnF* in *R. cantleyi and R. phalloides*). More surprisingly,
308 however, there were cases where a mitochondrial tRNA was retained despite loss of the corresponding
309 organellar aaRS and apparent retargeting of the cytosolic enzyme, particularly in the Rafflesiaceae. This
310 type of mismatch is unexpected because cytosolic aaRS are thought to generally be more discriminating
311 enzymes than their mitochondrial counterparts (8, 59, 68–70), making it more difficult to transition to a
312 new tRNA substrate. Our results suggest some cytosolic enzymes may be interacting with bacterial-type

313 tRNAs in these heterotrophic species, although we cannot rule out the possibility that cytosolic tRNAs
314 are being redundantly imported. Given that parasitic plant lineages tend to be fast evolving across all
315 three genomes (71), the cytosolic aaRS in these species may have evolved lower specificity and the
316 capability to charge organellar tRNAs. Overall, our findings suggest that the plastid tRNA content, and
317 not the mitochondrial tRNA content, drives retention of organellar aaRS, again placing the
318 photosynthetic plastid as a central factor in the retention of functionally redundant translation
319 machinery.

320 The identity of retained organellar aaRS versus those substituted by cytosolic counterparts was
321 seemingly non-random. GluRS and PheRS were the two mostly highly retained organellar aaRS across
322 these diverse lineages of plants (**Fig. 1**). As $tRNA^{Glu}$ is the most frequently retained tRNA in heterotrophic
323 plant plastomes due to its additional role in heme biosynthesis (50, 53), it is not surprising that GluRS
324 was among the most retained organellar aaRS in this study. The Rafflesiaceae were the only exception,
325 as they might have evolved retargeting of the entire heme pathway to the cytosol to compensate for
326 lack of plastid *trnE*. Organellar PheRS is likely retained because the cytosolic PheRS is unusual in that it is
327 a heteromeric complex with independently encoded α and β subunits (72). Therefore, functional
328 replacement of the organellar PheRS would require two independent acquisitions of TPs, as well as
329 stoichiometric import and assembly within the organelles. Such a coordinated set of changes may be
330 unlikely to evolve (24, 59). The one exception was *E. aphyllum,* for which we could not find an organellar
331 PheRS. However, this may be an artefact of low expression, as protein models for this species were
332 curated from RNA-seq data. At the other extreme, there was repeated loss of organellar AlaRS, which
333 was absent from all heterotrophs and the hemiparasitic *Cuscuta* species, possibly because plastomes of
334 heterotrophs are often AT-rich and thus have underrepresentation of amino acids with GC-rich codons
335 like Ala, Gly and Pro (47, 48). Thus, the corresponding organellar aaRS and tRNAs may be subject to

336 decreased functional constraint and be less likely to be retained. This hypothesis is supported by the fact
337 that we also see frequent loss of organellar GlyRS and ProRS in heterotrophs (**Fig. 1**).
338 Our results indicate that imported cytosolic aaRS and tRNAs are sufficient to sustain
339 mitochondrial translation in plants, supporting the hypothesis that the selective pressure to keep
340 organellar aaRS/tRNAs is maintained by translational demands associated with photosynthesis. Yet
341 other eukaryote lineages that never bore plastids often still retain organelle-specific tRNAs and aaRS for
342 their mitochondria (8, 9, 22, 24). This is even more puzzling when considering that various eukaryotic
343 lineages have jettisoned nearly all their mitochondrial tRNAs/aaRS, including cnidarians, ctenophores,
344 some fungi, and trypanosomes (12, 24, 73). One possible explanation is the logical parallel to our finding
345 that photosynthesis governs plastid translation rates and maintains gene redundancy; eukaryotic
346 lineages with high respiratory rates may require high mitochondrial gene expression and, therefore,
347 retain mitochondria-specific translation machinery. Indeed, mitochondrial gene expression tracks well
348 with respiratory rate and cellular activity across different tissues and metabolic states (74–76). This
349 connection to cell energetics would also explain why mitochondrial tRNAs/aaRS are disposable in
350 parasitic eukaryotes which presumably have lower energy demands, like trypanosomes and
351 heterotrophic plants. By studying non-model plant species with unusual physiology and life histories, we
352 were able to link plastid bioenergetic processes to the evolution of both organellar and nuclear
353 genomes. The sequencing of mitogenomes and characterization of metabolism in non-model eukaryotes
354 will provide exciting opportunities to conduct similar studies to better understand how mitochondrial
355 metabolism effects gene redundancy and elucidate the principles governing endosymbiotic integration.

356 ***Methods***

357 ***Data acquisition and ortholog search***

358 Gene models for species used in the study were collected from publicly available data sources
359 (**Supplemental Dataset 1**). When protein sequences were not available, we generated sequences from

360 transcriptome data using Transdecoder v5.5.0 (77), applying a minimum protein length of 50 amino
361 acids and retaining only protein sequences with blast hits to any sequence in the Uniprot Taxonomic
362 Divisions Plant database (https://ftp.uniprot.org/pub/databases/uniprot/knowledgebase
363 /taxonomic_divisions/). Then, the Transdecoder primary_transcript.fa script was used to extract only
364 the longest translated protein sequence from the primary transcript for each gene to reduce the search
365 space. SeqKit 2.3.0 (78) was used to create uniform gene handles for all files and remove stop codons
366 from sequences. Orthofinder v2.5.4 (37) was run on all sets of protein models with a user-defined
367 species tree (see **Fig. 1** for tree topology) to find groups of likely orthologous proteins, i.e. "orthogroups"
368 in two steps as follows:

369 orthofinder.py -f <folder_with_protein_models> -t 24 -a 6

370 orthofinder.py -ft <folder_with_results_> -s userdefined_speciestree.txt -t 24 -a 6

371 Orthogroups containing proteins of interest including cytosolic/organellar aaRS and tRNA-processing
372 enzymes characterized in *A. thaliana* (**Supplemental Dataset 2**) were extracted and checked for
373 completeness via RBH, using blastp from the command line BLAST 2.13.0+ toolkit (79) and custom
374 Python/Bash scripts. After ortholog discovery via Orthofinder, full sets of protein models (not just the
375 longest) were searched for *A. thaliana* protein sets using blastp to ensure narrowing of search space for
376 Orthofinder did not result in loss of matching gene models. In cases where available curated protein
377 models excluded alternative transcripts, this required obtaining transcriptome data and running
378 Transdecoder (see **Supplemental Dataset 1** for details). Note that ArgRS was not included in the
379 ortholog search as there are no distinct organellar versions of this enzyme in *A. thaliana.* One ArgRS
380 gene product is cytosol-targeted, and the other is triple-targeted to cytosol, mitochondria, and plastids
381 (9, 16). In the case of GlnRS, there is no organellar ortholog as plants typically use the bacterial-type
382 GatCAB amidotransferase system to charge Gln-tRNA indirectly in mitochondria and plastids (42). For
383 CysRS, tRNase Z and PRORP, sequences for both cytosolic and organellar proteins were clustered

384 together in one orthogroup likely to due to recent shared ancestry. For these genes, cytosolic versus
385 organellar versions were differentiated by aligning sequences in each orthogroup using MAFFT v7.490
386 (80) with the --auto setting, followed by manual trimming of the alignment in MEGA 11 (Tamura et al.,
387 2021). Maximum-likelihood tree building was conducted using RaXML v 8.2.12 (82) as follows:
388 raxmlHPC-PTHREADS-SSE3 -f a -# 100 -p 12345 -x 12345 -s <fasta file with orthogroup sequences> -n
389 <orthogroup> -m PROTGAMMALG -T 12
390 Inclusion in cytosolic versus organellar groups was then determined by relationship to known *A. thaliana*
391 cytosolic/nuclear/organellar proteins in maximum likelihood trees (**Figs. S7-10**).
392 Once we had a curated set of sequences for aaRS and processing enzymes for all species,
393 orthologs were then checked for completeness of length by alignment with *A. thaliana* orthologs by
394 MAFFT and subsequent plotting with a custom Python script. A "full-length" ortholog was defined as
395 having at least 50% alignment with any *A. thaliana* protein in the orthogroup, excluding pseudogenes.
396 Custom Python and Bash scripts are available at https://doi.org/10.5061/dryad.0cfxpnw7p.
397 ***In silico targeting predictions***
398 For TP analysis, all protein models available for each gene with similarity to the original Orthofinder hit
399 (blastp e-value < 1e-6) were used, not just the longest. All transcripts for cytosolic aaRS, a "positive
400 control" set of genes involved in mitochondrial oxidative phosphorylation and a "negative control" set of
401 genes involved in cell-wall synthesis were run through targeting prediction programs LOCALIZER-1.0.5
402 (83) and TargetP-2.0 (84), with commands as follows:
403 LOCALIZER.py -p -i <input file>
404 targetp -fasta <input file> -mature -org pl
405 Protein models for some species were often incomplete due to the fragmented nature of some of the
406 transcriptome assemblies. To ensure models missing the N-terminus were not used in retargeting
407 analysis, predictions for models that did not align within 100 amino acids of the start codon for at least

408 one *A. thaliana* ortholog were discarded. For transcript-by-transcript heatmaps (**Supplemental Dataset**
409 **6**), a custom R script was used to plot all transcripts for all species. For plotting in meta-heatmaps (**Fig.**
410 **S2**), only the most retargeted ortholog for each species was used. Orthologs were then binned based on
411 probability of targeting to the mitochondria or plastids. High confidence retargeting was defined as a >
412 0.25 probability of targeting for both LOCALIZER and TargetP-2.0. Moderate confidence was defined as a
413 > 0.25 probability for one of the algorithms. Low confidence retargeting was defined as neither
414 algorithm providing a probability > 0.25. For subsequent analysis of ratios (**Fig. 2, Fig. S1**), any ortholog
415 with a "moderate" or better confidence was counted as retargeted. Note that *R. cantleyi* and *S.*
416 *himalayana* proteins were excluded from predictions of targeting to plastids as these species likely lack
417 plastomes. All analyses including plotting and statistics were carried out using custom Bash, Python, and
418 R scripts available at https://doi.org/10.5061/dryad.6hdr7sr5x.

419 ***In-vivo localization of S. himalayana aaRS transit peptides via tobacco infiltration and confocal***
420 ***microscopy***

421 TPs were chosen based on having a moderate probability (see above) of targeting a protein to either the
422 mitochondria or plastid, and/or having an N-terminal extension preceding conserved sequence shared
423 with *A. thaliana.* We selected either the sequence predicted to be a TP by TargetP-2.0 plus 10 residues
424 or the entire N-terminal overhang plus the first 10 residues aligning with the functional portion of the *A.*
425 *thaliana* protein. These amino acid sequences were then converted into DNA sequence using EMBOSS
426 Backtranseq (85) with the *Nicotiana tabacum* codon usage table (see **Supplemental Dataset 3**). These
427 sequences were then used to generate constructs for *Agrobacterium* transformation of *N. benthamiana*
428 and confocal microscopy as described in (59). Constructs are available from Addgene (ID numbers
429 202646-202653).

430 ***Mitogenome assembly***

431 Prior to tRNA analysis, the mitogenomes of *S. himalayana* (SRR629601) and *R. cantleyi* (SRR629613)
432 were assembled *de novo* from raw reads available from the NCBI SRA database. To maximize the
433 likelihood of finding all mitogenome tRNAs, genomes were assembled in several ways. Reads were first
434 trimmed using FASTP (86), with either default parameters or "harsh" trimming where the --
435 qualified_quality_phred and --unqualified_percent_limit parameters were set to 20 and 10, respectively.
436 SPAdes v.3.15.4 (87) was then used to assemble trimmed reads *de novo,* with kmer sizes of 21, 33, 55,
437 and 77. To avoid potential assembly fragmentation due to excessive levels of sequence coverage, a
438 subset of reads (1 million from S. *himalayana* and 500,000 from *R. cantleyi*) were sampled and
439 assembled from each species as well.

440 ***tRNA analysis***

441 Excepting the Rafflesiaceae, for which we assembled mitogenomes *de novo,* the mitogenomes,
442 plastomes and accompanying GenBank (.gbk) annotation files for species of interest were downloaded
443 from NCBI (**Supplemental Dataset 1**). A custom Perl script was used to extract annotated tRNAs from
444 the .gbk files. Concurrently, tRNAscan-SE-2.0.9 (88, 89) was run on downloaded genomes (parameters -
445 BQH), and results were cross-checked with preexisting annotations from GenBank annotations
446 (**Supplemental Dataset 4**). For *S. himalayana* and *R. cantleyi*, the same tRNAscan-SE-2.0.9 pipeline was
447 used on all generated assemblies, and the consensus set of tRNAs was checked for plant origin using
448 blastn against all NCBI reference sequences and the *A. thaliana* mitogenome (NC_037304.1) and
449 plastome (NC_000932.1). tRNA presence/absence patterns were plotted along with the corresponding
450 aaRS presence/absence/retargeting data using a custom Python script. Scripts are available at
451 https://doi.org/10.5061/dryad.np5hqc009.

452 ***Rate Analysis***

453 Primary protein sequences for cytosolic and organellar aaRS from heterotroph *M. hypopitys* and related
454 autotroph *R. williamsianum* from the original Orthofinder search were used for comparison of rates,

455 with *A. thaliana* orthologs forming the outgroup. CysRS, LeuRS, ThrRS, and TrpRS were excluded from
456 the analysis because the organellar ortholog was missing or too short to create informative gene trees in
457 one or both of the test species. Predicted TPs were trimmed from sequences using output of TargetP
458 2.0. Sequences were then aligned using MAFFT v7.490 (80) and trimmed manually in MEGA 11 (81).
459 Maximum likelihood trees were built using RaXML v 8.2.12 (82) as follows:

460 raxmlHPC-PTHREADS-SSE3 -f a -# 100 -p 12345 -x 12345 -s <orthogroup_aligned.fa> -n <orthogroup> -m
461 PROTGAMMALG -T 12

462 For orthogroups containing only three ortholog sequences, a "dummy" outgroup sequence represented
463 by the *A. thaliana* ortholog with one branch-specific substitution was created to overcome RaXML's
464 input requirement of four or more taxa. The resulting trees were then evaluated for topology. In
465 instances where species-specific paralogs were not sister to one another, separate multiple sequence
466 alignments and trees were re-made for each paralog as described above. In instances where species-
467 specific paralogs were sister, both sequences were retained in the same tree. Trees were rooted to the
468 *A. thaliana* ortholog(s), then branch lengths to the last shared node from each *H. monotropa* and *R.*
469 *williamsianum* ortholog were extracted from the trees. For instances where two sister paralogs from the
470 same species were retained, the branch lengths were averaged. Then, the $\log_{10}$-transformed ratio of the
471 *H. monotropa/R williamsianum* branches was plotted, and mean branch length for cytosolic and
472 organellar enzymes was calculated. Statistical difference in mean branch ratio in cytosolic vs. organellar
473 enzymes was tested using a *t*-test. Tajima's relative rate tests were run in MEGA 11 using *A. thaliana* as
474 the outgroup. Data are available in **Supplemental Dataset 5.** Scripts, alignments, and maximum
475 likelihood trees are available at https://doi.org/10.5061/dryad.tb2rbp05r.

476

477 ***Acknowledgements***

478 We thank Jessica Warren for helpful discussion and feedback on this project. This work was supported
479 by grants MCB-20484, MCB-2048407, and MCB-1933590 from the National Science Foundation (NSF)
480 and utilized the RMACC Summit supercomputer (NSF ACI-1532235 and ACI-1532236). RAD was
481 supported by an NSF Postdoctoral Research Fellowship in Biology (IOS-2208908).

482

483 ***References***

484 1. A. N. Khachane, K. N. Timmis, V. A. P. Martins Dos Santos, Dynamics of reductive genome
485 evolution in mitochondria and obligate intracellular microbes. *Mol Biol Evol* **24**, 449–456 (2007).

486 2. P. J. Keeling, J. D. Palmer, Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**, 605–
487 618 (2008).

488 3. W. Martin, *et al.*, Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes
489 reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS* **99**,
490 12246–12251 (2002).

491 4. E. C. M. Nowack, *et al.*, Endosymbiotic gene transfer and transcriptional regulation of transferred
492 genes in Paulinella chromatophora. *Mol Biol Evol* **28**, 407–422 (2011).

493 5. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial
494 symbionts. *Annu Rev Genet* **42**, 165–190 (2008).

495 6. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat Rev*
496 *Microbiol* **10**, 13–26 (2012).

497 7. V. Rajendran, P. Kalita, H. Shukla, A. Kumar, T. Tripathi, Aminoacyl-tRNA synthetases: Structure,
498 function, and drug discovery. *Int J Biol Macromol* **111**, 400–414 (2018).

499 8. T. Salinas-Giegé, R. Giegé, P. Giegé, TRNA biology in mitochondria. *Int J Mol Sci* **16**, 4518–4559
500 (2015).

501 9. A. M. Duchêne, C. Pujol, L. Maréchal-Drouard, Import of tRNAs and aminoacyl-tRNA synthetases
502 into mitochondria. *Curr Genet* **55**, 1–18 (2009).

503 10. L. Bonnefond, *et al.*, Toward the full set of human mitochondrial aminoacyl-tRNA synthetases:
504 Characterization of AspRS and TyrRS. *Biochemistry* **44**, 4805–4816 (2005).

505 11. E. Tolkunova, H. Park, J. Xia, M. P. King, E. Davidson, The human lysyl-tRNA synthetase gene
506 encodes both the cytoplasmic and mitochondrial enzymes by means of an unusual: Alternative
507 splicing of the primary transcript. *Journal of Biological Chemistry* **275**, 35063–35069 (2000).

508 12. S. Shikha, R. Brogli, A. Schneider, N. Polacek, TRNA biology in trypanosomes. *Chimia (Aarau)* **73**,
509 395–405 (2019).

510 13. T. H. P. Tan, R. Pach, A. Crausaz, A. Ivens, A. Schneider, tRNAs in Trypanosoma brucei : Genomic
511 Organization, Expression, and Mitochondrial Import . *Mol Cell Biol* **22**, 3707–3717 (2002).

512 14. J. M. Warren, *et al.*, Rapid Shifts in Mitochondrial tRNA Import in a Plant Lineage with Extensive
513 Mitochondrial tRNA Gene Loss. *Mol Biol Evol* **38**, 5735–5751 (2021).

514 15. J. M. Warren, D. B. Sloan, Interchangeable parts: The evolutionarily dynamic tRNA population in
515 plant mitochondria. *Mitochondrion* **52**, 144–156 (2020).

516 16. A.-M. Duchêne, *et al.*, Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in
517 Arabidopsis thaliana. *PNAS* **102**, 16484–16489 (2005).

518 17. K. Fujishima, A. Kanai, tRNA gene diversity in the three domains of life. *Front Genet* **5** (2014).

519 18. M. Lynch, B. Koskella, S. Schaack, Mutation Pressure and the Evolution of Organelle Genomic
520 Architecture. *Science (1979)* **311**, 1727–1730 (2006).

521 19. S. Schaack, E. K. H. Ho, F. MacRae, Disentangling the intertwined roles of mutation, selection and
522 drift in the mitochondrial genome. *Philosophical Transactions of the Royal Society B: Biological*
523 *Sciences* **375** (2020).

524 20. K. Crosby, D. R. Smith, Does the Mode of Plastid Inheritance Influence Plastid Genome
525 Architecture? *PLoS One* **7** (2012).

526 21. K. H. Wolfe, W.-H. Li, P. M. Sharp, "Rates of nucleotide substitution vary greatly among plant
527 mitochondrial, chloroplast, and nuclear DNAs (plant molecular evolution/molecular
528 clock/mutation rate/organelle DNA/inverted repeat)" (1987).

529 22. J. R. Adrion, P. S. White, K. L. Montooth, The roles of compensatory evolution and constraint in
530 aminoacyl trna synthetase evolution. *Mol Biol Evol* **33**, 152–161 (2016).

531 23. F. S. Barreto, *et al.*, Genomic signatures of mitonuclear coevolution across populations of
532 Tigriopus californicus. *Nat Ecol Evol* **2**, 1250–1257 (2018).

533 24. W. Pett, D. V Lavrov, Cytonuclear Interactions in the Evolution of Animal Mitochondrial tRNA
534 Metabolism. *Genome Biol Evol* **7**, 2089–2101 (2015).

535 25. D. B. Sloan, R. A. DeTar, J. M. Warren, Aminoacyl-tRNA synthetase evolution within the dynamic
536 tripartite translation system of plant cells. *Genome Biol Evol* **15** (2023).

537 26. S. Wicke, G. M. Schneeweiss, C. W. dePamphilis, K. F. Müller, D. Quandt, The evolution of the
538 plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol* **76**,
539 273–297 (2011).

540 27. M. Michaud, V. Cognat, A. M. Duchêne, L. Maréchal-Drouard, A global picture of tRNA genes in
541 plant genomes. *Plant Journal* **66**, 80–93 (2011).

542 28. E. S. Forsythe, *et al.*, Organellar transcripts dominate the cellular mRNA pool across plants of
543 varying ploidy levels. *PNAS* **119** (2022).

544 29. J. Mergner, *et al.*, Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* **579**, 409–
545 414 (2020).

546 30. J. Těšitel, Functional biology of parasitic plants: A review. *Plant Ecol Evol* **149**, 5–20 (2016).

547 31. L. Teixeira-Costa, C. C. Davis, Life history, diversity, and distribution in parasitic flowering plants.
548 *Plant Physiol* **187**, 32–51 (2021).

549 32. S. Wicke, J. Naumann, "Molecular Evolution of Plastid Genomes in Parasitic Flowering Plants" in
550 *Advances in Botanical Research*, (Academic Press Inc., 2018), pp. 315–347.

551 33. D. R. Smith, S. R. Asmail, Next-generation sequencing data suggest that certain
552 nonphotosynthetic green plants have lost their plastid genomes. *New Phytologist* **204**, 7–11
553 (2014).

554 34. H. T. Li, *et al.*, Plastid phylogenomic insights into relationships of all flowering plant families. *BMC*
555 *Biol* **19** (2021).

556 35. J. Molina, *et al.*, Possible loss of the chloroplast genome in the parasitic flowering plant Rafflesia
557 lagascae (Rafflesiaceae). *Mol Biol Evol* **31**, 793–803 (2014).

558 36. D. L. Nickrent, Y. Ouyang, R. J. Duff, C. W. Depamphilis, Do nonasterid holoparasitic flowering
559 plants have plastid genomes? *Plant Mol Biol* **34**, 717–729 (1997).

560 37. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics.
561 *Genome Biol* **20** (2019).

562 38. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons
563 dramatically improves orthogroup inference accuracy. *Genome Biol* **16** (2015).

564 39. G. Canino, *et al.*, Arabidopsis encodes four tRNase Z enzymes. *Plant Physiol* **150**, 1494–1502
565 (2009).

566 40. B. Gutmann, A. Gobert, P. Giegé, PRORP proteins support RNase P activity in both organelles and
567 the nucleus in Arabidopsis. *Genes Dev* **26**, 1022–1027 (2012).

568 41. A. Gobert, *et al.*, A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*
569 **17**, 740–744 (2010).

570 42. C. Pujol, *et al.*, Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and
571 chloroplastic Gln-tRNAGln synthesis in plants. *Proc Natl Acad Sci U S A* **105**, 6481–6485 (2008).

572 43. A. Nagao, T. Suzuki, T. Katoh, Y. Sakaguchi, T. Suzuki, Biogenesis of glutaminyl-mt tRNA Gln in
573 human mitochondria. *PNAS* **106**, 16209–16214 (2009).

574 44. L. Echevarría, *et al.*, Glutamyl-tRNAGln amidotransferase is essential for mammalian
575 mitochondrial translation in vivo. *Biochemical Journal* **460**, 91–101 (2014).

576 45. R. Giegé, G. Eriani, The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Res*
577 **51**, 1528–1570 (2023).

578 46. L. F. Ceriotti, M. E. Roulet, M. V. Sanchez-Puerta, Plastomes in the holoparasitic family
579 Balanophoraceae: Extremely high AT content, severe gene content reduction, and two
580 independent genetic code changes. *Mol Phylogenet Evol* **162** (2021).

581 47. H. J. Su, *et al.*, Novel genetic code and record-setting AT-richness in the highly reduced plastid
582 genome of the holoparasitic plant Balanophora. *Proc Natl Acad Sci U S A* **116**, 934–943 (2019).

583 48. M. I. Schelkunov, M. S. Nuraliev, M. D. Logacheva, Rhopalocnemis phalloides has one of the most
584 reduced and mutated plastid genomes known. *PeerJ* **2019** (2019).

585 49. X. Chen, *et al.*, Comparative Plastome Analysis of Root-and Stem-Feeding Parasites of Santalales
586 Untangle the Footprints of Feeding Mode and Lifestyle Transitions. *Genome Biol Evol* **12**, 3663–
587 3676 (2019).

588 50. R. Tanaka, K. Kobayashi, T. Masuda, Tetrapyrrole Metabolism in Arabidopsis thaliana. *Arabidopsis*
589 *Book* (2011) https:/doi.org/10.1199/tab.0145.

590 51. K. Kobayashi, T. Masuda, Transcriptional regulation of tetrapyrrole biosynthesis in arabidopsis
591 thaliana. *Front Plant Sci* **7** (2016).

592 52. M. Kayama, *et al.*, Highly Reduced Plastid Genomes of the Non-photosynthetic
593 Dictyochophyceans Pteridomonas spp. (Ochrophyta, SAR) Are Retained for tRNA-Glu-Based
594 Organellar Heme Biosynthesis. *Front Plant Sci* **11** (2020).

595 53. A. C. Barbrook, C. J. Howe, S. Purton, Why are plastid genomes retained in non-photosynthetic
596 organisms? *Trends Plant Sci* **11**, 101–108 (2006).

597 54. S. M. Ng, *et al.*, Comparative analysis of nucleus-encoded plastid-targeting proteins in Rafflesia
598 cantleyi against photosynthetic and non-photosynthetic representatives reveals orthologous
599 systems with potentially divergent functions. *Sci Rep* **8** (2018).

600 55. L. Cai, *et al.*, Deeply Altered Genome Architecture in the Endoparasitic Flowering Plant Sapria
601 himalayana Griff. (Rafflesiaceae). *Current Biology* **31**, 1002-1011.e9 (2021).

602 56. M. D. Logacheva, M. I. Schelkunov, V. Y. Shtratnikova, M. V. Matveeva, A. A. Penin, Comparative
603 analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives.
604 *Sci Rep* **6** (2016).

605 57. M. I. Schelkunov, M. S. Nuraliev, M. D. Logacheva, Genomic comparison of non-photosynthetic
606 plants from the family Balanophoraceae with their photosynthetic relatives. *PeerJ* **9** (2021).

607 58. L. F. Ceriotti, L. Gatica-Soria, M. V. Sanchez-Puerta, Cytonuclear coevolution in a holoparasitic
608 plant with highly disparate organellar genomes. *Plant Mol Biol* **109**, 673–688 (2022).

609 59. J. M. Warren, *et al.*, Rewiring of aminoacyl-tRNA synthetase localization and interactions in plants
610 with extensive mitochondrial tRNA gene loss. *Mol Biol Evol* **40** (2023).

611 60. M. W. Murcha, *et al.*, Protein import into plant mitochondria: Signals, machinery, processing, and
612 regulation. *J Exp Bot* **65**, 6301–6335 (2014).

613 61. B. Bölter, En route into chloroplasts: preproteins' way home. *Photosynth Res* **138**, 263–275
614 (2018).

615 62. P. Chotewutmontri, K. Holbrook, B. D. Bruce, "Plastid Protein Targeting: Preprotein Recognition
616 and Translocation" in *International Review of Cell and Molecular Biology*, (Elsevier Inc., 2017), pp.
617 227–294.

618 63. R. W. Christian, S. L. Hewitt, G. Nelson, E. H. Roalson, A. Dhingra, Plastid transit peptides-where
619 do they come from and where do they all belong? Multi-genome and pan-genomic assessment of
620 chloroplast transit peptide evolution. *PeerJ* **8** (2020).

621 64. S. Reinbothe, *et al.*, Trna-dependent import of a transit sequence-less aminoacyl-trna synthetase
622 (Leurs2) into the mitochondria of arabidopsis. *Int J Mol Sci* **22** (2021).

623 65. M. W. Murcha, *et al.*, Plant-specific preprotein and amino acid transporter proteins are required
624 for tRNA import into mitochondria. *Plant Physiol* **172**, 2471–2490 (2016).

625 66. E. Skippington, T. J. Barkman, D. W. Rice, J. D. Palmer, Miniaturized mitogenome of the parasitic
626 plant viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. *PNAS*
627 **112**, E3515–E3524 (2015).

628 67. G. Petersen, A. Cuenca, I. M. Møller, O. Seberg, Massive gene loss in mistletoe (Viscum,
629 Viscaceae) mitochondria. *Sci Rep* **5** (2015).

630 68. A. Fender, *et al.*, Loss of a primordial identity element for a mammalian mitochondrial
631 aminoacylation system. *Journal of Biological Chemistry* **281**, 15980–15986 (2006).

632 69. Y. Kumazawa, H. Himeno, K.-I. Miura, K. Watanabe, Unilateral Aminoacylation Specificity
633 between Bovine Mitochondria and Eubacteria. *The Journal of Biochemistry* **109**, 421–427 (1991).

634 70. L. Bonnefond, M. Frugier, R. Giegé, J. Rudinger-Thirion, Human mitochondrial TyrRS disobeys the
635 tyrosine identity rules. *RNA* **11**, 558–562 (2005).

636 71. L. Bromham, P. F. Cowman, R. Lanfear, Parasitic plants have increased rates of molecular
637 evolution across all three genomes. *BMC Evol Biol* **13** (2013).

638 72. M. Safro, N. Moor, O. Lavrik, Phenylalanyl-tRNA Synthetases. *Madame Curie Bioscience Database*
639 (2013) (May 24, 2023).

640 73. K. M. Haen, W. Pett, D. V. Lavrov, Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA
641 synthetases and mtDNA-encoded tRNAs in cnidaria. *Mol Biol Evol* **27**, 2216–2219 (2010).

642 74. I. Martin, O. Vinas, T. Mampel, F. Villarroya, Effects of cold environment on mitochondrial
643 genome expression in the rat: evidence for a tissue-specific increase in the liver, independent of
644 changes in mitochondrial gene abundance. *Biochemical Journal* **296**, 231–234 (1993).

645 75. R. F. Hevner, M. T. T. Wong-Riley, Mitochondrial and Nuclear Gene Expression for Cytochrome
646 Oxidase Subunits Are Disproportionately Regulated by Functional Activity in Neurons. *The Journal*
647 *of Neuroscience* **13**, 1805–1819 (1993).

648 76. E. A. Amiott, J. A. Jaehning, Mitochondrial Transcription Is Regulated via an ATP "Sensing"
649 Mechanism that Couples RNA Abundance to Respiration. *Mol Cell* **22**, 329–338 (2006).

650 77. B. J. Haas, Transdecoder v5.5.0 (2018) (May 14, 2023).

651 78. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file
652 manipulation. *PLoS One* **11** (2016).

653 79. C. Camacho, *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10** (2009).

654 80. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence
655 alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).

656 81. K. Tamura, G. Stecher, S. Kumar, MEGA11: Molecular Evolutionary Genetics Analysis Version 11.
657 *Mol Biol Evol* **38**, 3022–3027 (2021).

658 82. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large
659 phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

660 83. J. Sperschneider, *et al.*, LOCALIZER: Subcellular localization prediction of both plant and effector
661 proteins in the plant cell. *Sci Rep* **7** (2017).

662 84. J. J. A. Armenteros, *et al.*, Detecting sequence signals in targeting peptides using deep learning.
663 *Life Sci Alliance* **2** (2019).

664 85. F. Madeira, *et al.*, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic*
665 *Acids Res* **50**, W276–W279 (2022).

666 86. S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor in
667 *Bioinformatics*, (Oxford University Press, 2018), pp. i884–i890.

668 87. A. Bankevich, *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell
669 sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).

670 88. P. P. Chan, T. M. Lowe, "tRNAscan-SE: Searching for tRNA genes in genomic sequences" in
671 *Methods in Molecular Biology*, (Humana Press Inc., 2019), pp. 1–14.

672 89. P. P. Chan, B. Y. Lin, A. J. Mak, T. M. Lowe, TRNAscan-SE 2.0: Improved detection and functional
673 classification of transfer RNA genes. *Nucleic Acids Res* **49**, 9077–9096 (2021).

Supplemental Dataset 1: Species and datasets used for gene presence/absence, Species and datasets used for retargeting analysis, Species and datasets used for tRNA content analysis.

Supplemental Dataset 2. Genes of interest and counts of gene copies for each species.

Supplemental Dataset 3: *S. himalayana* transit peptide sequences for transient expression and colocalization in *N. benthamiana*

Supplemental Dataset 4: tRNA content of organelle genomes.

Supplemental Dataset 5: Relative branch lengths and Tajima's relative rate test for aaRS in Ericales spp.

Supplemental Dataset 6: Heatmaps for probability of subcellular targeting to plastids or mitochondria on a protein-by-protein basis.

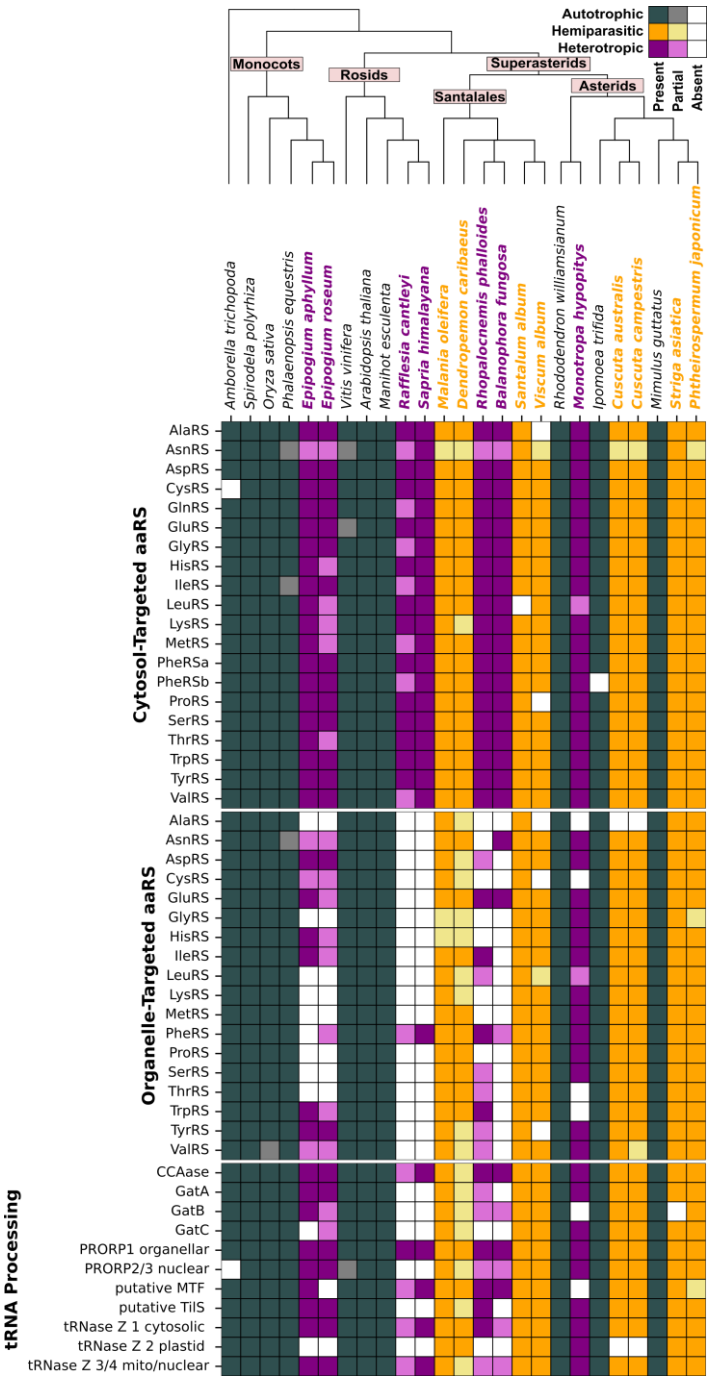


Figure 1: Presence/absence matrix of selected organellar and cytosolic aaRS and tRNA-processing enzymes shows major losses of organellar enzymes in heterotrophic lineages. Hue of boxes indicates metabolism (gray = autotroph, orange = hemiparasite, purple = heterotroph), shade indicates degree of presence/absence (white = no sequence found, light colors = partial sequence found, dark colors = complete or near complete sequence found). Note that there is no organellar GlnRS as plants typically use the bacterial-type GatCAB amidotransferase system to charge organellar Gln-tRNA. ArgRS is not presented as it has no distinct organellar enzyme; gene products are either cytosol-targeted, or triple-targeted to the cytosol, mitochondria, and plastids.

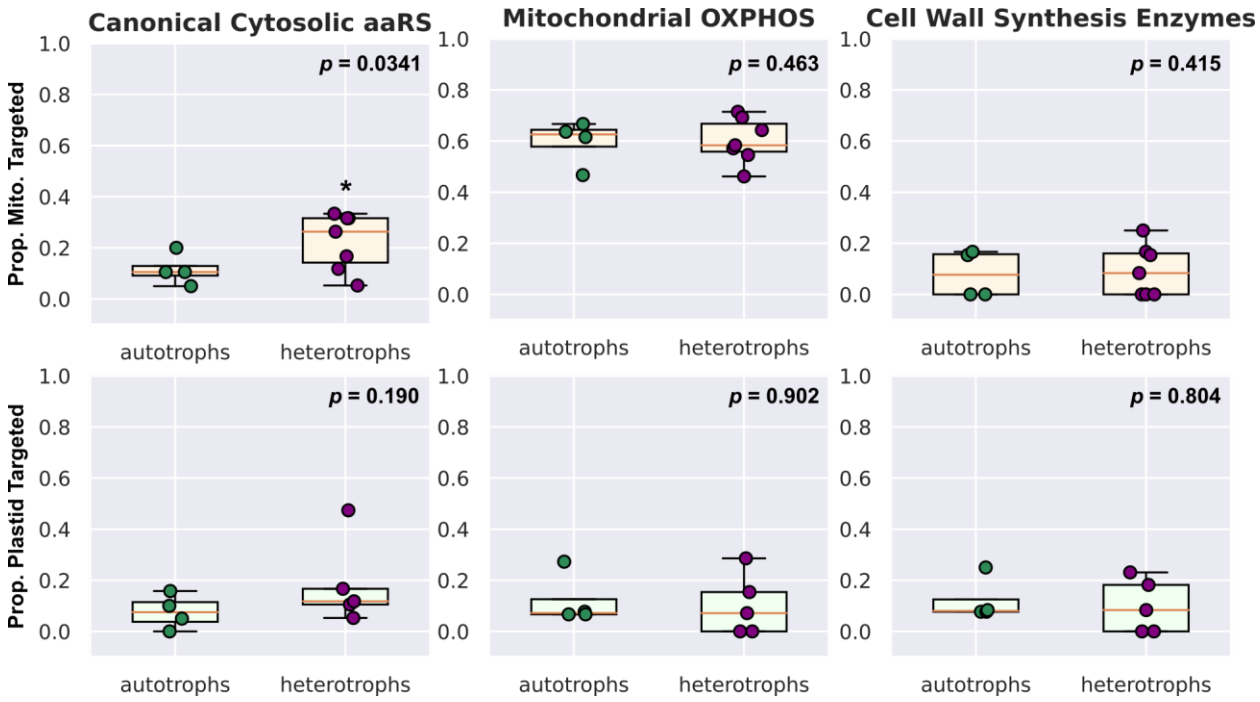


Figure 2: Proportion of retargeted cytosolic aaRS is higher in heterotrophs than autotrophs based on *in-silico* predictions. Proportion of enzymes with predicted retargeting (probability ≥ 0.25) to either mitochondria (top row) or plastids (bottom row) based on the outputs of LOCALIZER and/or TargetP. Autotrophs and heterotrophs are plotted in green and purple, respectively. Targeting predictions for proteins involved in mitochondrial oxidative phosphorylation (OPXHOS) are included as a positive control for mitochondrial targeting. Predictions for cell-wall synthesis associated enzymes are included as a negative control for organelle targeting. Reported *p*-values from *t*-tests or Mann-Whitney U tests depending on distribution of the data. Note that *R. cantleyi* and *S. himalayana* are not included in plastid retargeting analysis due to suspected loss of the plastome.

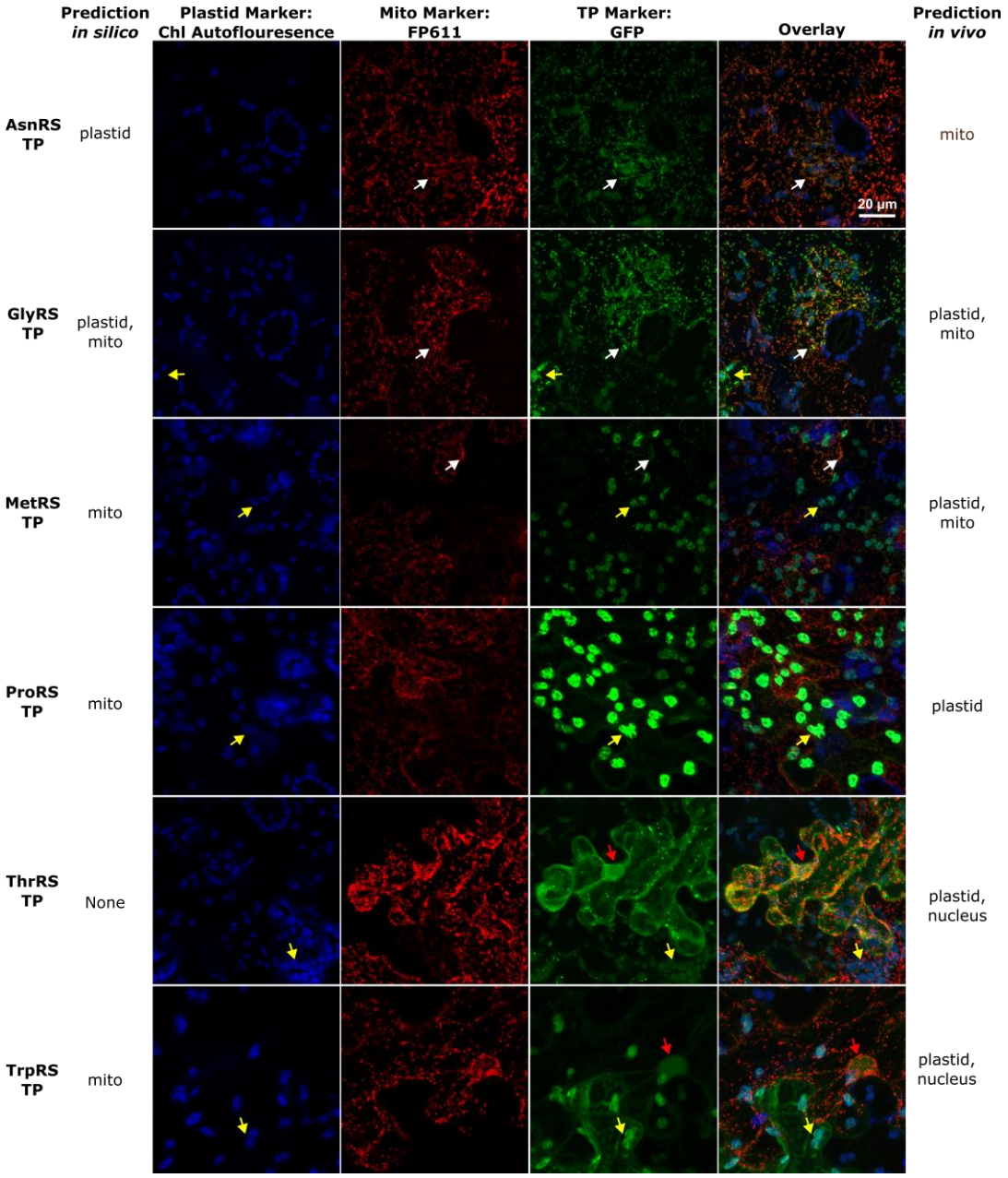


Figure 3: Cytosolic-type aaRS transit peptides from heterotroph *S. himalayana* can retarget enzymes to organelles in *Nicotiana benthamiana*. Transit peptides (TPs) from *S. himalayana* fused to GFP (middle right column) were transiently expressed in *N. benthamiana.* Chlorophyll autoflorescence (far left column) was used for colocalization with plastids. As a positive control for mitochondrial targeting, eqFP611 was tagged with the TP from mitochondrial isovaleryl-CoA dehydrogenase (middle left). The overlay of all three channels on far right shows mitochondrial localization for GFP tagged with AsnRS, GlyRS, and MetRS TPs. Minor/weak mitochondrial localization for ProRS TP. MetRS and ProRS TPs additionally target GFP to plastids. TrpRS and ThrRS TPs target GFP to plastids, cytosol, and nucleus. Note presence of punctate structures for ThrRS TP, possibly inclusion bodies. Arrows indicate good examples of plastid (yellow) and mitochondrial (white) and potential nuclear (red) localization. Scale indicated in top right panel is the same for all images.

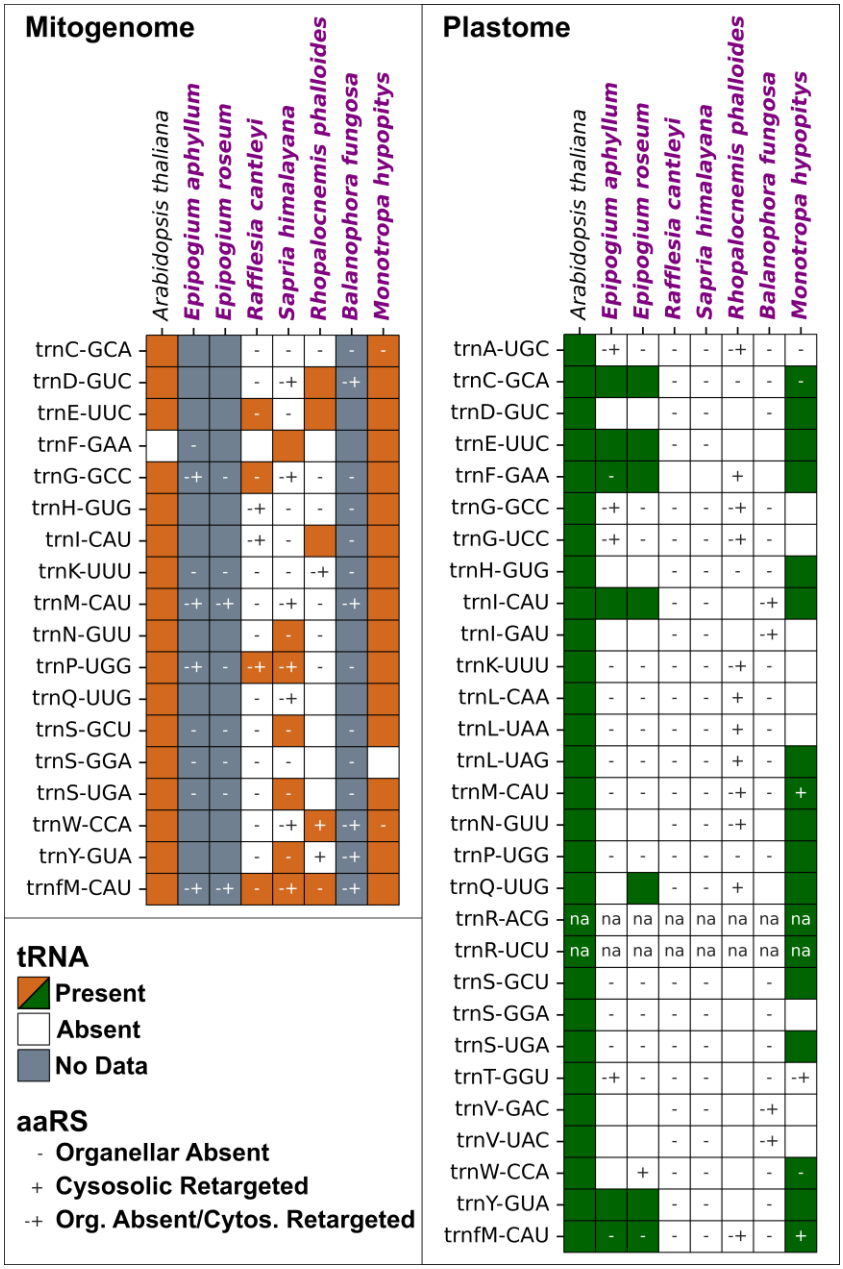
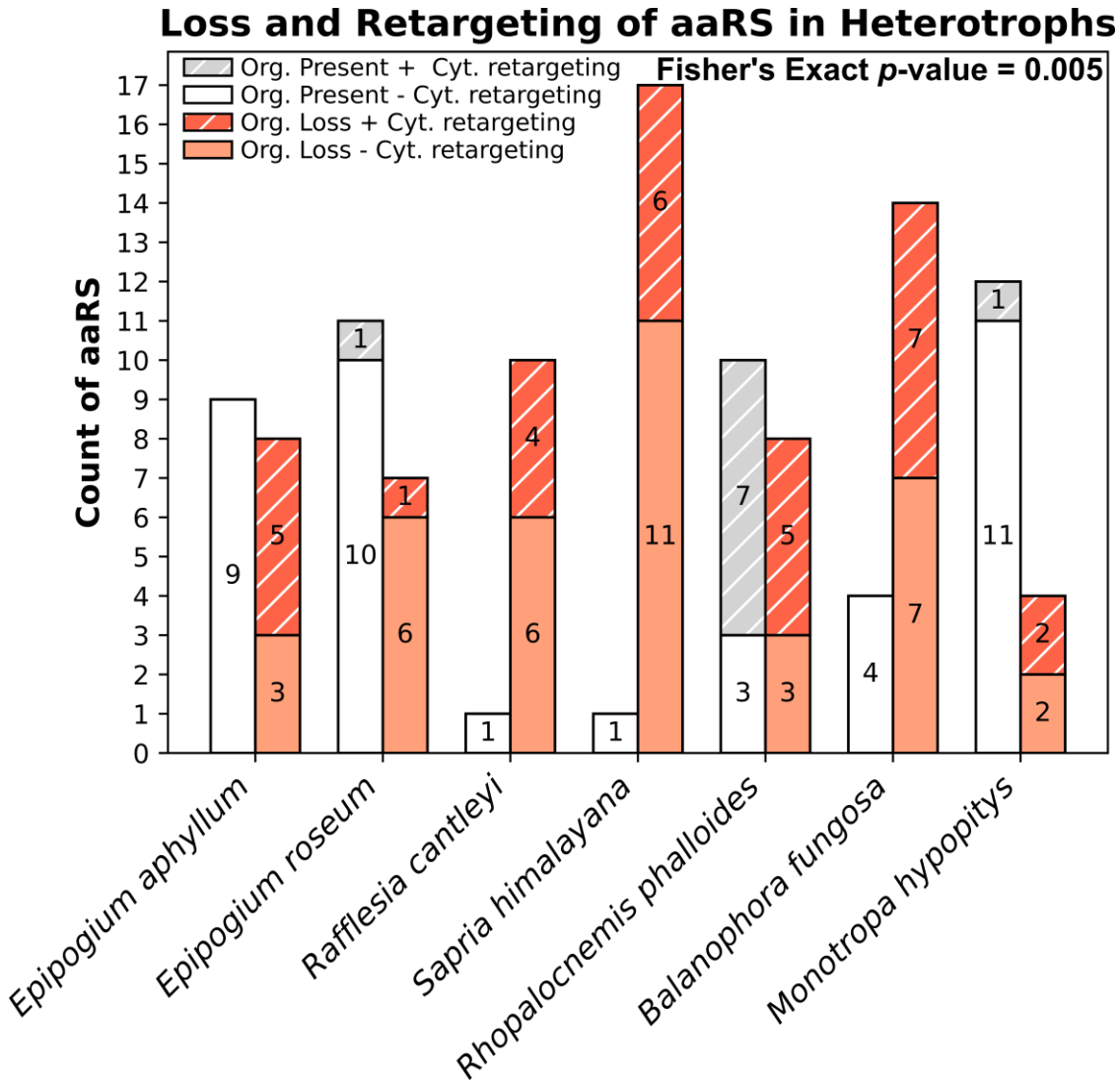
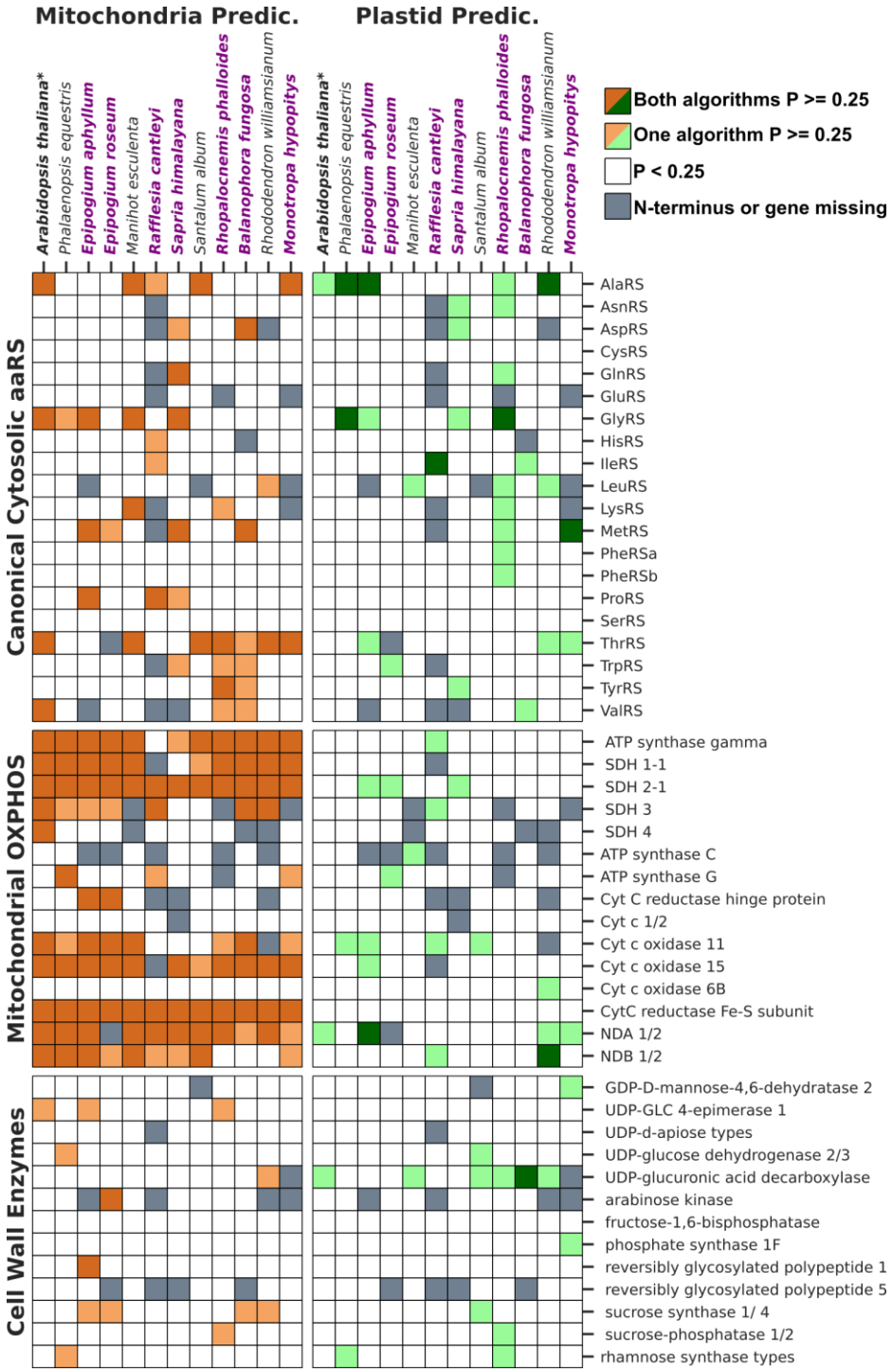


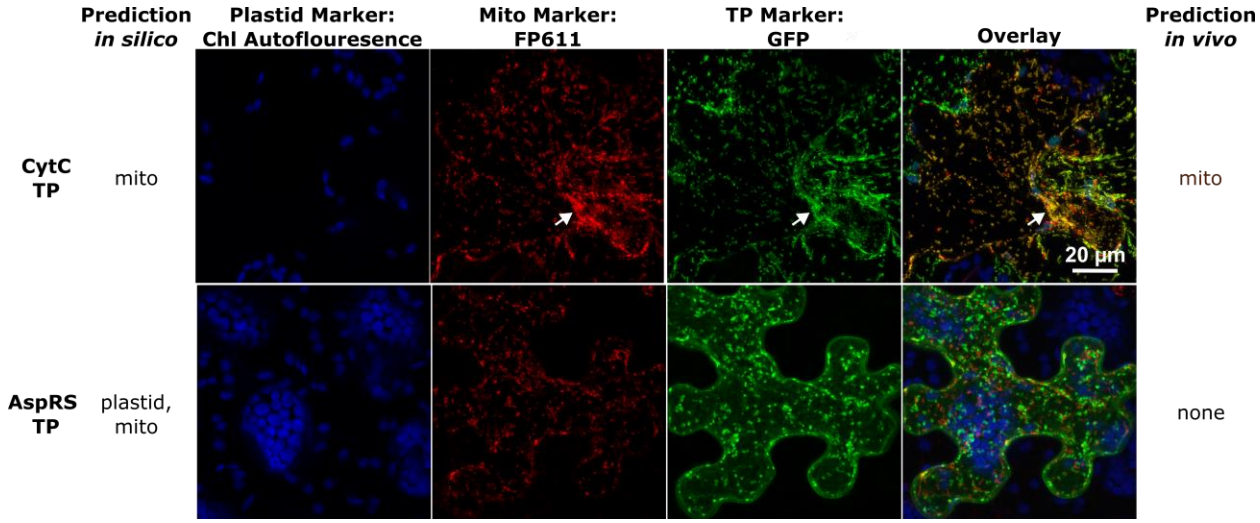
Figure 4: Instances of both concordance and mismatch between organellar tRNA gene loss and aaRS loss/retargeting. Heterotrophic species are written in purple. *Arabidopsis thaliana* is included for reference. Left: mitogenome tRNAs. Right: plastome tRNAs. Color of boxes indicates presence/absence of tRNA. The – and + annotation denotes loss of corresponding organellar aaRS and retargeting of cytosolic tRNA respectively as described for figure 4. Retargeting predictions are based on individual predictions for each organelle. Note that loss and retargeting of the corresponding aaRS was not assessed for *trnR* as there are no organelle-specific aaRS orthologs responsible for charging this tRNA. Loss of GatCAB and associated cytosolic GlnRS retargeting are presented for *trnQ*. Additionally, plastid targeting predictions for Rafflesiaceae are not included here because of the suspected loss of the plastome in this taxon.



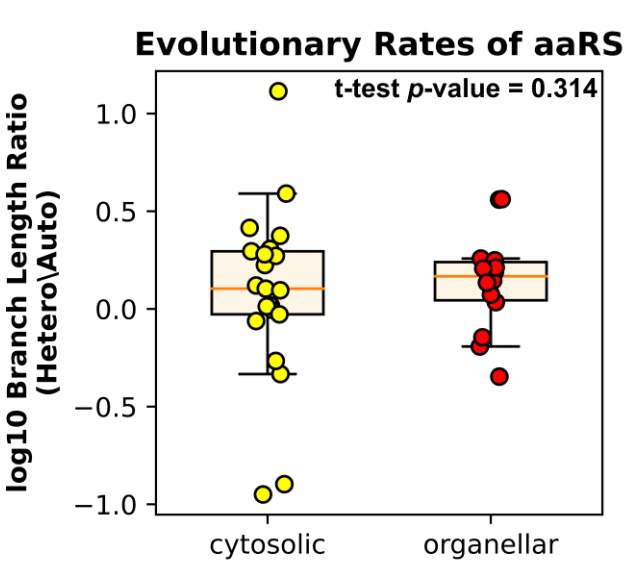
Supplemental Figure 1: Significant overlap between loss of organellar aaRS and retargeting of cytosolic counterpart in heterotrophs. Reported counts refer to number of aaRS "families" defined by amino acids (e.g., GluRS, PheRS, TrpRS, etc.). Organellar gene presence/absence determined as for figure 1. Partial sequences for organellar enzymes are counted as present. Predicted retargeting (probability ≥ 0.25) to either mitochondria or plastids was based on the outputs of LOCALIZER and/or TargetP. Note that aaRS where cytosolic targeting could not be tested (i.e., enzyme was not present, or N-terminus was incomplete) were excluded. Organellar enzyme present = shades of gray. Organellar enzyme absent = shades of orange. Cytosolic enzyme retargeting = hashing. No retargeting = solid bars. Reported *p*-value is from a Fisher's exact test for correlation between loss of organellar aaRS and retargeting of the corresponding cytosolic aaRS. Note that *R. cantleyi* and *S. himalayana* proteins are excluded from predictions of targeting to plastids as these species lack plastid genomes. Loss of GatCAB and associated cytosolic GlnRS retargeting are included in the counts.



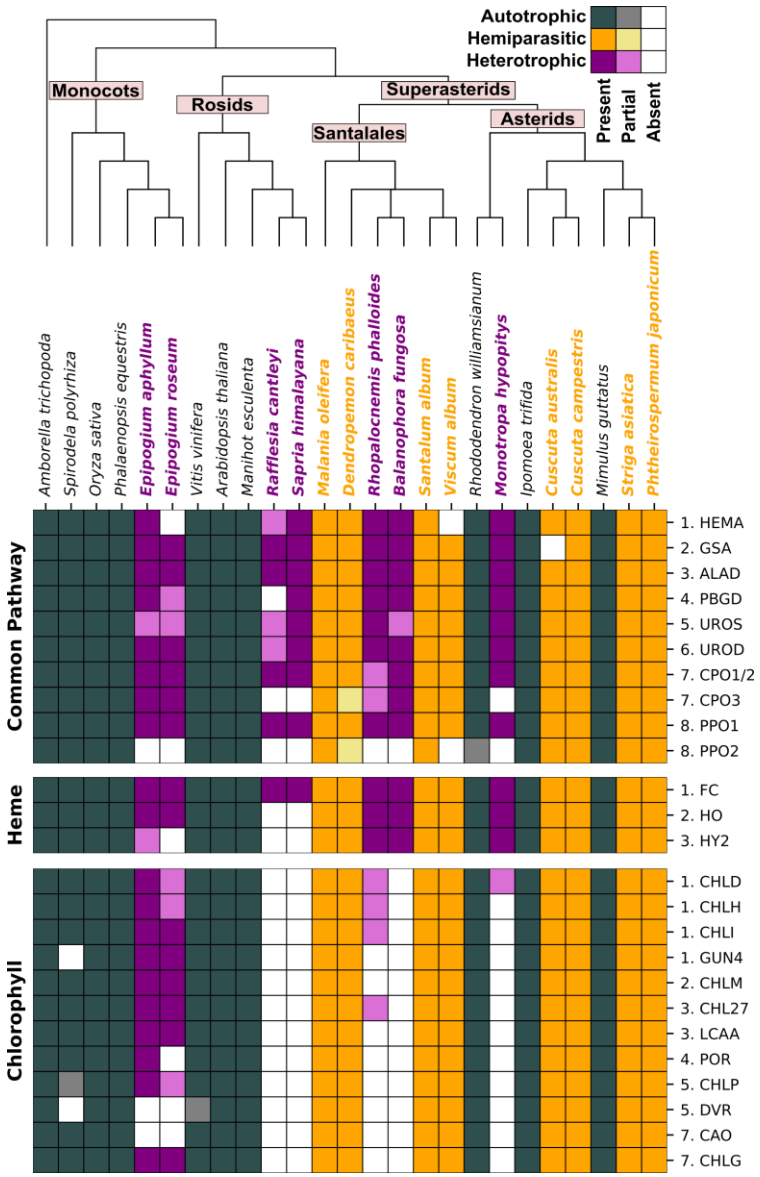
Supplemental Figure 2: Targeting predictions for cytosolic-type aaRS. Deeper shades of green (plastid, right) and orange (mitochondrion, left) indicate probability of retargeting to that organelle is ≥ 0.25 based on both LOCALIZER-1.0.5 and TargetP-2.0. Lighter shades of orange/green indicate one of the two algorithms yielded a probability ≥ 0.25. White indicates probability for both algorithms was < 0.25. Gray indicates that the gene/transcript was not present or that the N-terminus was incomplete, so no targeting prediction could be made. Heterotrophs are written in purple, closely related autotrophs in black. *A. thaliana* is included as a reference.



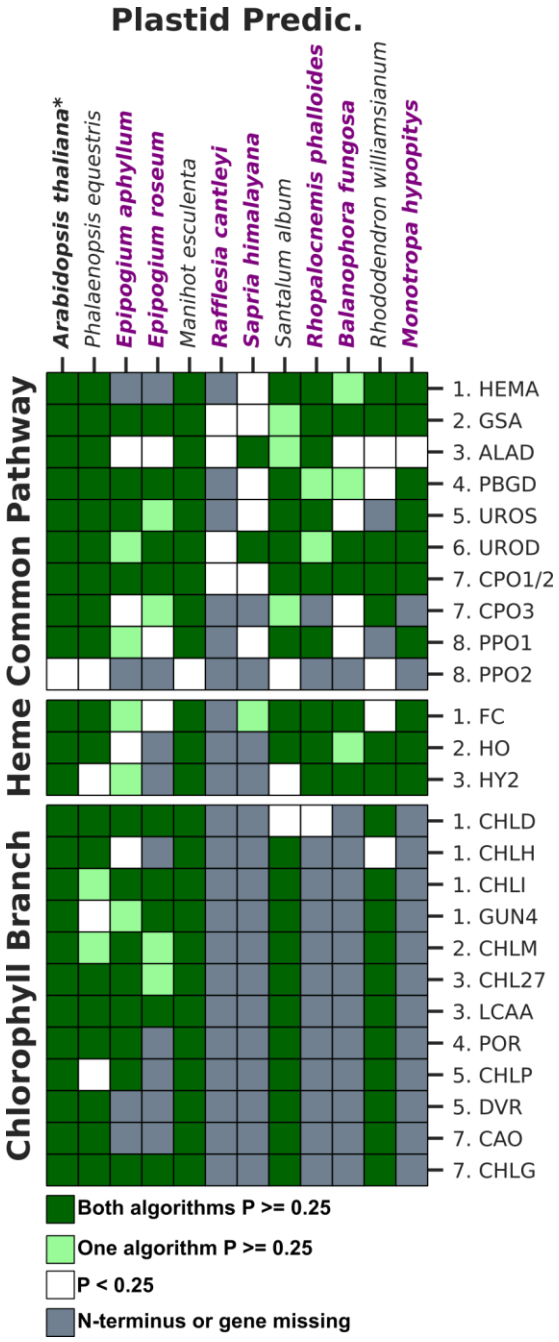
Supplemental Figure 3: Cytosolic-type aaRS transit peptides from heterotroph *S. himalayana* transiently expressed in *Nicotiana benthamiana*. Transit peptides (TPs) from *S. himalayana* fused to GFP (middle right column) were transiently expressed in *N. benthamiana.* Chlorophyll autofluorescence (far left column) was used for colocalization with plastids. As a positive control for mitochondrial targeting, eqFP611 was tagged with the TP from mitochondrial isovaleryl-CoA dehydrogenase (middle left). Overlay of all three channels on far right. Arrows indicate good examples mitochondrial (white) localization. From top down: Cytochrome C (CytC) TP was included as a positive control showing mitochondrial targeting of a *S. himalayana* TP. AspRS TP (bottom row) primarily shows GFP fluorescence from non-mitochondrial punctate structures in the cytosol, potentially inclusion bodies. Thus, fluorescence patterns in this line appear to be an expression artifact and are not informative for addressing the question of targeting. Scale indicated in top right panel is the same for all images.



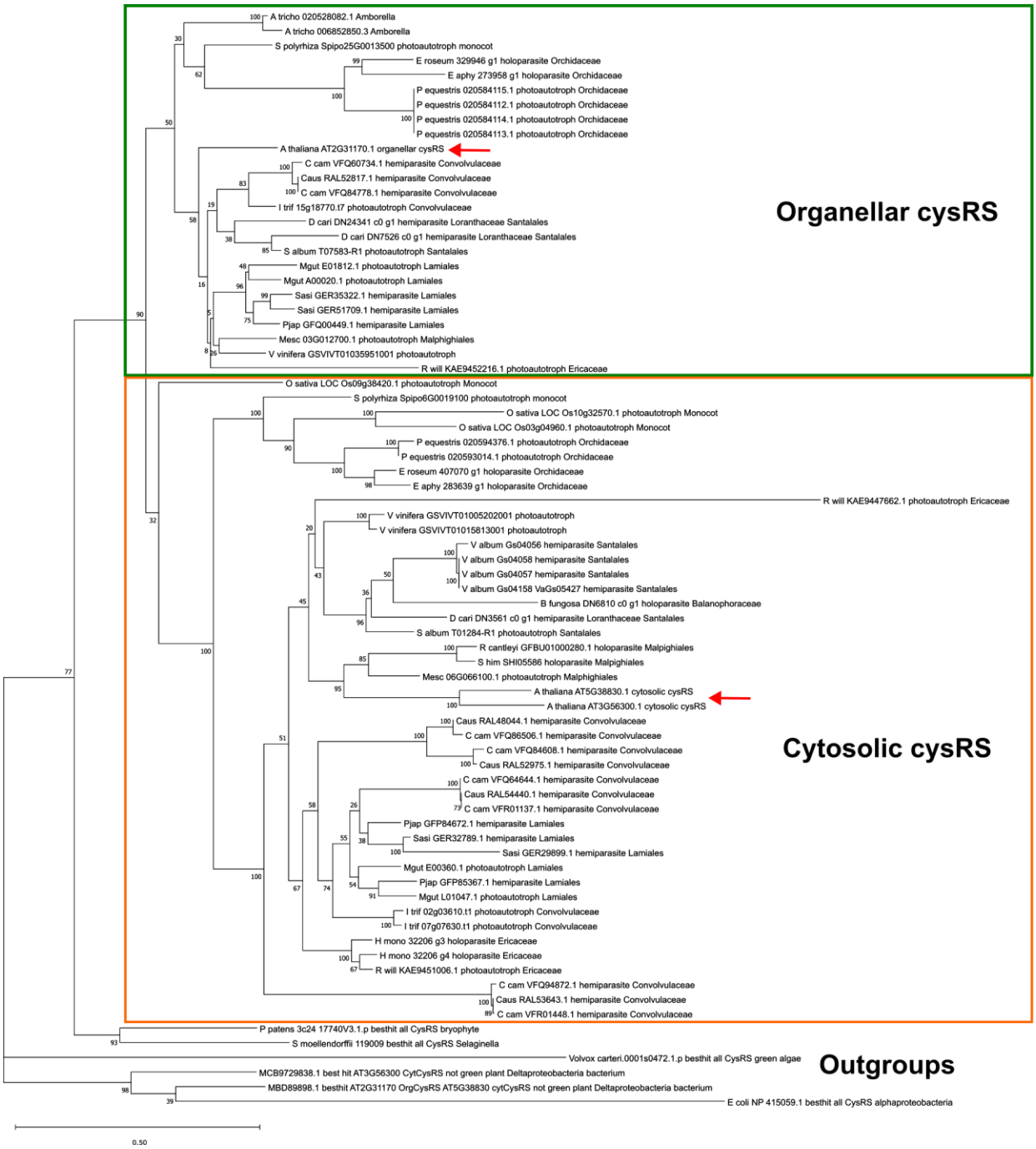
Supplemental Figure 4: Relative rates of sequence evolution of cytosolic and organellar aaRS are similar in heterotroph compared to autotrophic relative. Each point represents an aaRS branch length of *M. hypopitys* (heterotrophic) normalized to branch length of *R. williamsianum* (autotrophic) for canonically cytosolic and organellar aaRSs ($\log_{10}$ scale).



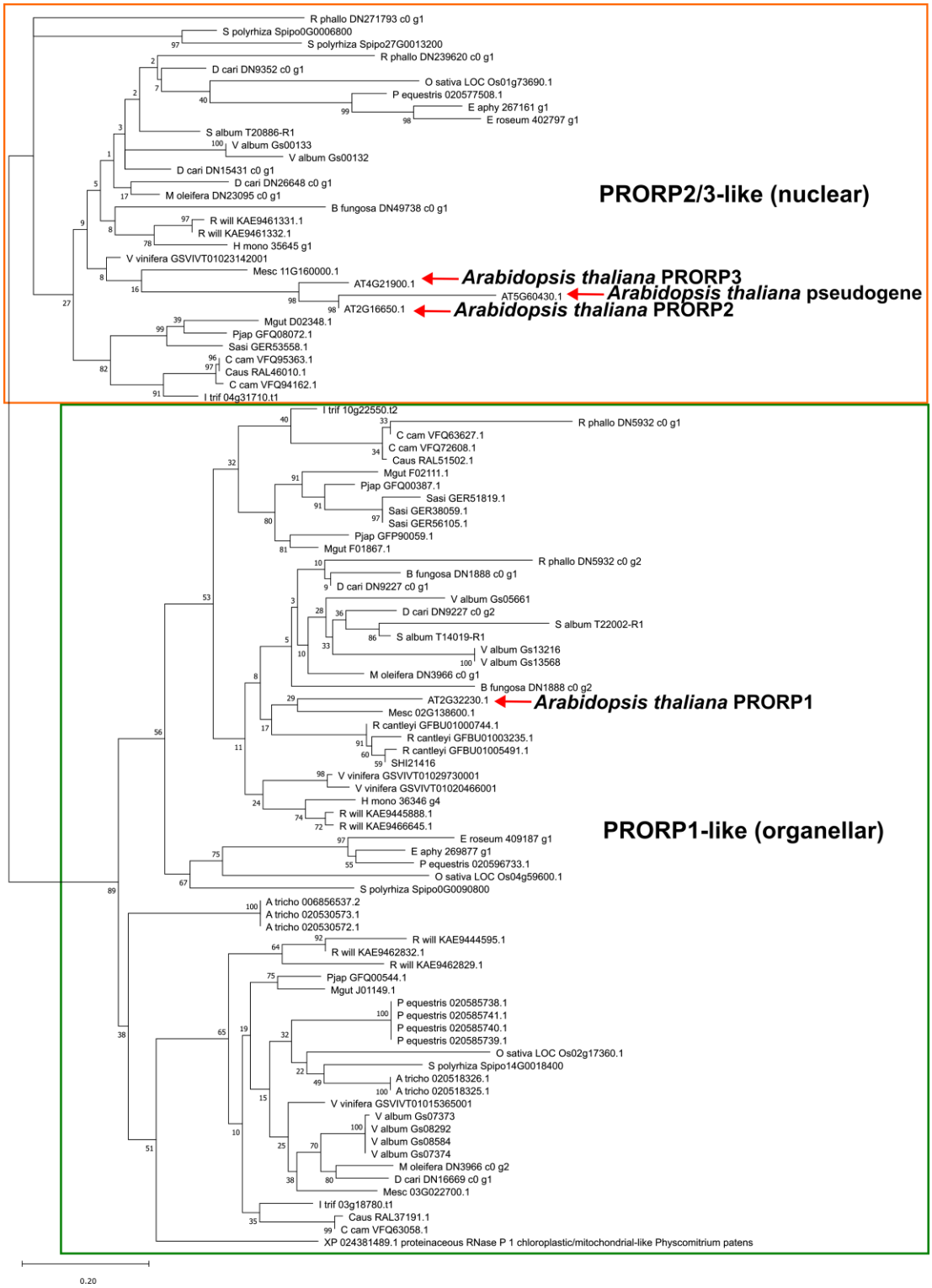
Supplemental Figure 5: Heterotrophic plants often lose chlorophyll biosynthesis enzymes yet retain heme biosynthesis enzymes in the tetrapyrrole pathway. “Common Pathway” = Enzymes for the base pathway for both heme and chlorophyll. “Heme” = Enzymes for heme (FC) and phytochromobilin (HO and HY2) biosynthesis. “Chlorophyll” = enzymes for the biosynthesis of chlorophyll A and B. Numbering of protein names represents their order within the pathway. For complete names of enzymes and corresponding *A. thaliana* gene ID, see supplemental dataset 2. Hue of boxes indicates metabolism (gray = autotroph, orange = hemiparasite, purple = heterotroph), shade indicates degree of presence/absence (white = no sequence found, light colors = partial sequence found, dark colors = complete or near complete sequence found).



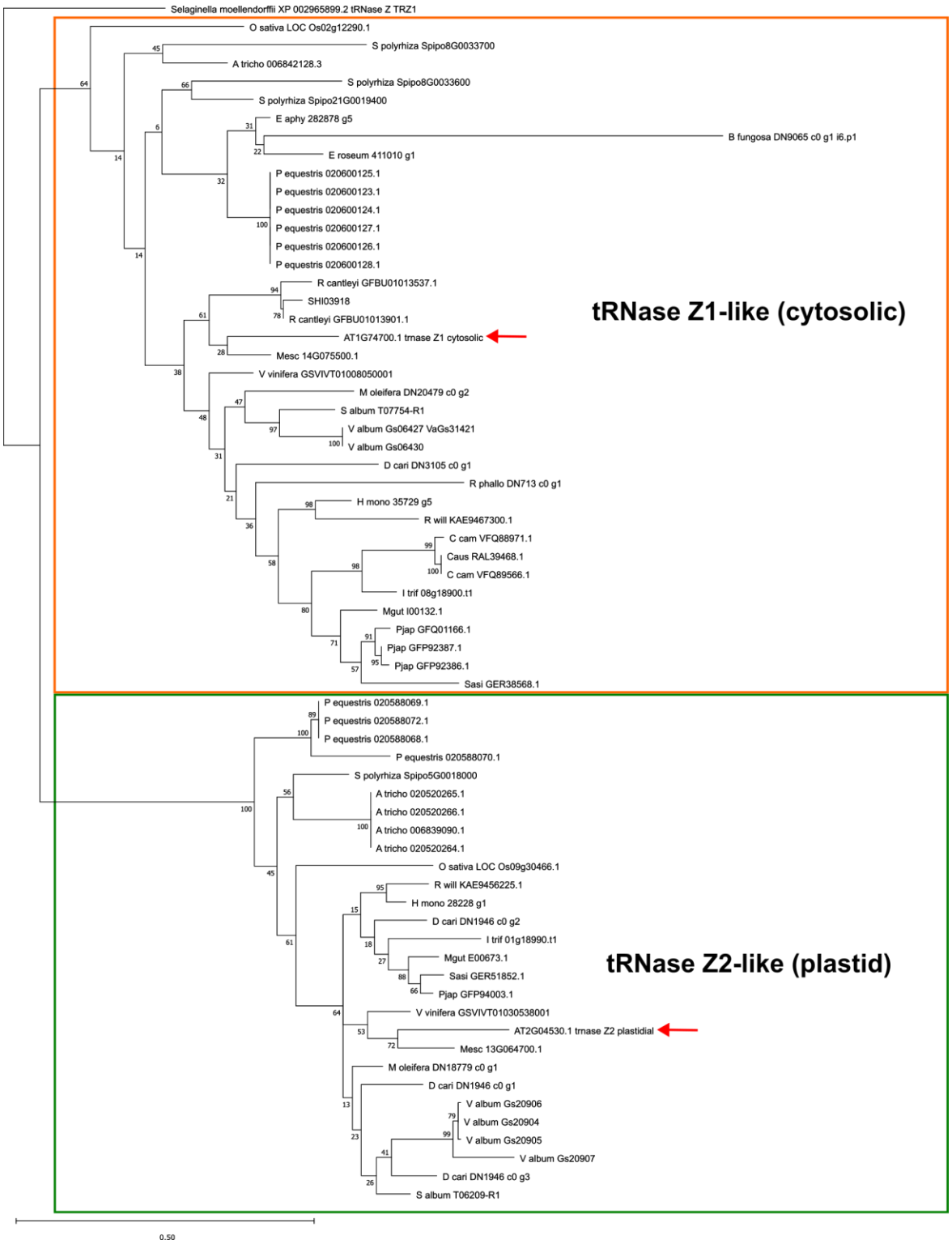
Supplemental Figure 6: Targeting predictions for tetrapyrrole pathway enzymes suggests potential relocation of pathway to cytosol in *Rafflesiaceae*. Deeper shades of green indicate probability of targeting to the plastid is ≥ 0.25 based on both LOCALIZER and TargetP. Lighter shades of green indicate one of the two algorithms yielded a probability ≥ 0.25. White indicates probability for both algorithms was < 0.25. Gray indicates that the gene/transcript was not present or that the N-terminus was incomplete, so no targeting prediction could be made. Enzymes are divided based on whether they are involved in the common (shared) pathway, the heme/phytochromobilin branch, or the chlorophyll branch. Numbering of protein names indicates their order within the pathway. Heterotrophs are written in purple, closely related autotrophs in black. *A. thaliana* is included as a reference.



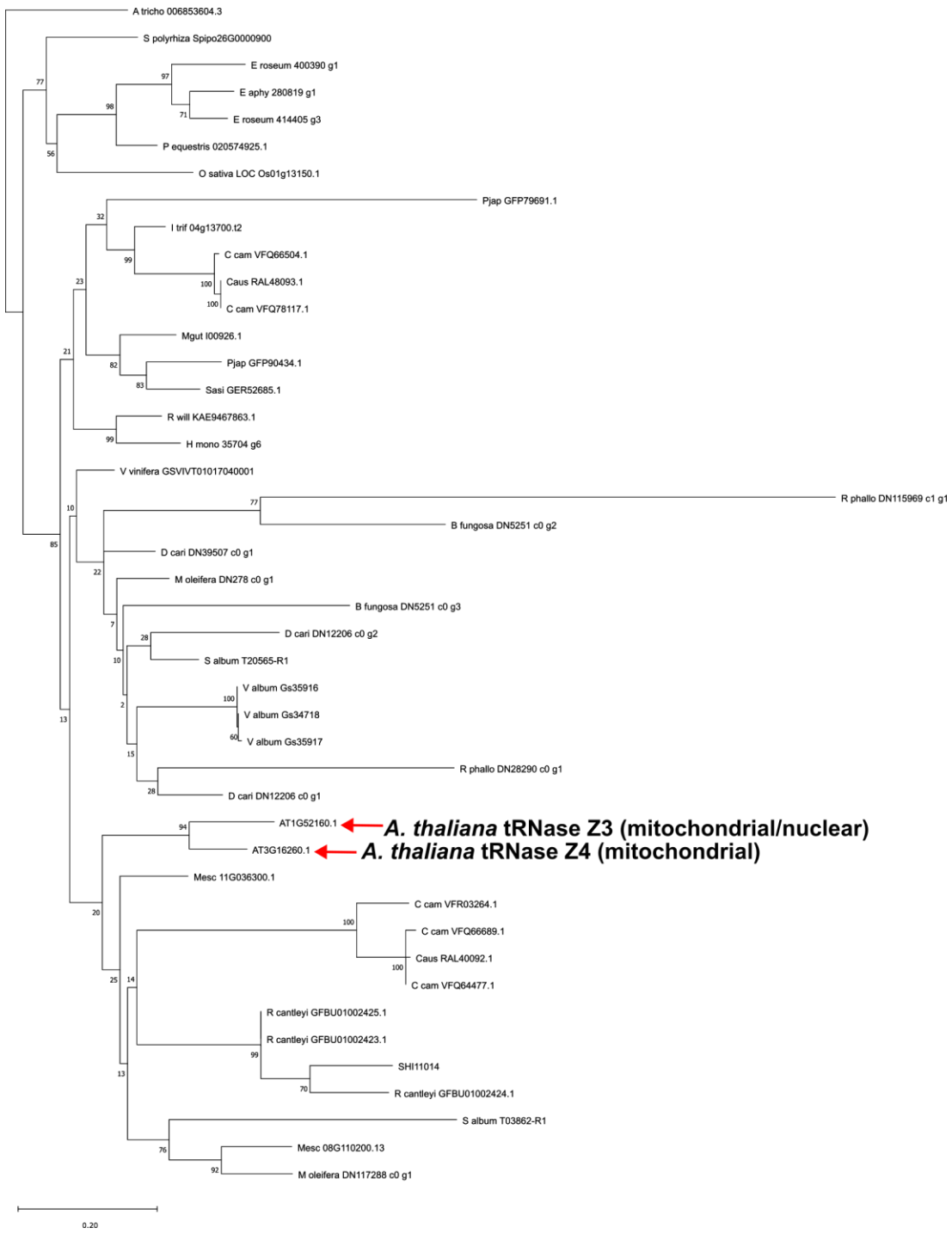
Supplemental Figure 7: Maximum likelihood tree for CysRS. Bootstrap values are indicated at nodes. Branch length labels represent average substitutions per site. Red arrow denotes *A. thaliana* sequences for organellar and cytosolic CysRS. Orthologs for *A. thaliana* organellar CysRS (AT2G31170) outlined in green, orthologs for *A. thaliana* cytosolic CysRS (AT3G56300, AT5G38830) outlined in orange.



Supplemental Figure 8: Maximum likelihood tree for Protein Only RNase Ps (PRORPs). Bootstrap values are indicated at nodes. Branch length labels represent average substitutions per site. Red arrow denotes *A. thaliana* sequences for nuclear and organellar PRORPs. Orthologs for *A. thaliana* organellar PRORP1 (AT2G32230) outlined in green, Orthologs for *A. thaliana* cytosolic PRORP2/3 (AT2G16650, AT4G21900) outlined in orange.



Supplemental Figure 9: Maximum likelihood tree for tRNase Z 1/2. Bootstrap values are indicated at nodes. Branch length labels represent average substitutions per site. Red arrow denotes *A. thaliana* sequences for tRNAse Z type proteins. Orthologs for *A. thaliana* cytosolic tRNase Z1 (AT1G74700) are outlined in orange, and orthologs for *A. thaliana* plastid tRNase Z2 (AT2G04530) are outlined in green.



Supplemental Figure 10: Maximum likelihood tree for tRNase Z 3/4. Bootstrap values are indicated at nodes. Branch length labels represent average substitutions per site. Red arrow denotes *A. thaliana* sequences for tRNAse Z type proteins.