Photosynthetic demands on translational machinery drive retention of redundant tRNA metabolism in plant organelles

Rachael Ann DeTar[1,], Joanna Chustecki[2], Anna Martinez-Hottovy[2], Luis Federico Ceriotti[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, tRNA, photosynthesis, aminoacyl-tRNA synthetase

Abstract

Eukaryotic nuclear genomes often encode distinct sets of translation machinery for function in the cytosol vs. organelles (mitochondria and plastids). This 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 plants because translation machinery, including aminoacyl-tRNA synthetases (aaRS), is often dual-targeted to the plastids and mitochondria. These organelles have different functions, 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 eukaryotes that lack plastids. Thus, we investigated the evolution of nuclear-encoded organellar and cytosolic aaRS and tRNA maturation enzymes across a broad sampling of angiosperms, including non-photosynthetic (heterotrophic) plant species with reduced plastid gene expression, to test the hypothesis that translational demands associated with photosynthesis constrain the evolution of 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 clear 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 functionally redundant proteins that are 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 essential genes to the nucleus via horizontal transfer or outright loss of some genes that are no longer useful after transition to an intracellular lifestyle. This trajectory has been observed repeatedly for various endosymbionts and organelles, including mitochondria, plastids, and more recently acquired intracellular bacteria such as the cyanobacterial-derived chromatophores of *Paulinella* amoeba or the many heritable endosymbionts in insects (1–6). However, there are some endosymbiotically derived genes that are recalcitrant to loss despite the presence of host counterparts with comparable functions. Redundant genes may be maintained in the organelle genomes or in the nuclear genome alongside genes encoding proteins for the same role in other cellular compartments. These observations elicit some interesting questions; 1) Why are some functionally redundant genes retained? 2) Do these redundant genes evolve differently depending on the cellular compartment where they operate?

The genes that encode the machinery required for organelle translation are widely retained, as eukaryotes maintain distinct translation systems for nuclear and mitochondrial transcripts. Translation becomes even more complicated in eukaryotes bearing three genomes. For example, plant cells contain nuclear, mitochondrial, and plastid genomes. The translation machinery associated with each of these cellular compartments includes ribosomal RNAs (rRNAs), ribosomal proteins, transfer RNAs (tRNAs), and the accompanying enzymes involved in tRNA processing and aminoacylation. The aminoacyl-tRNA synthetases (aaRS) that catalyze the “charging” of a tRNA with its cognate amino acid tend to be highly conserved, with distinct versions for function in the organelles and cytosol (7). The degree of 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 CCAses (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 that 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 more likely reflect differences in selection, rather than other factors like
47 compartment-specific mutation rates or effective population size (18–21). Previous studies have

48 quantified rates of evolution for these enzymes in various animal lineages and a small number of plants.
49 Mitochondria-specific aaRS in animals exhibit lower sequence conservation than their cytosolic
50 counterparts (22–24). In contrast, plant organellar aaRS exhibit equivalent or greater sequence
51 conservation compared to their cytosolic counterparts (25). Furthermore, plastid genomes (plastomes)
52 tend to retain a set of 30 tRNA genes that is fully sufficient for translation (26, 27). Plastomes also
53 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 tRNA processing and charging enzymes by analyzing evolution of this machinery across
63 diverse lineages of autotrophic and heterotrophic plants. It should be noted there is a gradient of
64 heterotrophy amongst plants – for the purposes of this study, we use “heterotrophic” to describe plants
65 that do not photosynthesize at all and exclusively rely on parasitism for acquisition of carbon. They do
66 this by tapping directly into host plants or by exploiting mycorrhizal fungi (mycoheterotrophy) (30, 31).
67 We 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

been lost from the plastomes of these plant species (32), providing a good system to examine the relationship between evolutionary redundancy of aaRS, tRNA content of the organelle genomes, and trophic lifestyle.

Results

Heterotrophic plants exhibit extensive loss of organellar aaRS and tRNA processing enzymes but not ribosomal protein subunits

We searched for organellar and cytosolic aaRS orthologs in a broad sampling of angiosperm species with a range of trophic lifestyles and plastome reduction (**Supp. Dataset 1**). For each hemiparasitic/heterotrophic species in the analysis, we included the closest fully autotrophic relative with a nuclear genome or transcriptome available, excepting the Balanophoraceae where we used hemiparasite *Santalum album* as the reference autotroph. In total, we sampled 7 heterotrophs, 8 hemiparasites, and 10 autotrophs, including reference species *Arabidopsis thaliana.* Among heterotrophs, we included the endoparasites *Rafflesia cantleyi* and *Sapria himalayana* from the family Rafflesiaceae, likely the most extreme examples of plant heterotrophy, as several studies have suggested wholesale loss of the plastome in these species (33–36).

We found orthologs across these diverse taxa by running Orthofinder (37, 38) on protein models derived from nuclear genome or transcriptome data to find groups of likely orthologous proteins based on similarity and phylogeny, i.e., “orthogroups”. Then we extracted sequences that fall into an orthogroup with known *A. thaliana* aaRS (**Supp. Dataset 2**). It is important to note that the number of homologous sequences found for each species in this analysis may not directly reflect the absolute number of genes or proteins due to the fragmented nature of gene models made from transcriptome data and the inclusion of two independent sets of protein models for *Balanophora fungosa* and *S. himalayana.* Unexpectedly, many lineages of heterotrophs did not have any orthologs for many of the organellar aaRS. This loss of organelle-specific orthologs in heterotrophs was surprising as even species

96 entirely lacking the plastome must still conduct tRNA aminoacylation within mitochondria. Further
97 investigation using reciprocal-best-hit (RBH) searches with BLAST between *A. thaliana* and each species
98 supported that these were true losses. These losses primarily occurred in purely heterotrophic plants,
99 while hemiparasites retained most of their organellar aaRS. (**Fig. 1**). Yet, within the heterotrophs, there
100 was a gradient of loss. The two endophytic species from Rafflesiaceae (*R. cantleyi* and *S. himalayana*)
101 exhibited the most striking loss of organellar aaRS, retaining only PheRS. In contrast, obligate
102 mycoheterotroph *Monotropa hypopitys* retained more aaRS than any other heterotroph, missing only
103 AlaRS, CysRS, ThrRS, and TrpRS (**Fig. 1**).
104 This striking pattern led us to investigate the retention of other enzymes involved in tRNA
105 maturation and processing; see (15) and **Supplemental Dataset 2** for a comprehensive list.
106 Unsurprisingly, the CCAse enzyme responsible for adding CCA tails to tRNAs in all compartments was
107 retained across all species. Two enzymes that maintain distinct cytosolic and organellar versions, tRNase
108 Z (involved in 3' end processing) and Protein Only RNase P (PRORP; required for 5' end processing),
109 exhibited two different trends. We found that plastid tRNase Z (tRNase Z2) (39) was missing from all
110 heterotrophs except *M. hypopitys* and hemiparasites of the *Cuscuta* genus yet was retained in the
111 autotrophic relatives. Conversely, the organellar PRORP1 (40, 41) appeared to be universally retained
112 while cytosolic PRORP2/3 orthologs were missing from the Rafflesiaceae. We also investigated the
113 retention of bacterial-like organelle-specific machinery with no cytosolic counterpart, including the
114 methionyl-tRNA formyltransferase (MTF), tRNA-Ile lysidine synthetase (TilS), and the bacterial-type
115 glutamine amidotransferase (GatCAB) complex. We found orthologs for the putative *A. thaliana* MTF in
116 all species except the heterotrophs *Epipogium roseum* and *M. hypopitys,* and putative TilS orthologs
117 were absent only from the heterotrophs Rafflesiaceae and *B. fungosa*. GatCAB is the key complex used
118 by organelles in land plants to accurately charge $tRNA^{Gln}$ in the absence of a dedicated organellar GlnRS
119 (42–44). This involves a two-step process, whereby $tRNA^{Gln}$ is initially mis-acylated with Glu by GluRS.

120 GatCAB then catalyzes the transamidation of Glu to Gln. Most of our test species, even heterotrophs,
121 retained genes for at least some of the GatCAB subunits. However, Rafflesiaceae species appear to have
122 lost all subunits in this complex.

123 To assess whether other functionally redundant translation machinery was lost in heterotrophic
124 lineages, we analyzed presence/absence of nuclear-encoded riboproteins with exclusive targeting to a
125 given compartment (Table S1, Supp. Dataset 2, "Riboproteins"). While there was minor loss of cytosolic,
126 mitochondrial, and plastid subunits in all heterotrophic lines, the Rafflesiaceae lineages exhibited a
127 dramatic reduction in plastid subunits, which is expected because these lineages appear to have
128 dispensed with the plastid genome entirely and consequently no longer need plastid ribosomes. This
129 finding reveals a key contrast between the fate of organellar aaRS and riboproteins in response to loss of
130 photosynthesis and may reflect the fact that aaRS are shared between the two organelles, whereas the
131 riboproteins are compartment-specific.

132 Generally, we observed reductive evolution in all heterotrophs via loss of organellar aaRS and
133 bacterial-type enzymes for tRNA modification. This raises the question as to how gene expression is
134 accomplished in the organelles of these species. One possibility is that cytosolic enzymes are retargeted
135 to the organelles to functionally replace organellar enzymes.

136 ***Heterotrophic plants may compensate for loss of organellar aaRS by retargeting cytosolic***
137 ***counterparts***

138 To test for cytosolic aaRS retargeting, we applied two *in silico* targeting prediction programs to protein
139 models for all cytosolic aaRS in our heterotrophs and their close autotrophic relatives (**Fig. 2, Fig. S1)**.
140 Our analysis revealed that canonically cytosolic aaRS are disproportionately retargeted to mitochondria
141 in obligate heterotrophs compared to autotrophs (p = 0.018). There was also a non-significant trend
142 towards retargeting to the plastid (p = 0.11). Targeting of mitochondrial OXPHOS components and
143 plastid housekeeping proteins including subunits of the caseinolytic protease (Clp) complex and the

144 acetyl-CoA carboxylase (ACC) complex involved in producing malonyl-CoA for fatty acid biosynthesis
145 were included as known organellar positive controls (Supp. Dataset 2, “targeting controls”). Protein
146 degradation and fatty acid biosynthesis are essential plastid functions that are still necessary in the
147 absence of photosynthesis. Cell wall synthesis enzymes were included as negative controls. Of the
148 controls, the only group that showed a different degree of targeting than expected was the
149 mitochondrial OXPHOS components, where recovery of proteins with known mitochondrial targeting
150 was low (~60%) for both autotrophs and heterotrophs (Supp. Table 2). Thus, these *in silico* approaches
151 appear to successfully (albeit imperfectly) predict subcellular localization in both heterotrophic and
152 autotrophic plants.

153 We would specifically expect that a given cytosolic aaRS is more likely to be retargeted if its
154 organellar counterpart is lost. We examined this possibility and found a clear and statistically significant
155 correlation between loss of an organellar aaRS and retargeting of the corresponding cytosolic enzyme to
156 one or both compartments (p = 0.005, **Fig. 4, Fig. S2**).

157 The variable quality of gene models for each species and high false negative rate for retargeting
158 prediction confound interpretation of retargeting for individual enzymes in each species (**Fig. 4**, **Fig. S3**).
159 However, when looking only at *S. himalayana*, a species with relatively high-quality genome-based
160 models, we observed some interesting phenomena. Cytosolic AlaRS, GlnRS, MetRS, ProRS, and TrpRS
161 were predicted to be retargeted to mitochondria. Retargeting of cytosolic GlnRS could explain how this
162 species copes with loss of the GatCAB complex (**Fig. 1**). The cytosolic AspRS, GluRS, and GlyRS were
163 predicted to be dual targeted to both organelles, while cytosolic TyrRS and the PheRSα subunit were
164 predicted to be targeted to the plastid despite the likely absence of a plastid genome. All other cytosolic
165 enzymes were either not predicted to be retargeted or were missing a large portion of the N-terminus.
166 The fact that some cytosolic aaRS corresponding to lost organellar aaRS were not predicted to be
167 retargeted to the mitochondria and that some cytosolic aaRS were targeted to a plastid presumed to

168 lack gene expression may reflect an underestimation of mitochondrial retargeting due to the constraints
169 of *in silico* predictions and the potential divergence of TP sequences in Rafflesiaceae. Indeed, an
170 investigation of subunits for the protein import apparatus of plastids and mitochondria revealed that
171 the Rafflesiaceae are missing some key components of the plastid import apparatus (Fig. S6), some of
172 which are responsible for TP binding and recognition.

173 To assess whether our *in silico* predictions hold true *in vivo,* we conducted targeting assays on a
174 subset of aaRS (see **Supplemental Dataset 3** for sequence information) by inducing transient expression
175 of N-terminal fusions of *S. himalayana* cytosolic aaRS transit peptides (TPs) to green fluorescent protein
176 (GFP) in *Nicotiana benthamiana* (**Fig. 3, Fig. S4-S5**). We used eqFP611 tagged with the TP from
177 mitochondrial isovaleryl-CoA dehydrogenase (IVD) as a positive control for mitochondrial targeting.
178 Infiltration with single expression vector with just IVD_FP611 was used as a negative control (**Fig. S4**).
179 We found that proteins predicted to be retargeted *in silico* were almost universally retargeted to one or
180 both organelles *in vivo*. In some cases, putative TPs retargeted GFP to a different organelle than
181 predicted *in silico.* AsnRS displayed the plastid targeting that was predicted *in silico* and additional
182 mitochondrial targeting *in vivo* (**Fig. 3**). GluRS and GlyRS were predicted to be dual-targeted but
183 exhibited targeting to only one of the two organelles. The only two TPs for which organellar GFP
184 localization completely mismatched the *in silico* prediction (but still exhibited organelle targeting) were
185 ProRS and TrpRS, which both showed strong plastid GFP fluorescence *in vivo* despite mitochondrial
186 predictions. These results were consistently observed with two different cloning and expression
187 strategies (Fig. 3, Fig. S5)**.** The cytosolic AspRS TP was the only one we cloned that did not exhibit
188 targeting to either organelle *in vivo* despite a clear biological expectation that it would have evolved
189 retargeting (**Fig. S4**). GFP tagged with this TP formed punctate structures in the cytosol, which may be
190 artefactual aggregates resulting from overexpression. Generally, *in vivo* assays provided evidence for

191 retargeting of cytosolic aaRS to at least one organelle and suggested that *in silico* prediction tools may
192 not be adept at distinguishing plastid from mitochondrial targeting.

193 Generally, organellar tRNAs differ from their cytosolic counterparts in evolutionary history,
194 sequence, and posttranscriptional modification (8). Identity elements that govern the specificity of
195 interaction between a tRNA and its conjugate aaRS may also differ between cytosolic and organellar
196 systems (45). However, in many systems (including plants), cytosolic tRNAs can be imported into the
197 organelles (8, 9, 14). Thus, our observation of retargeting opens some interesting questions – are the
198 retargeted aaRS charging organellar tRNAs? Or is there replacement of the entire organellar system of
199 tRNA aminoacylation, including both tRNAs and aaRS? We investigated the tRNA gene content of the
200 organelles of these species to gain insight into which tRNA substrates may be in use.

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

202 To investigate the possibility of enzyme-substrate mismatch, we cataloged the presence/absence of
203 organellar tRNA-encoding genes (*trn*) for all sampled heterotrophic species (**Fig. 4, Supplemental**
204 **Dataset 4**)**.** Generally, there was strong correspondence between retention of organellar aaRS and
205 plastome/mitogenome tRNA genes. As the Rafflesiaceae species are thought to lack plastomes entirely,
206 they likely have no plastome-encoded tRNAs (**Fig. 4**). The Balanophoraceae species used in this study
207 also have been reported to exhibit loss of all plastid tRNA genes, except for an extremely divergent *trnE*
208 in *B. fungosa* (46–49). The Rafflesiaceae and Balanophoraceae have additionally lost most of their
209 mitogenome-encoded tRNAs. Correspondingly, these two families also have the most reduced sets of
210 organellar aaRS. At the opposite extreme, *M. hypopitys* retains most of its mitochondrial and plastid
211 tRNA genes and corresponding organellar aaRS.

212 However, there are some incidences where a cytosolic enzyme looks to be charging an
213 organellar tRNA due to loss of the organellar aaRS (**Fig. 4**). For example, the few remaining mitogenome-
214 encoded tRNAs in the Rafflesiaceae are missing their organellar aaRS partner (except *S. himalayana*

215 tRNA^{Phe}) and are likely charged by a cytosolic enzyme. Alternatively, they may be pseudogenes that have
216 been functionally replaced by an imported tRNA. For both compartments, there are numerous instances
217 where there is no encoded tRNA despite retention of the organellar aaRS (**Fig. 4**). For example, *R.*
218 *phalloides* retains its organellar SerRS, PheRS, and GlnRS but has lost the corresponding mitochondrial
219 tRNAs. No retargeting of cytosolic enzymes is predicted in these cases, suggesting the organellar enzyme
220 may be acting on an imported cytosolic tRNA.

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

222 Although wholesale loss of many organellar aaRS in heterotrophs precludes analysis of sequence
223 conservation, *M. hypopitys* retained the majority of organellar aaRS despite exhibiting an obligate
224 heterotrophic lifestyle. Thus, we quantified sequence conservation of organellar and cytosolic aaRS in
225 *M. hypopitys* and its autotrophic relative *Rhododendron williamsianum.* We built gene trees and ran
226 relative rate analysis to test for accelerated evolution on the *M. hypopitys* branch (**Supplemental**
227 **Dataset 5**). We also directly compared branch length ratios of *M. hypopitys* to *R. williamsianum* in
228 cytosolic and organellar aaRS (**Fig. S7, Supplemental Dataset 5).** Neither analysis showed a statistically
229 significant decrease in sequence conservation in organellar enzymes of *M. hypopitys* (**Fig. S7**). However,
230 the deep rooting of our tree on *A. thaliana* may result in poor resolution of the sequence evolution
231 between *M. hypopitys* and *R. williamsianum.* Additionally, the transition to heterotrophy on the *M.*
232 *hypopitys* branch may have occurred relatively recently compared to the split from *R. williamsianum*,
233 thus diluting any signals of accelerated evolution.

234 **Discussion**

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

236 Our primary objective in this study was to test whether the high translational demands of the
237 photosynthetic chloroplast create functional constraint on the evolution of shared translation systems in
238 plastids and mitochondria. To do so, we investigated relative rates of protein evolution in cytosolic and

239 organellar aaRS in heterotrophic plants compared to their photosynthetic relatives. To our surprise,
240 heterotrophic plants underwent widespread loss of organellar aaRS and many bacterial-like enzymes
241 involved in tRNA processing. These losses suggest an extreme relaxation of selection pressures in the
242 absence of a photosynthesizing plastid and support our hypothesis that the translational demands in
243 chloroplasts are the linchpin responsible for the unusually strong functional constraint on the redundant
244 organelle-specific tRNA metabolism in plants. These correlated effects on mitochondrial translation
245 appear to be stronger when the same proteins are shared by mitochondria and plastids (e.g., aaRS)
246 because we found that the most reduced heterotrophs from Rafflesiaceae lost many plastid-specific
247 riboproteins but not their mitochondrial-specific counterparts.

248 There was variation in the degree of loss across the heterotrophs, often correlated with the
249 number of genes retained in the plastome, particularly tRNAs (see **Supp. Dataset 1** for
250 plastome/mitogenome content). For example, the most aaRS losses occurred in the Rafflesiaceae, all of
251 which are obligate endoparasites and have likely lost their plastomes entirely (33–36, 50). The fewest
252 aaRS losses occurred in *M. hypopitys*, which has the highest number of plastome-encoded genes of all
253 the heterotrophic species we sampled – and notably the greatest number of retained plastid tRNAs (51).
254 Thus, *M. hypopitys* may not be as far along the path of reductive evolution as the other species, which
255 might provide an additional explanation as to why our comparative rate analyses on the remaining
256 organellar aaRS in *M. hypopitys* did not support accelerated evolution compared to cytosolic enzymes
257 (**Fig. S7**). The Balanophoraceae and *Epipogium* species exhibited intermediate degrees of aaRS and
258 plastome gene loss, although the Balanophoraceae species used in this study have lost all plastid tRNAs
259 save for a truncated *trnE* in *B. fungosa* that is likely retained for use in the first step of heme
260 biosynthesis *(46).* An increased plastome mutation rate caused by loss of DNA repair machinery (52, 53)
261 in the Balanophoraceae may explain accelerated plastome (tRNA) gene losses relative to nuclear (aaRS)

262 gene loss. Generally, loss of organellar aaRS correlates well with reduction of the plastome across
263 heterotrophs.

264 ***Loss of organellar aaRS is associated with retargeting of cytosolic aaRS***

265 The wholesale loss of organellar aaRS in heterotrophs was unexpected, as most non-photosynthetic
266 plants still contain a minimal plastome to support heme and fatty acid biosynthesis (32, 54) and fully
267 intact mitogenomes (see **Supp. Dataset 1**). This conundrum prompted us to investigate if cytosolic aaRS
268 are retargeted to organelles in heterotrophs to compensate for gene loss, as has been shown for
269 aaRS/tRNAs in mitochondria of other plant species (55). Retargeting a protein to the organelles
270 generally involves obtaining an N-terminal TP sequence that can be recognized by organellar protein
271 import complexes (56–58). We discovered that many cytosolic aaRS in the heterotrophs had N-terminal
272 sequences predicted to target to one or both organelles (**Fig. 2-3, S2-S5**). Therefore, the import of
273 cytosolic aaRS likely compensates for the loss of organellar aaRS and even the loss of the GatCAB
274 complex in Rafflesiaceae via retargeting of cytosolic GlnRS.

275 There was not perfect correspondence between loss of organellar aaRS and predicted
276 retargeting of cytosolic aaRS. However, the extent of retargeting may have been underestimated for
277 two main reasons. Firstly, TP sequence composition and protein import machinery vary considerably by
278 plant lineage (59), making *in silico* predictions more challenging in uncharacterized species. Thus, some
279 organelle-imported cytosolic aaRS may eschew the use of typical TPs that are key for targeting
280 prediction *in silico* (60). Indeed, some organellar proteins are known to not rely at all on N-terminal TPs
281 for import (61, 62). TP sequence divergence could also be accelerated by alteration of the sequence or
282 subunit composition of protein import machinery. Notably, our presence/absence analysis of plastid and
283 mitochondrial protein import machinery revealed some critical subunits for plastid protein import are
284 lost in the heterotrophs, especially the Rafflesiaceae, which could influence which sequences are
285 recognized as transit peptides (**Fig. S6**). For example, Toc159, one of the key outer envelope import

286 receptors involved in TP binding and recognition, is missing from the Rafflesiaceae (63). This subunit
287 specializes in binding and import of proteins involved in photosynthesis, whereas Toc132, which is
288 retained in these species, specializes in “non-green” protein import (64, 65). The Rafflesiaceae are also
289 missing Tic236, the protein that spans the space between the envelope membranes to physically
290 connect the inner and outer envelope (66). This raises interesting questions about the nature of plastid
291 protein import and the ultrastructure of envelope membranes in Rafflesiaceae. Other heterotrophs in
292 addition to the Rafflesiaceae are also missing some subunits present in *A. thaliana*; *Epipogium* species,
293 *M. hypopitys* and the Rafflesiaceae lack one or more Tic20 and Tic22 paralogs. The Tic22 protein family
294 is highly conserved and expressed, with a hypothesized role in outer membrane maturation in addition
295 to its role in protein import (67, 68). Study of Tic-22 loss-of-function mutants revealed that these
296 proteins are not essential, at least in *A. thaliana* (67, 69). Tic20 forms the structural basis of the inner
297 pore of the translocon, with Tic20-I/Tic20-IV being the essential isoforms for plant survival (63, 70–72).
298 Similar trends have also recently been found by Guo et al. 2024 (73). These alterations to the
299 composition of the TIC-TOC complex may reduce the specificity of plastid protein targeting predictions.
300 In contrast, the heterotrophic lines retained almost all TIM-TOM subunits forming the mitochondrial
301 translocon. The divergence of the TIC-TOC complex in some heterotrophs may result in accelerated
302 evolution of TPs, thus, targeting predictions made based on characteristics of photosynthetic plants may
303 not be fully accurate.

304 Secondly, uncertain protein models for some of our species may have led to true N-terminal
305 sequences being absent or incomplete. Retargeting of cytosolic aaRS may occur through two different
306 mechanisms; 1) alternative splicing or alternative transcription/translation start sites within one gene
307 resulting in cytosolic and organellar isoforms and 2) duplicated genes within the genome encoding
308 cytosolic and organellar paralogs. Our ability to investigate these two mechanisms was limited as we
309 generally only had transcriptome or genome information but not both for each species. However,

310 observations of transcriptomes assembled with consideration for alternative splice isoforms indicate
311 that both mechanisms are potentially at play (**Supplemental Dataset 6**). For example, *Rhopalocnemis*
312 *phalloides* has two separate paralogs for cytosolic ThrRS, one of which is predicted to be mitochondrial-
313 targeted (Rphallo_DN1669_c0_g1) while the other is not (Rphallo_DN1669_c0_g2). Conversely, *R.*
314 *phalloides* TyrRS is encoded by a single locus (Rphallo_DN204_c0_g1) with a multitude of splice
315 isoforms, some of which are putatively mitochondrial-targeted. Genome sequencing and long-read
316 mRNA sequencing would allow for more thorough analysis of which mechanism is most common in each
317 species.

318 While almost all the TPs that we investigated with *in vivo* microscopy assays exhibited targeting
319 to mitochondria and/or plastids, the specific organelle sometimes differed from the *in silico* prediction.
320 The ProRS and TrpRS results in particular were unexpected as the *in silico* predictions were strongly
321 mitochondrial (**Supp. Dataset 3**), but we observed clear plastid localization *in vivo* in two separate
322 rounds of imaging with different vectors (**Fig 3, Fig. S4, S5)**. There were other TPs where targeting
323 predictions were only partly supported by GFP localization experiments, including GluRS and GlyRS
324 which exhibited targeting to only one of the two predicted organelles, and AsnRS which was only
325 predicted to have plastid localization but additionally exhibited mitochondrial localization *in vivo.*

326 These results may reflect the fact that target peptide prediction models like those used in this
327 study do not have complete precision or sensitivity for a given subcellular compartment even when
328 applied to benchmark datasets. This is not surprising, as changing just a few amino acids in a transit
329 peptide can alter protein targeting from singular to dual targeting *in vivo* (74). Furthermore, predictions
330 made by a classifier trained on a broad sampling of TPs from diverse plant species cannot accurately
331 predict every TP for an individual species, since species-specific features may not be represented
332 enough to be detected by the model. Even if we assume our microscopy data are a perfect reflection of
333 "true" targeting in the native system, we might expect some inaccuracies in our targeting predictions

334 just based on the sensitivity and precision of the prediction algorithms (TargetP-2.0 and LOCALIZER). Of
335 the two models, TargetP-2.0 specifically identifies mitochondrial and plastid transit peptides with 85-
336 86% sensitivity and 87-90% precision and differentiates proteins lacking TPs with over 98% sensitivity
337 and precision (75, 76). The publication describing LOCALIZER does not directly quantify the tool's ability
338 to distinguish proteins with no transit peptide but has 60-73% sensitivity and 60-80% precision for
339 mitochondrial and plastid predictions respectively, counting proteins with predicted dual-targeting as
340 positive hits for both plastid and mitochondria.

341 The artificial nature of heterologous expression in *N. benthamian* can also create artefacts. We
342 can see that in some cases predictions for a given TP can change when it is fused to GFP as opposed to
343 the native aaRS sequence (Supp. Dataset 3). This implies that more than just the N-terminal extension of
344 a protein determines targeting and may create some artefacts in assessing targeting based on TP-GFP
345 fusions. Furthermore, alternative splicing or translation initiation sites can be a mechanism of
346 alternative targeting (77–79). Alternative translation initiation may require RNA cis regulatory factors
347 i.e., 5′ UTR sequences not captured in our heterologous expression system (80). Additionally, the import
348 of a heterologously expressed TP-tagged GFP can change based on the age of leaves during infiltration,
349 the efficiency of protein folding, protein expression levels, and truncation of an internal targeting
350 sequence (81–83). As comparison of the *in silico* and heterologous expression assays is effectively
351 combining the errors created by both methods, we would not expect perfect concordance between the
352 methods regarding the specific identity of the retargeted organelle. Rather, these methods are
353 complementary tools to assess general retargeting to the endosymbiotic organelles rather than
354 differentiating between TPs specifically targeting mitochondria or plastids from those that are dual
355 targeted. An additional method for future investigation of targeting to plastids or mitochondria would
356 be to subfractionate tissue from one or more of these heterotrophic plants species and conduct
357 comparative high throughput proteomics on each fraction.

358 ***Potential rewiring of aaRS-tRNA interactions***

359 Retargeting of cytosolic aaRS led us to investigate organelle genome tRNA content and whether
360 there was substrate-enzyme mismatch in the organelles, i.e. instances where the cytosolic enzyme was
361 interacting with an organellar tRNA or vice-versa. Our interpretations presume that loss of organellar
362 tRNA genes is compensated by import of the corresponding cytosolic tRNAs (9, 14, 84). We found that
363 the retention of organellar aaRS in heterotrophs generally matched the retention of plastid tRNAs but
364 not necessarily mitochondrial tRNAs (**Fig. 4**). This is further supported by observations in hemiparasite
365 *Viscum album*, which has lost most of its mitochondrial tRNAs, yet retains most organellar aaRS in
366 conjunction with a nearly full set of plastid tRNAs (32, 85, 86). Substrate-enzyme mismatch was rare
367 within the plastid and usually involved loss of a plastid tRNA yet continued presence of the
368 corresponding organellar aaRS (e.g., see *R. phalloides* where all plastid tRNAs are lost but a handful of
369 organellar aaRS remain). There were also instances of organellar aaRS apparently charging imported
370 cytosolic tRNAs in the mitochondria (e.g., *trnF* in *R. cantleyi and R. phalloides*). More surprisingly,
371 however, there were cases where a mitochondrial tRNA was retained despite loss of the corresponding
372 organellar aaRS and apparent retargeting of the cytosolic enzyme, particularly in the Rafflesiaceae. This
373 type of mismatch is unexpected because cytosolic aaRS are thought to generally be more discriminating
374 enzymes than their mitochondrial counterparts (8, 55, 87–89), making it more difficult to transition to a
375 new tRNA substrate. However, our results suggest some cytosolic enzymes may be interacting with
376 bacterial-type tRNAs in these heterotrophic species, although we cannot rule out the possibility that
377 cytosolic tRNAs are being redundantly imported. Indeed, our findings raise fascinating questions about
378 the process of functionally replacing tRNAs and interacting enzymes in these dynamic systems (55, 90),
379 which could be explored with comparative and biochemical approaches in closely related species still
380 undergoing these transitions. Given that parasitic plant lineages tend to be fast evolving across all three
381 genomes (91), the cytosolic aaRS in these species may have evolved lower specificity and the capability

382 to charge organellar tRNAs. Overall, our findings suggest that the plastid tRNA content, and not the
383 mitochondrial tRNA content, drives retention of organellar aaRS, again placing the photosynthetic
384 plastid as a central factor in the retention of functionally redundant translation machinery.

385 The identity of retained organellar aaRS versus those substituted by cytosolic counterparts was
386 seemingly non-random. GluRS and PheRS were the two mostly highly retained organellar aaRS across
387 these diverse lineages of plants (**Fig. 1**). Organellar PheRS is likely retained because the cytosolic PheRS
388 is unusual in that it is a heteromeric complex with independently encoded α and β subunits (92).
389 Therefore, functional replacement of the organellar PheRS would require two independent acquisitions
390 of TPs, as well as stoichiometric import and assembly within the organelles. Such a coordinated set of
391 changes may be unlikely to evolve (24, 55). The one exception was *E. aphyllum,* for which we could not
392 find an organellar PheRS. However, this may be an artefact of low expression, as protein models for this
393 species were curated from RNA-seq data. As $tRNA^{Glu}$ is the most frequently retained tRNA in
394 heterotrophic plant plastomes due to its additional role in the first step of heme biosynthesis (54, 93,
395 94), it is not surprising that GluRS was among the most retained organellar aaRS in this study.
396 Intriguingly, the *Rafflesiaceae* are the only species that lost the organellar GluRS, as well as plastid *trnE*.
397 Yet, genes encoding heme synthesis enzymes are retained in these species (92), opening interesting
398 questions into how they carry out the first step in the pathway. In *S. himalayana*, we found retargeting
399 of the cytosolic GluRS to the plastid, which along with presumed import of the cytosolic $tRNA^{Glu}$, could
400 functionally replace the first step in the heme biosynthesis pathway. This has the interesting implication
401 that the glutamyl-tRNA reductase catalyzing the next step in heme biosynthesis is non-discriminating
402 towards cytosolic tRNAs. The organellar GluRS is also necessary for mis-aminoacylation of $tRNA^{Gln}$ prior
403 to transamidation of Glu to Gln by the GatCAB complex(96). Cytosolic GlnRS is retargeted to the
404 mitochondria in *S. himalayana*, thus allowing for direct aminoacylation of $tRNA^{Gln}$ assuming there is also

405 import of cytosolic $tRNA^{Gln}$. This likely precipitated the loss of the GatCAB complex and negated the
406 need for a non-discriminating organellar GluRS.

407 At the other extreme, there was repeated loss of organellar AlaRS, which was absent from all
408 heterotrophs and the hemiparasitic *Cuscuta* species, possibly because plastomes of heterotrophs are
409 often AT-rich and thus have underrepresentation of amino acids with GC-rich codons like Ala, Gly and
410 Pro (47, 48). Thus, the corresponding organellar aaRS and tRNAs may be subject to decreased functional
411 constraint and be less likely to be retained. This hypothesis is supported by the fact that we also see
412 frequent loss of organellar GlyRS and ProRS in heterotrophs (**Fig. 1**).

413 Our results indicate that imported cytosolic aaRS and tRNAs are sufficient to sustain
414 mitochondrial translation in plants, supporting the hypothesis that the selective pressure to keep
415 organellar aaRS/tRNAs is maintained by translational demands associated with photosynthesis. Yet
416 other eukaryote lineages that never bore plastids often still retain organelle-specific tRNAs and aaRS for
417 their mitochondria (8, 9, 22, 24). This is even more puzzling when considering that various eukaryotic
418 lineages have jettisoned nearly all their mitochondrial tRNAs/aaRS, including cnidarians, ctenophores,
419 some fungi, and trypanosomes (12, 24, 97). One possible explanation is the logical parallel to our finding
420 that photosynthesis governs plastid translation rates and maintains gene redundancy; eukaryotic
421 lineages with high respiratory rates may require high mitochondrial gene expression and, therefore,
422 retain mitochondria-specific translation machinery. Indeed, mitochondrial gene expression tracks well
423 with respiratory rate and cellular activity across different tissues and metabolic states (98–100). This
424 connection to cell energetics would also explain why mitochondrial tRNAs/aaRS are dispensable in
425 parasitic eukaryotes like trypanosomes and heterotrophic plants, which presumably have lower energy
426 demands. By studying non-model plant species with unusual physiology and life histories, we were able
427 to link plastid bioenergetic processes to the evolution of both organellar and nuclear genomes. The
428 sequencing of mitogenomes and characterization of metabolism in non-model eukaryotes will provide

429 exciting opportunities to conduct similar studies to better understand how mitochondrial metabolism
430 effects gene redundancy and elucidate the principles governing endosymbiotic integration.

431 ***Methods***

432 ***Data acquisition and ortholog search***

433 Gene models for species used in the study were collected from publicly available data sources
434 (**Supplemental Dataset 1**). When protein sequences were not available, we generated sequences from
435 transcriptome data using Transdecoder v5.5.0 (101), applying a minimum protein length of 50 amino
436 acids and retaining only protein sequences with blast hits to any sequence in the Uniprot Taxonomic
437 Divisions Plant database (https://ftp.uniprot.org/pub/databases/uniprot/knowledgebase
438 /taxonomic_divisions/). Then, the Transdecoder primary_transcript.fa script was used to extract only
439 the longest translated protein sequence from the primary transcript for each gene to reduce the search
440 space. SeqKit 2.3.0 (102) was used to create uniform gene handles for all files and remove stop codons
441 from sequences. Orthofinder v2.5.5 (37) was run on all sets of protein models with a user-defined
442 species tree (see **Fig. 1** for tree topology) to find groups of likely orthologous proteins, i.e. "orthogroups"
443 in two steps as follows:

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

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

446 Orthogroups containing proteins of interest including cytosolic/organellar aaRS and tRNA-processing
447 enzymes characterized in *A. thaliana* (**Supplemental Dataset 2**) were extracted and checked for
448 completeness via RBH against all protein models for a given species using BLASTP 2.15.0+ (103) and
449 custom Python/Bash scripts. After ortholog discovery via Orthofinder, full sets of protein models (not
450 just the longest) were searched for *A. thaliana* protein sets using BLASTP to ensure that narrowing of the
451 search space for Orthofinder had not resulted in loss of matching gene models. In cases where available
452 curated protein models excluded alternative transcripts, transcriptome data were obtained and

453 Transdecoder was run to predict protein-coding sequences (see **Supplemental Dataset 1** for details).
454 Note that ArgRS was not included in the ortholog search as there are no distinct organellar versions of
455 this enzyme in *A. thaliana.* One ArgRS gene product is cytosol-targeted, and the other is triple-targeted
456 to cytosol, mitochondria, and plastids (9, 16). In the case of GlnRS, there is no organellar ortholog as
457 plants typically use the bacterial-type GatCAB amidotransferase system to charge Gln-tRNA indirectly in
458 mitochondria and plastids (42). For CysRS, tRNase Z and PRORP, sequences for both cytosolic and
459 organellar proteins were clustered together in one orthogroup likely due to recent shared ancestry. For
460 these genes, cytosolic vs. organellar versions were differentiated by aligning sequences in each
461 orthogroup using MAFFT v7.525 (104) with the --auto setting, followed by trimming of the alignment
462 using trimAl v1.5.rev0 on the -automated1 setting (105). Alignments were then viewed using MEGA 11
463 (Tamura et al., 2021). Maximum-likelihood tree building was conducted using RAxML v 8.2.12 (107) as
464 follows:
465 raxmlHPC-PTHREADS-SSE3 -f a -# 100 -p 12345 -x 12345 -s <fasta file with orthogroup sequences> -n
466 <orthogroup> -m PROTGAMMALG -T 12
467 Inclusion in cytosolic versus organellar groups was then determined by relationship to known *A. thaliana*
468 cytosolic/nuclear/organellar proteins in maximum-likelihood trees (**Fig. S8-10**).
469 Once we had a curated set of protein sequences for all species, they were checked for
470 completeness of length by alignment with *A. thaliana* orthologs by MAFFT and subsequent visualization
471 with a custom Python script. A "full-length" ortholog was defined as having at least 50% alignment with
472 any *A. thaliana* protein in the orthogroup, excluding pseudogenes. Custom Python and Bash scripts are
473 available via Dryad (https://doi.org/10.5061/dryad.0cfxpnw7p).
474 ***In silico targeting predictions***
475 For TP analysis, all protein models available for each gene with similarity to the original Orthofinder hit
476 (BLASTP e-value < 1e-6) were used, not just the longest. All transcripts for cytosolic aaRS, a "positive

477 control" set of genes involved in mitochondrial oxidative phosphorylation/ plastid housekeeping and a
478 "negative control" set of genes involved in cell-wall synthesis were run through targeting prediction
479 programs LOCALIZER-1.0.5 (76) and TargetP-2.0 (75), with commands as follows:

480 LOCALIZER.py -p -i <input file>

481 targetp -fasta <input file> -mature -org pl

482 Protein models for some species were often incomplete due to the fragmented nature of some of the
483 transcriptome assemblies. To ensure models missing the N-terminus were not used in retargeting
484 analysis, predictions for models that did not align without gaps for 15 residues within 100 amino acids of
485 the start codon for at least one *A. thaliana* ortholog were discarded. For transcript-by-transcript
486 heatmaps (**Supplemental Dataset 6**), a custom R script was used to plot all transcripts for all species. For
487 Fig. S3 (which shows predicted targeting for a multitude of different proteins) only the most retargeted
488 ortholog for each species was used. Orthologs were then binned based on probability of targeting to the
489 mitochondria or plastids. We chose a probability of 0.5 as a threshold for LOCALIZER to match the
490 original publication describing the tool(76). TargetP-2.0 uses a "winner take all" strategy to declare one
491 subcellular localization based on the location with the highest likelihood score, which is not compatible
492 with assessing dual targeting (75, 108). Thus, we used the lowest likelihood score for known organelle-
493 targeted *A. thaliana* proteins from our control dataset identified by TargetP-2.0 to be plastid or
494 mitochondria targeted (~0.5 for AT1G02410 cytochrome c oxidase 11). High confidence retargeting was
495 defined as a > 0.5 probability/likelihood of targeting for both LOCALIZER and TargetP-2.0. Moderate
496 confidence was defined as a > 0.5 probability/likelihood for one of the algorithms. Low confidence
497 retargeting was defined as neither algorithm providing a probability/likelihood > 0.5. For subsequent
498 analysis of ratios (**Fig. 2, Fig. S1-S2**), any ortholog with a "moderate" or better confidence was counted
499 as retargeted. Note that *R. cantleyi* and *S. himalayana* proteins were excluded from predictions of
500 targeting to plastids as these species likely lack plastomes. All analyses including plotting and statistics

501 were carried out using custom Bash, Python, and R scripts available via Dryad

502 (https://doi.org/10.5061/dryad.6hdr7sr5x).

503 ***In vivo localization of S. himalayana aaRS transit peptides via tobacco infiltration and confocal***

504 ***microscopy***

505 TPs were chosen based on having an at least moderate probability (see above) of targeting a protein to

506 either the mitochondria or plastid and/or having an N-terminal extension preceding conserved sequence

507 shared with *A. thaliana.* We selected either the sequence predicted to be a TP by TargetP-2.0 plus 10

508 residues or the entire N-terminal overhang plus the first 10 residues aligning with the functional portion

509 of the *A. thaliana* protein. These amino acid sequences were then converted into DNA sequences using

510 EMBOSS Backtranseq (109) with the *Nicotiana tabacum* codon usage table (see **Supplemental Dataset**

511 **3**). These sequences were then used to generate constructs for *Agrobacterium* transformation using a

512 newly designed dual-expression vector containing both the mitochondrial control IVD-FP611 and GFP

513 fused to the C-terminus of our TP of interest. Some TPs were also originally cloned into the single-gene

514 expression vector described in (55). Transformation into *N. benthamiana* and confocal microscopy was

515 constructed as described in (55). With our dual expression vectors, there was occasionally low GFP

516 and/or FP611 signal relative to chlorophyll autofluorescence (see Fig. 3 GlyRS and MetRS). Thus, we

517 scaled pixel intensity using look-up tables (LUTs) for all images to better see localization for low

518 expression constructs. It is important to note that in cases where GFP fluorescence is low and LUTs are

519 scaled, overlap of the chlorophyll florescence and GFP spectra can create the false impression of plastid

520 GFP targeting (see negative control in Fig. S4 for an example). Thus, we only determined a TP to be

521 plastid targeted if we either observed 1) GFP-fluorescing stromules in transformed plastids, an

522 indication of true GFP signal as these structures occlude thylakoids or 2) Increased GFP florescence in

523 plastids of transformed cells compared to untransformed cells in the same frame. Often guard cells

524 would not be transformed, so we also used these as an in-frame negative control. Note that *in silico*

525 predictions shown are for the TP attached to GFP, not the original protein model. Relative to the whole
526 protein models, the GFP predictions yielded higher targeting probabilities and affected classification for
527 MetRS (LOCALIZER likelihood score > 0.5 for mitochondria) and AsnRS (TargetP-2.0 likelihood score > 0.5
528 for plastid), but there were no substantive changes for any of the other transit peptides. For more
529 information see supplemental dataset 3. Constructs are available from Addgene; see supplemental
530 dataset 3 for ID numbers.

531 ***Mitogenome assembly***

532 Prior to tRNA analysis, the mitogenomes of *S. himalayana* (SRR629601) and *R. cantleyi* (SRR629613)
533 were assembled *de novo* from raw reads available from the NCBI SRA database. To maximize the
534 likelihood of finding all mitogenome tRNAs, genomes were assembled in several ways. Reads were first
535 trimmed using FASTP (110), with either default parameters or "harsh" trimming where the --
536 qualified_quality_phred and --unqualified_percent_limit parameters were set to 20 and 10, respectively.
537 SPAdes v.3.15.4 (111) was then used to assemble trimmed reads *de novo,* with kmer sizes of 21, 33, 55,
538 and 77. To avoid potential assembly fragmentation due to excessive levels of sequence coverage, a
539 subset of reads (1,000,000 from S. *himalayana* and 500,000 from *R. cantleyi*) were sampled and
540 assembled from each species as well.

541 ***tRNA analysis***

542 Excepting the Rafflesiaceae, for which we assembled mitogenomes *de novo,* the GenBank (.gbk) files for
543 mitogenomes and plastomes from species of interest were downloaded from NCBI (**Supplemental**
544 **Dataset 1**). A custom Perl script was used to extract annotated tRNAs from the .gbk files. tRNAscan-SE-
545 2.0.9 (112, 113) was run on downloaded genomes (parameters -BQH), and results were cross-checked
546 with preexisting GenBank annotations (**Supplemental Dataset 4**). For *S. himalayana* and *R. cantleyi*, the
547 same tRNAscan-SE-2.0.9 pipeline was used on all generated assemblies, and the consensus set of tRNAs
548 was checked for plant origin using BLASTN against all NCBI reference sequences and the *A. thaliana*

549 mitogenome (NC_037304.1) and plastome (NC_000932.1). tRNA presence/absence patterns were
550 plotted along with the corresponding aaRS presence/absence/retargeting data using a custom Python
551 script. Scripts are available via Dryad (https://doi.org/10.5061/dryad.np5hqc009). The highly divergent
552 and truncated trnE for *B. fungosa* was added based on prior publication(46).

553 ***Rate Analysis***

554 Primary protein sequences for cytosolic and organellar aaRS from heterotroph *M. hypopitys* and related
555 autotroph *R. williamsianum* from the original Orthofinder search were used for comparison of rates,
556 with *A. thaliana* orthologs forming the outgroup. CysRS, LeuRS, ThrRS, and TrpRS were excluded from
557 the analysis because the organellar ortholog was missing or too short to create informative gene trees in
558 one or both test species. Predicted TPs were trimmed from sequences using output of TargetP-2.0.
559 Sequences were then aligned using MAFFT v7.490 (104) and trimmed manually in MEGA 11 (106).
560 Maximum likelihood trees were built using RAxML v 8.2.12 (107) as follows:

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

563 For orthogroups containing only three ortholog sequences, a “dummy” outgroup sequence represented
564 by the *A. thaliana* ortholog with one branch-specific substitution was created to overcome RAxML’s
565 input requirement of four or more taxa. The resulting trees were then evaluated for topology. In
566 instances where species-specific paralogs were not sister to one another, separate multiple sequence
567 alignments and trees were re-made for each paralog as described above. In instances where species-
568 specific paralogs were sister, both sequences were retained in the same tree. Trees were rooted to the
569 *A. thaliana* ortholog(s), then branch lengths to the last shared node from each *H. monotropa* and *R.*
570 *williamsianum* ortholog were extracted from the trees. For instances where two sister paralogs from the
571 same species were retained, the branch lengths were averaged. Then, the $\log_{10}$-transformed ratio of the
572 *H. monotropa/R williamsianum* branches was plotted, and mean branch lengths for cytosolic and

organellar enzymes were calculated. Statistical difference in mean branch ratio in cytosolic vs. organellar enzymes was tested using a *t*-test. Tajima's relative rate tests were run in MEGA 11 using *A. thaliana* as the outgroup. Data is available in **Supplemental Dataset 5.** Scripts, alignments, and maximum likelihood trees are available via Dryad (https://doi.org/10.5061/dryad.tb2rbp05r).

Acknowledgements

We thank Jessica Warren for helpful discussion and feedback on this project and Anna Pratt for helpful advice on the cloning approach. We thank undergraduate researchers Xiaorui Lou and Janna Novak for their assistance. This work was supported by grants MCB-2322154, MCB-2048407, and MCB-1933590 from the National Science Foundation (NSF) and utilized the RMACC Summit supercomputer (NSF ACI-1532235 and ACI-1532236). RAD was supported by an NSF Postdoctoral Research Fellowship in Biology (IOS-2208908). JC was supported by the University of Nebraska Foundation.

References

1. A. N. Khachane, K. N. Timmis, V. A. P. Martins Dos Santos, Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. *Mol Biol Evol* **24**, 449–456 (2007).
2. P. J. Keeling, J. D. Palmer, Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**, 605–618 (2008).
3. W. Martin, *et al.*, Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS* **99**, 12246–12251 (2002).
4. E. C. M. Nowack, *et al.*, Endosymbiotic gene transfer and transcriptional regulation of transferred genes in Paulinella chromatophora. *Mol Biol Evol* **28**, 407–422 (2011).
5. N. A. Moran, J. P. McCutcheon, A. Nakabachi, Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**, 165–190 (2008).
6. J. P. McCutcheon, N. A. Moran, Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**, 13–26 (2012).
7. V. Rajendran, P. Kalita, H. Shukla, A. Kumar, T. Tripathi, Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. *Int J Biol Macromol* **111**, 400–414 (2018).

602 8. T. Salinas-Giegé, R. Giegé, P. Giegé, TRNA biology in mitochondria. *Int J Mol Sci* **16**, 4518–4559
603 (2015).

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

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

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

611 12. S. Shikha, R. Brogli, A. Schneider, N. Polacek, TRNA biology in trypanosomes. *Chimia (Aarau)* **73**,
612 395–405 (2019).

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

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

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

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

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

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

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

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

629 21. K. H. Wolfe, W.-H. Li, P. M. Sharp, Rates of nucleotide substitution vary greatly among plant
630 mitochondrial, chloroplast, and nuclear DNAs. *Proc. Nati. Acad. Sci. USA* **84**, 9054–9058 (1987).

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

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

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

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

639 26. S. Wicke, G. M. Schneeweiss, C. W. dePamphilis, K. F. Müller, D. Quandt, The evolution of the
640 plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol Biol* **76**,
641 273–297 (2011).

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

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

646 29. J. Mergner, *et al.*, Mass-spectrometry-based draft of the Arabidopsis proteome. *Nature* **579**, 409–
647 414 (2020).

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

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

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

653 33. D. R. Smith, S. R. Asmail, Next-generation sequencing data suggest that certain
654 nonphotosynthetic green plants have lost their plastid genomes. *New Phytologist* **204**, 7–11
655 (2014).

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

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

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

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

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

666 39. G. Canino, *et al.*, Arabidopsis encodes four tRNase Z enzymes. *Plant Physiol* **150**, 1494–1502
667 (2009).

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

670 41. A. Gobert, *et al.*, A single Arabidopsis organellar protein has RNase P activity. *Nat Struct Mol Biol*
671 **17**, 740–744 (2010).

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

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

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

678 45. R. Giegé, G. Eriani, The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Res*
679 **51**, 1528–1570 (2023).

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

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

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

687 49. X. Chen, *et al.*, Comparative Plastome Analysis of Root-and Stem-Feeding Parasites of Santalales
688 Untangle the Footprints of Feeding Mode and Lifestyle Transitions. *Genome Biol Evol* **12**, 3663–
689 3676 (2019).

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

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

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

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

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

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

703 56. B. Bölter, En route into chloroplasts: preproteins' way home. *Photosynth Res* **138**, 263–275
704 (2018).

705 57. P. Chotewutmontri, K. Holbrook, B. D. Bruce, "Plastid Protein Targeting: Preprotein Recognition
706 and Translocation" in *International Review of Cell and Molecular Biology*, (Elsevier Inc., 2017), pp.
707 227–294.

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

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

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

715 61. W. Neupert, J. M. Herrmann, Translocation of proteins into mitochondria. *Annu Rev Biochem* **76**,
716 723–749 (2007).

717 62. M. H. Chen, L. F. Huang, H. M. Li, Y. R. Chen, S. M. Yu, Signal peptide-dependent targeting of a
718 rice α-amylase and cargo proteins to plastids and extracellular compartments of plant cells. *Plant*
719 *Physiol* **135**, 1367–1377 (2004).

720 63. L. G. L. Richardson, D. J. Schnell, Origins, function, and regulation of the TOC-TIC general protein
721 import machinery of plastids. *J Exp Bot* **71**, 1226–1238 (2020).

722 64. H. Inoue, C. Rounds, D. J. Schnell, The molecular basis for distinct pathways for protein import
723 into Arabidopsis chloroplasts. *Plant Cell* **22**, 1947–1960 (2010).

724 65. S. Bischof, *et al.*, Plastid proteome assembly without Toc159: Photosynthetic protein import and
725 accumulation of N-Acetylated plastid precursor proteins. *Plant Cell* **23**, 3911–3928 (2011).

726 66. Y. L. Chen, *et al.*, TIC236 links the outer and inner membrane translocons of the chloroplast.
727 *Nature* **564**, 125–129 (2018).

728 67. A. R. Kasmati, *et al.*, Evolutionary, Molecular and Genetic Analyses of Tic22 Homologues in
729 Arabidopsis thaliana Chloroplasts. *PLoS One* **8** (2013).

730 68. J. Tripp, *et al.*, Structure and conservation of the periplasmic targeting factor Tic22 protein from
731 plants and cyanobacteria. *Journal of Biological Chemistry* **287**, 24164–24173 (2012).

732 69. M. Rudolf, *et al.*, In vivo function of Tic22, a protein import component of the intermembrane
733 space of chloroplasts. *Mol Plant* **6**, 817–829 (2013).

734 70. A. R. Kasmati, M. Töpel, R. Patel, G. Murtaza, P. Jarvis, Molecular and genetic analyses of Tic20
735 homologues in Arabidopsis thaliana chloroplasts. *Plant Journal* **66**, 877–889 (2011).

736 71. E. Kovács-Bogdán, J. P. Benz, J. Soll, B. Bölter, Tic20 forms a channel independent of Tic110 in
737 chloroplasts. *BMC Plant Biol* **11** (2011).

738 72. H. Liu, A. Li, J. D. Rochaix, Z. Liu, Architecture of chloroplast TOC–TIC translocon supercomplex.
739 *Nature* **615**, 349–357 (2023).

740 73. X. Guo, H. Wang, D. Lin, Y. Wang, X. Jin, Cytonuclear evolution in fully heterotrophic plants:
741 lifestyles and gene function determine scenarios. *BMC Plant Biol* **24**, 989 (2024).

742 74. C. Carrie, J. Whelan, Widespread dual targeting of proteins in land plants: When, where, how and
743 why. *Plant Signal Behav* **8** (2013).

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

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

748 77. M. Sharma, B. Bennewitz, R. B. Klösgen, Rather rule than exception? How to evaluate the
749 relevance of dual protein targeting to mitochondria and chloroplasts. *Photosynth Res* **138**, 335–
750 343 (2018).

751 78. L. Xu, C. Carrie, S. R. Law, M. W. Murcha, J. Whelan, Acquisition, conservation, and loss of dual-
752 targeted proteins in land plants. *Plant Physiol* **161**, 644–662 (2013).

753 79. D. B. Sloan, J. M. Warren, A. M. Williams, S. A. Kuster, E. S. Forsythe, Incompatibility and
754 Interchangeability in Molecular Evolution. *Genome Biol Evol* **15** (2023).

755 80. Y. R. Li, M. J. Liu, Prevalence of alternative AUG and non-AUG translation initiators and their
756 regulatory effects across plants. *Genome Res* **30**, 1418–1433 (2020).

757 81. S. K. Tanz, I. Castleden, I. D. Small, A. Harvey Millar, Fluorescent protein tagging as a tool to
758 define the subcellular distribution of proteins in plants. *Front Plant Sci* **4** (2013).

759 82. J. Jeong, B. Moon, I. Hwang, D. W. Lee, GREEN FLUORESCENT PROTEIN variants with enhanced
760 folding are more efficiently imported into chloroplasts. *Plant Physiol* **190**, 238–249 (2022).

761 83. M. Sharma, B. Bennewitz, R. B. Klösgen, Dual or not dual?—comparative analysis of fluorescence
762 microscopy-based approaches to study organelle targeting specificity of nuclear-encoded plant
763 proteins. *Front Plant Sci* **9** (2018).

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

766 85. E. Skippington, T. J. Barkman, D. W. Rice, J. D. Palmer, Miniaturized mitogenome of the parasitic
767 plant viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. *PNAS*
768 **112**, E3515–E3524 (2015).

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

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

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

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

777 90. I. Small, *et al.*, The Strange Evolutionary History of Plant Mitochondrial tRNAs and Their
778 Aminoacyl-tRNA Synthetases. *J Hered* **90:3**, 333–337 (1999).

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

781 92. M. Safro, N. Moor, O. Lavrik, Phenylalanyl-tRNA Synthetases. *Madame Curie Bioscience Database*
782 (2013).

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

785 94. A. Schon, *et al.*, The RNA required in the first step of chlorophyll biosynthesis is a chloroplast
786 glutamate transfer-RNA. *Nature* **322**, 281–284 (1986).

787 95. X. Chen, *et al.*, Balanophora genomes display massively convergent evolution with other extreme
788 holoparasites and provide novel insights into parasite–host interactions. *Nat Plants* (2023).
789 https://doi.org/10.1038/s41477-023-01517-7.

790 96. A. Schon, C. G. Kannangara, S. Gough, D. Sollt, Protein biosynthesis in organelles requires
791 misaminoacylation of tRNA. (1988).

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

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

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

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

802 101. B. J. Haas, Transdecoder v5.5.0. (2018). Available at:
803 https://github.com/TransDecoder/TransDecoder [Accessed 14 May 2023].

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

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

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

809 105. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment
810 trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

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

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

815 108. O. Emanuelsson, H. Nielsen, S. Brunak, G. Von Heijne, Predicting subcellular localization of
816 proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**, 1005–1016 (2000).

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

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

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

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

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

827

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

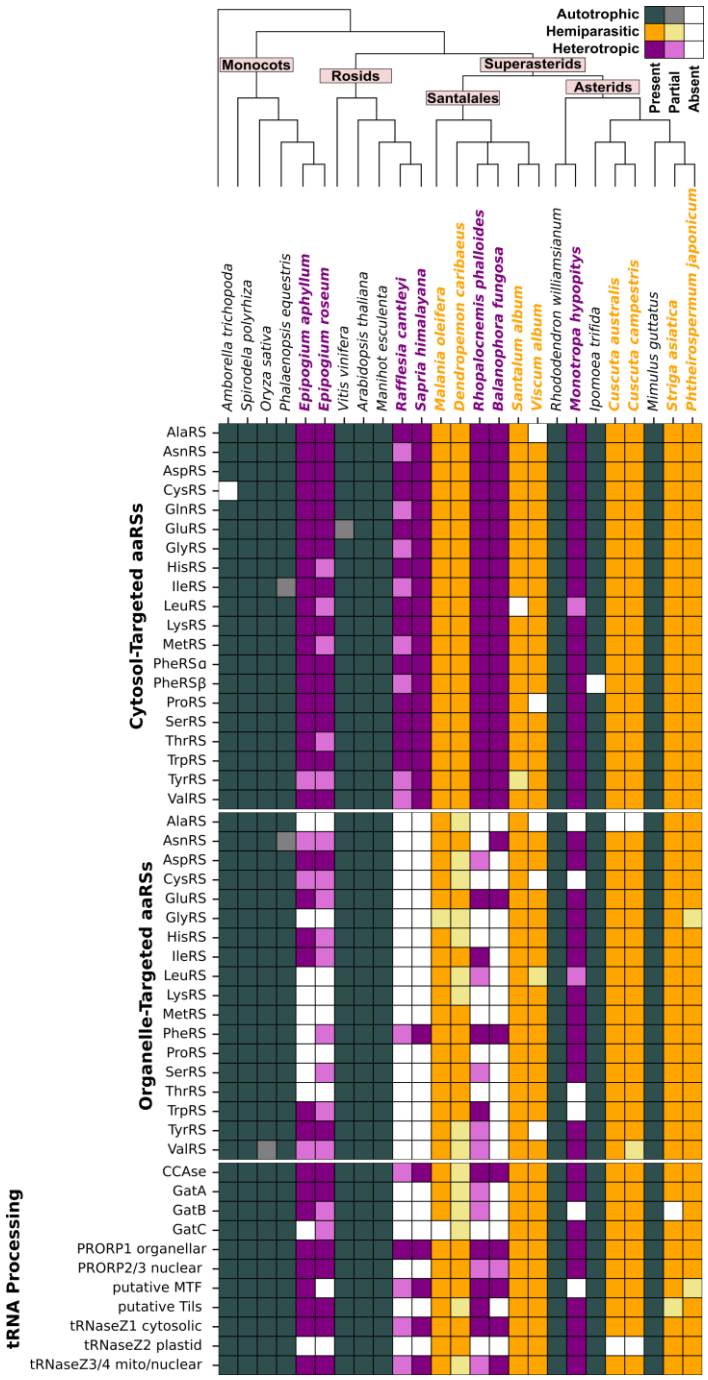


Figure 1: Presence/absence matrix of selected organellar and cytosolic aaRS and tRNA-processing enzymes shows major losses of organellar enzymes in heterotrophic lineages. Hue of boxes indicates metabolism (gray = autotroph, orange = hemiparasite, purple = heterotroph), shade indicates degree of presence/absence (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.

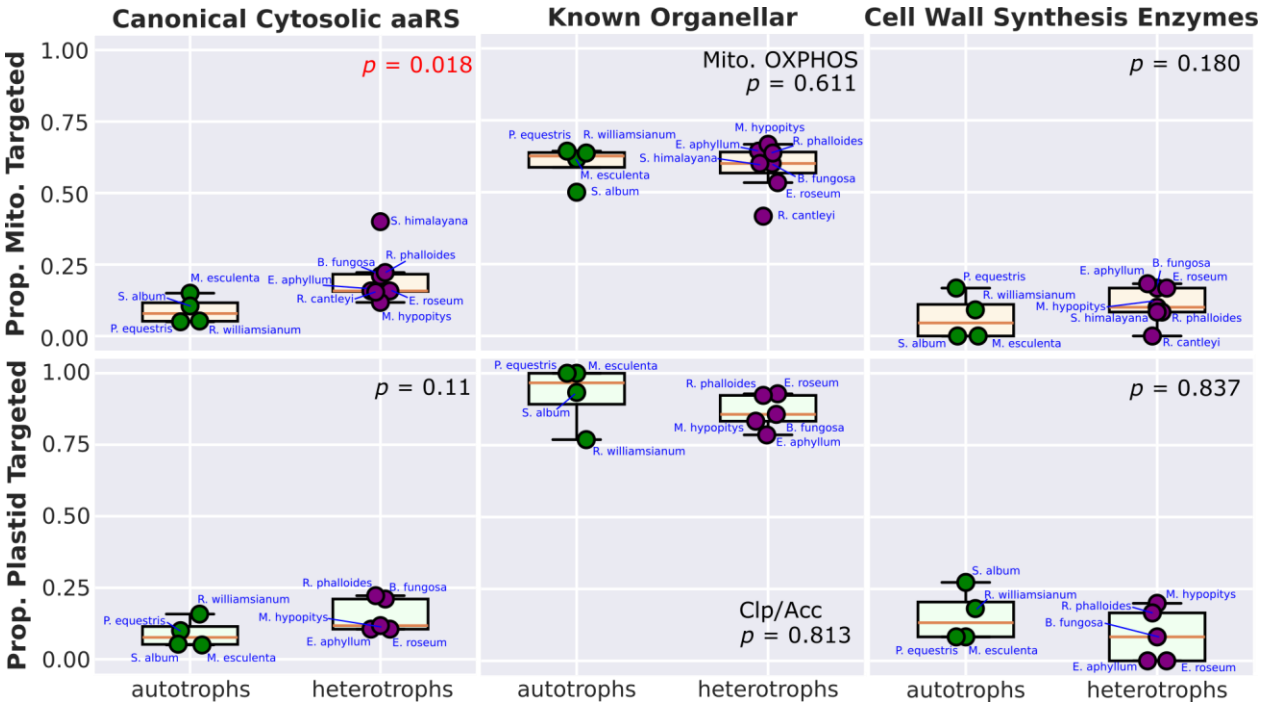


Figure 2: Proportion of retargeted cytosolic aaRS is higher in heterotrophs than autotrophs based on *in-silico* predictions. Proportion of enzymes with predicted retargeting (probability ≥ 0.5) to either mitochondria (top row) or plastids (bottom row) based on the outputs of LOCALIZER and/or TargetP. Autotrophs and heterotrophs are plotted in green and purple, respectively. Targeting predictions for proteins involved in mitochondrial oxidative phosphorylation (OXPHOS) are included as a positive control for mitochondrial targeting, and proteins involved in non-photosynthesis related housekeeping roles (ACC and Clp subunits) are included as a positive control for plastid 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, but they were predicted to have retargeted 7.7% and 30.0% of their cytosolic aaRS to the plastid, respectively. These species have lost too many plastid housekeeping genes to assess proportion targeted. For both species, no cell wall synthesis enzymes exhibited plastid retargeting.

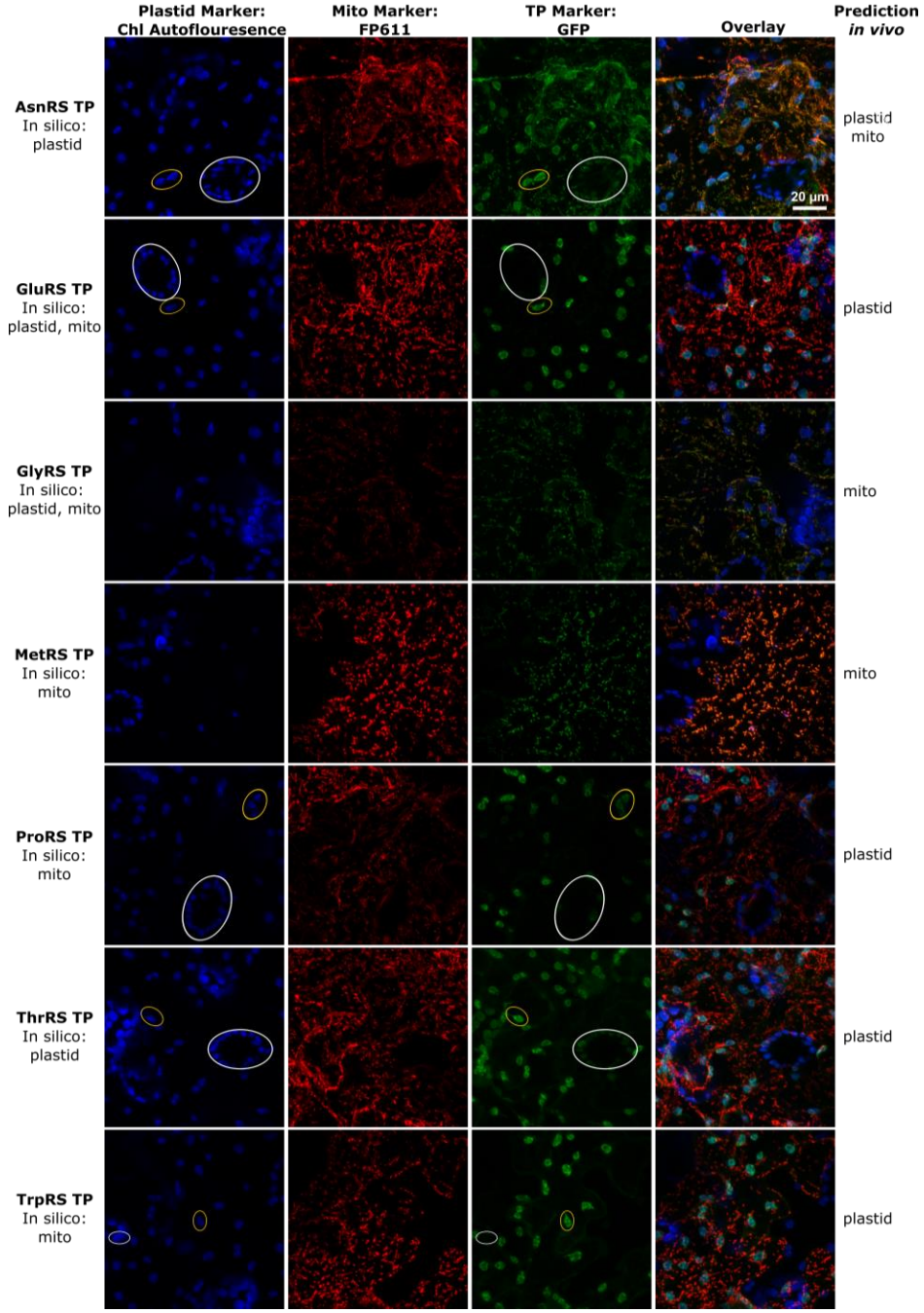


Figure 3: Cytosolic-type aaRS transit peptides from heterotroph *S. himalayana* can retarget enzymes to organelles in *Nicotiana benthamiana*. Transit peptides (TPs) from *S. himalayana* fused to GFP (middle right column) were transiently expressed in *N. benthamiana.* Chlorophyll autofluorescence (far 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). The overlay of all three channels is shown on far right. Scale indicated in top right panel is the same for all images. As a means of differentiating true GFP florescence from chlorophyll autofluorescence, yellow ovals are shown in panels to indicate plastids in transformed cells, whereas white ovals indicate plastids in non-

transformed cells. Note the presence of GFP-fluorescing stromules in transformed plastids, an indication of true GFP signal. For more information, see methods and supplemental dataset 3.

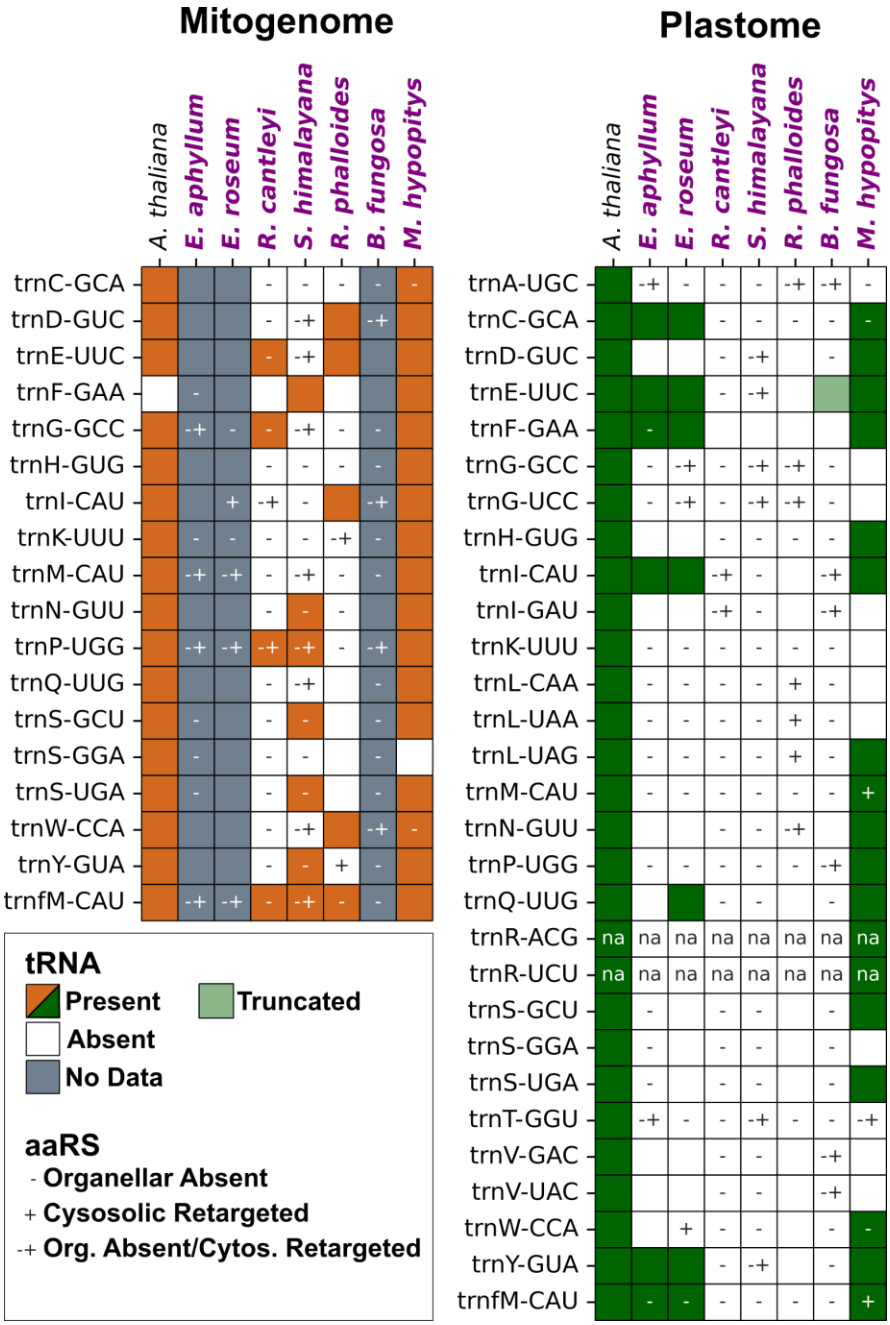


Figure 4: Instances of both concordance and mismatch between organellar tRNA gene loss and aaRS loss/retargeting. Heterotrophic species are written in purple. *Arabidopsis thaliana* is included for reference. Left: mitogenome tRNAs. Right: plastome tRNAs. Color of boxes indicates presence/absence of tRNA. The – and + annotation denotes loss of corresponding organellar aaRS and retargeting of cytosolic tRNA respectively as described for Fig. 2 and Fig. S2. Retargeting predictions are specific 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*.

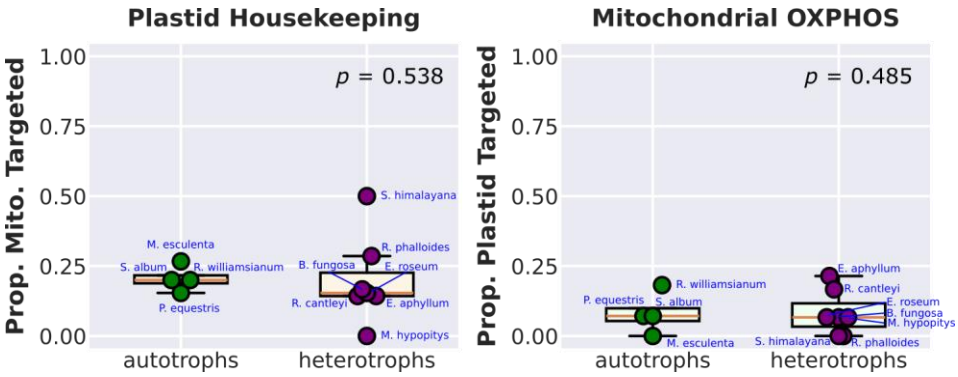
Supplemental Table 1: Count of Orthogroups Containing Riboprotein Ortholog(s)

Species	Plastid Specific	Mito. Specific	Cytosol Specific
Sapria himalayana	1	64	77
Rafflesia cantleyi	0	56	75
Arabidopsis thaliana	32	66	82

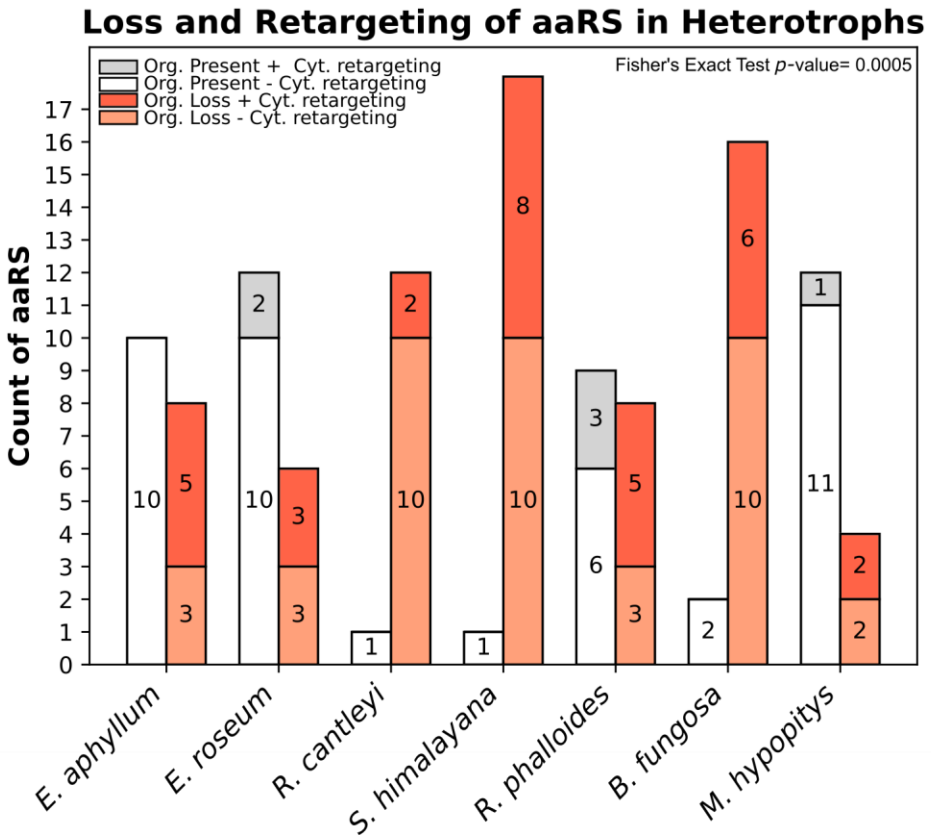
Supplemental Table 2: aaRS targeting and recovery of control groups

Protein Class	Mean % Mito-targeted		Mean % Plastid-targeted	
	Autotrophs	Heterotrophs	Autotrophs	Heterotrophs
Cytosolic aaRS	8.9%	20.3%	9%*	15.2%*
Mitochondrial OXPHOS	59.9%	58.5%	8.1%	8.3%
Plastid Housekeeping	20.5%	19.9%	92.6%*	86.6%*
Cell Wall Synth.	6.4%	11.2%	15.5%*	9%*

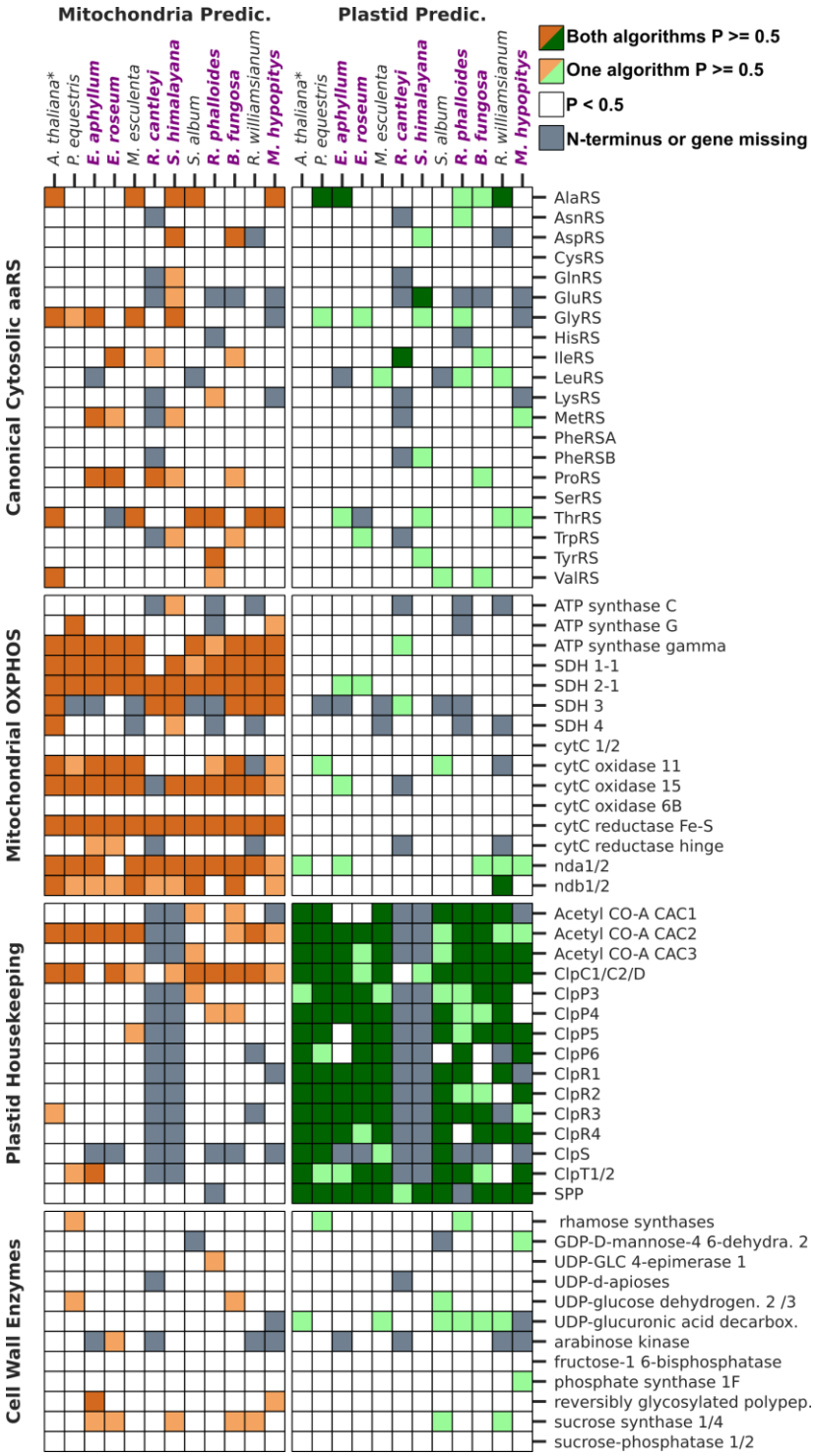
***Rafflesiaceae not included in calculation**



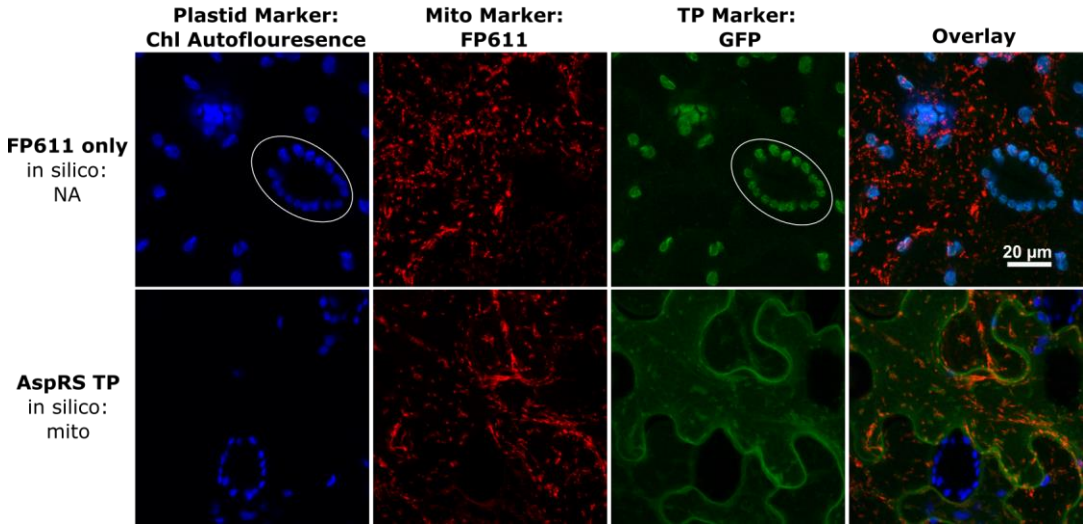
Supplemental Figure 1: Proportion of Plastid Housekeeping Genes with Predicted Mitochondrial Retargeting (Left), and Proportion of Mitochondrial OXPHOS proteins with Predicted Plastid Retargeting (Right). Proportion of enzymes with predicted retargeting (probability ≥ 0.5) based on the outputs of LOCALIZER and/or TargetP. Autotrophs and heterotrophs are plotted in green and purple, respectively. 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. It should be noted that proportions of plastid housekeeping genes in the Rafflesiaceae represent predictions for only two proteins; all other proteins are missing from these lineages, see Fig. S3.



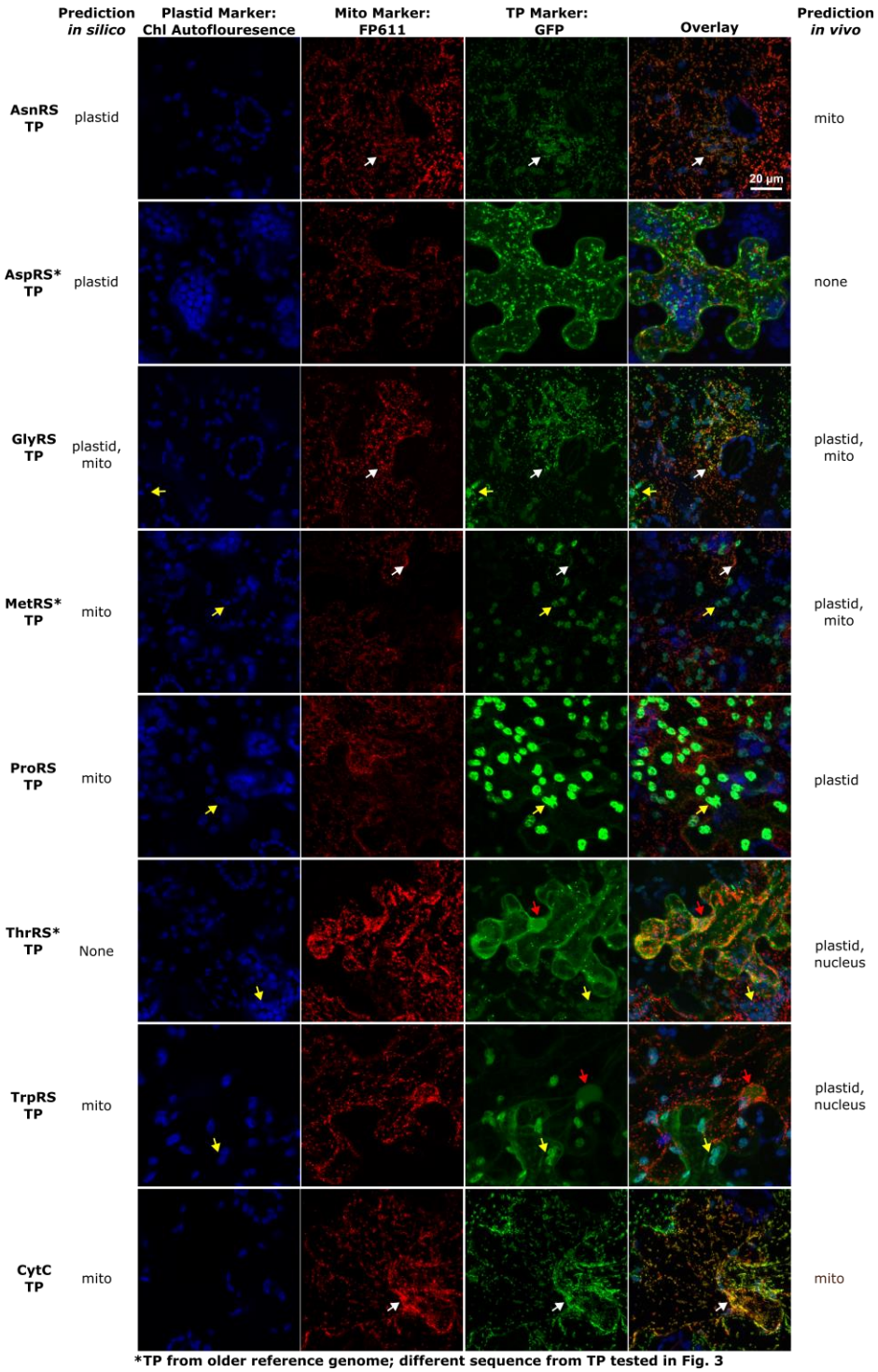
Supplemental Figure 2: Significant overlap between loss of organellar aaRS and retargeting of cytosolic counterpart in heterotrophs. Reported counts refer to number of aaRS "families" defined by amino acids (e.g., GluRS, PheRS, TrpRS, etc.). Organellar gene presence/absence determined as for Fig. 1. Partial sequences for organellar enzymes are counted as present. Predicted retargeting (probability ≥ 0.5) 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 = darker shading. No retargeting = lighter shading. 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 3: Targeting predictions for cytosolic-type aaRS. Deeper shades of green (plastid, right) and orange (mitochondrion, left) indicate probability of retargeting to that organelle is ≥ 0.5 based on both LOCALIZER and TargetP. Lighter shades of orange/green indicate one of the two algorithms yielded a probability ≥ 0.5. White indicates probability for both algorithms was < 0.5. 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. **Arabidopsis thaliana* is included as a reference.

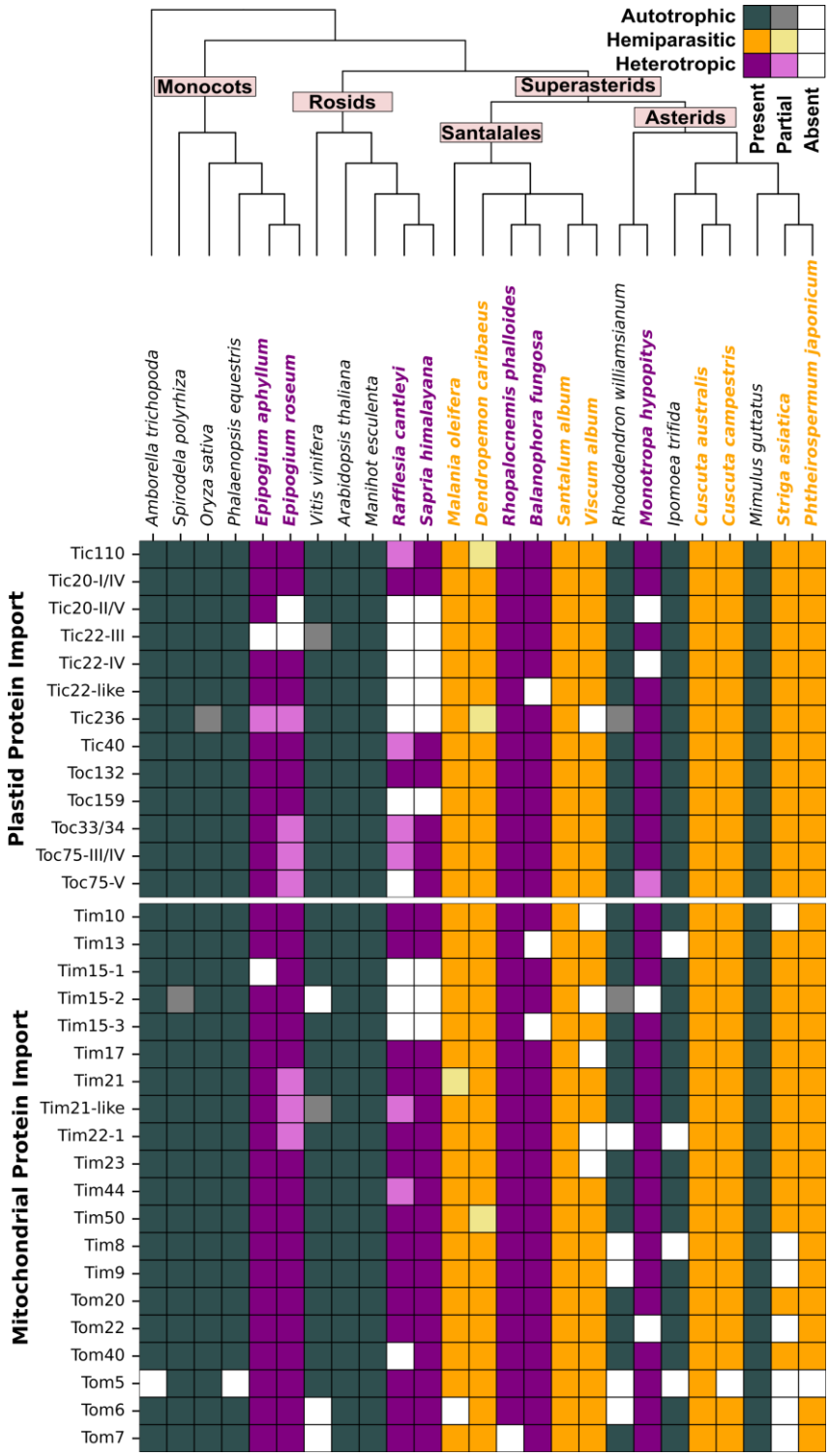


Supplemental Figure 4: Cytosolic-type aaRS transit peptides from heterotroph *S. himalayana* transiently expressed in *Nicotiana benthamiana*. Transit peptides (TPs) from *S. himalayana* fused to GFP (middle right column) were transiently expressed in *N. benthamiana.* Chlorophyll autofluorescence (far 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 (IVD) (middle left). Overlay of all three channels on far right. Top row: IVD_FP611 infiltration only- negative control. White ovals indicate plastids in untransformed cells, allowing assessment of bleed-over chlorophyll autofluorescence in the GFP channel. See main text for how true plastid localization was distinguished from bleed-over signal. Bottom row: Expression of AspRS TP led to apparent aggregates of GFP in the cytosol, potentially resulting from an artefact due to overexpression. Scale indicated in top right panel is the same for all images.

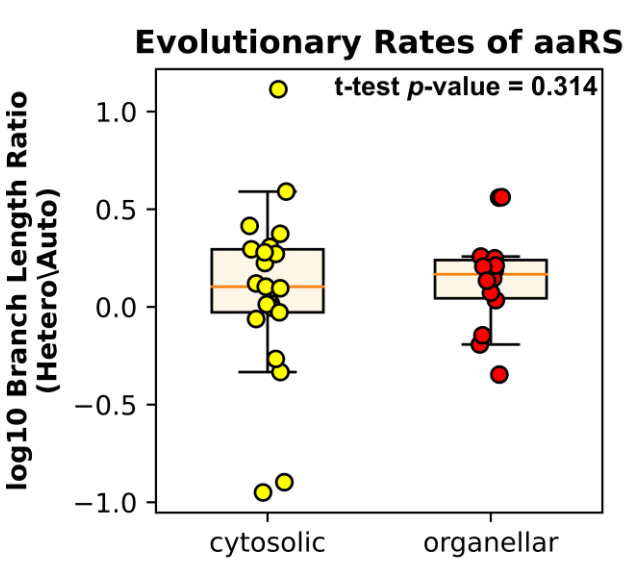


Supplemental Figure 5: Initial round of microscopy data using single expression vector PK7FWG2. Transit peptides (TPs) from *S. himalayana* fused to GFP (middle right column) were transiently expressed in *N. benthamiana.* Chlorophyll autofluorescence (far left column) was used for colocalization with plastids. As a positive control for mitochondrial targeting, eqFP611 was tagged with the TP from mitochondrial isovaleryl-CoA dehydrogenase (middle left). The overlay of all three channels on far right. Arrows indicate

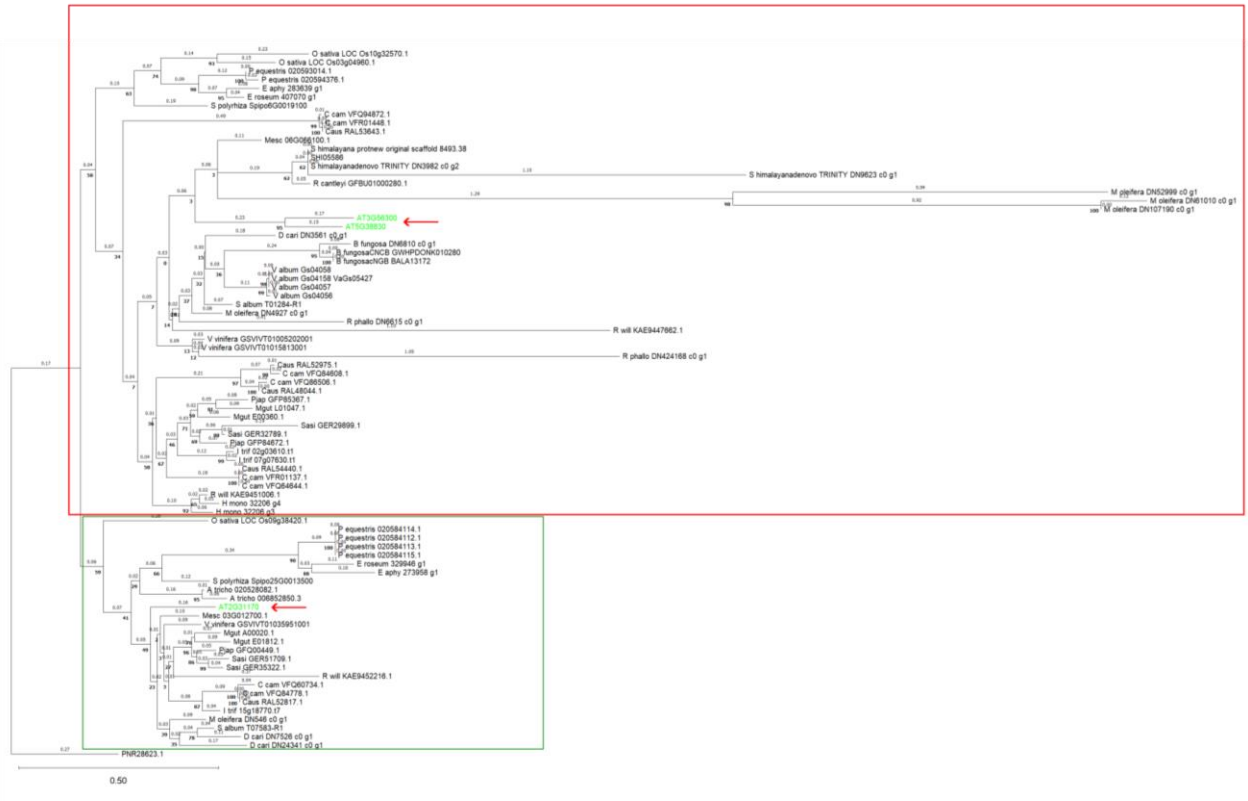
examples of plastid (yellow), mitochondrial (white) and potential nuclear (red) localization. Scale indicated in top right panel is the same for all images.



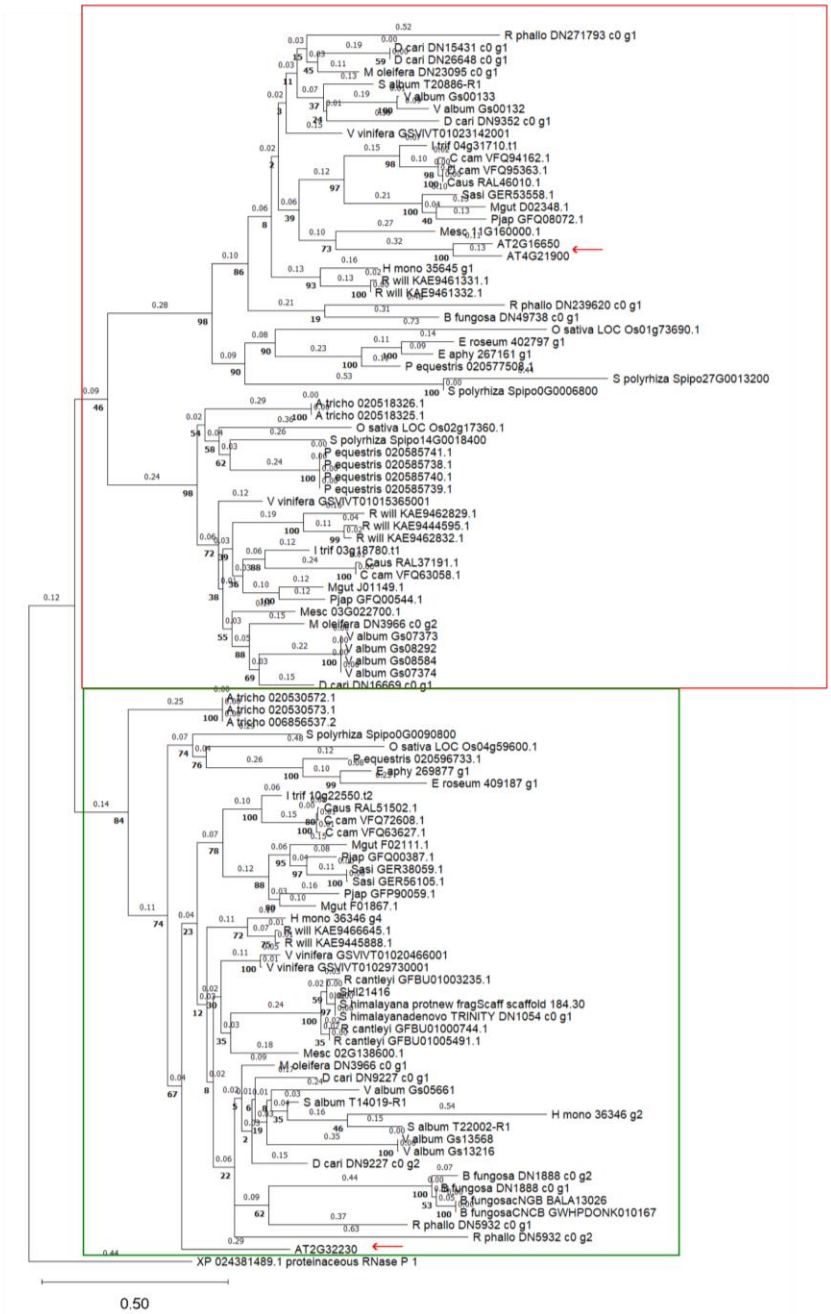
Supplemental Figure 6: Loss of key subunits for plastid import apparatus in *Rafflesiaceae*. For complete names of proteins and corresponding *A. thaliana* gene ID, see supplemental dataset 2. Hue of boxes indicates metabolism (gray = autotroph, orange = hemiparasite, purple = heterotroph), shade indicates degree of presence/absence (white = no sequence found, light colors = partial sequence found, dark colors = complete or near complete sequence found).



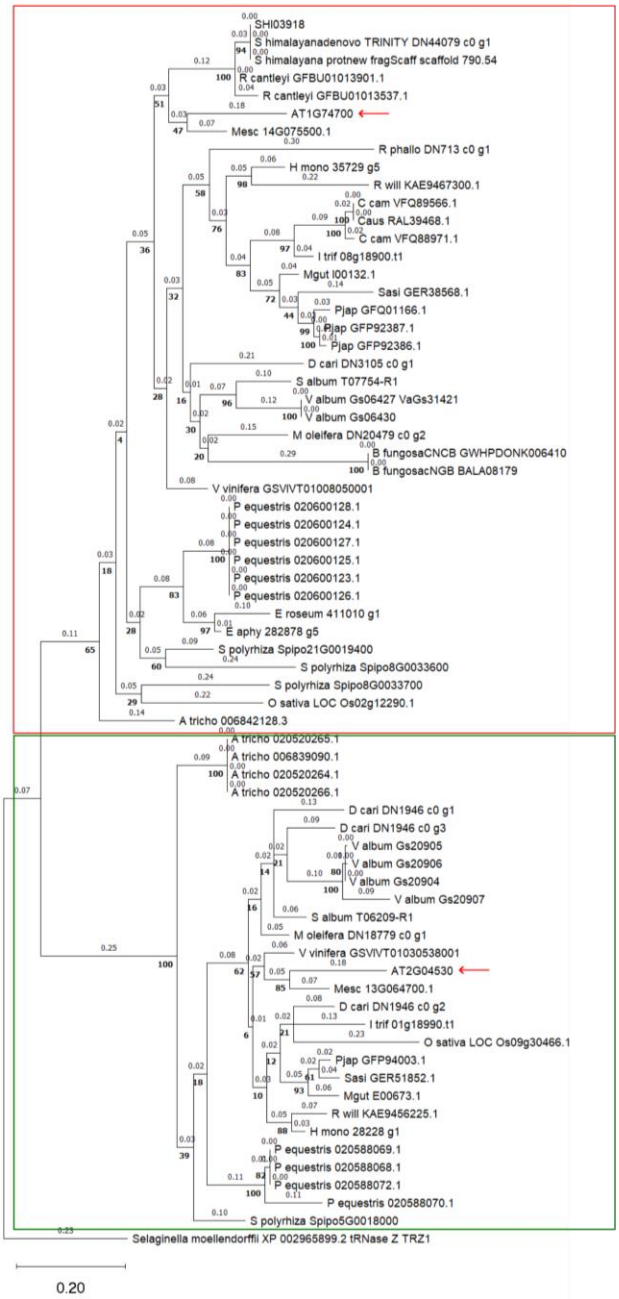
Supplemental Figure 7: Relative rates of sequence evolution of cytosolic and organellar aaRS are similar in heterotroph compared to autotrophic relative. Each point represents an aaRS branch length of *M. hypopitys* (heterotrophic) normalized to branch length of *R. williamsianum* (autotrophic) for canonically cytosolic and organellar aaRSs ($\log_{10}$ scale).



Supplemental Figure 8: 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 red.



Supplemental Figure 9: 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 red.



Supplemental Figure 10: 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 red, and orthologs for *A. thaliana* plastid tRNase Z2 (AT2G04530) are outlined in green.