Dual targeting of aminoacyl-tRNA synthetases to the mitochondrion and complex plastid in chlorarachniophytes

Yoshihisa Hirakawa, Fabien Burki and Patrick J. Keeling*

Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada

*Author for correspondence (pkeeling@mail.ubc.ca)

Accepted 18 September 2012
Journal of Cell Science 125, 6176–6184

doi: 10.1242/jcs.116533

Summary

In plants, many nucleus-encoded proteins are targeted to both mitochondria and plastids, and this process is generally mediated by ambiguous N-terminal targeting sequences that are recognized by receptors on both organelles. In many algae, however, plastids were acquired by secondarily engulfing green or red algae, which were retained within the endomembrane system. Protein targeting to these secondary plastids is more complex, and because they do not reside directly in the cytoplasm, dual targeting cannot function as it does in plant cells. Here we investigate dual targeting of aminoacyl-tRNA synthetases (aaRSs) in chlorarachniophytes, which are complex algae that possess secondary plastids and a relict nucleus derived from a green algal endosymbiont. Chlorarachniophytes have four genome-containing compartments, but almost all the aaRSs are nucleus-encoded and present in fewer than four copies (some as few as two), suggesting multiple targeting. We characterized the subcellular localization of two classes, HisRS (three copies) and GlyRS (two copies), using GFP fusion proteins. In both cases, one copy was dually targeted to mitochondria and plastids, but unlike plants this was mediated by translation initiation variants. We also found that the periplastidal compartment (the relict green algal cytoplasm) lacks both GlyRS and a cognate tRNA, suggesting that pre-charged host tRNAs are imported into this compartment. Leader analysis of other aaRSs suggests that alternative translation is a common strategy for dual targeting in these complex cells. Overall, dual targeting to mitochondria and plastids is a shared feature of plastid-bearing organisms, but the increased complexity of trafficking into secondary plastids requires a different strategy.

Key words: Endosymbiosis, Protein targeting, Chlorarachniophyte, Plastid, Mitochondrion

Introduction

In plant and algal cells, protein synthesis occurs in three subcellular compartments: the cytoplasm, mitochondria, and plastids. Each of these compartments requires a full set of transfer RNAs (tRNAs) and aminoacyl-tRNA synthetases (aaRSs), which catalyze the attachment of an amino acid to the cognate tRNA. In vascular plants, nuclear and plastid genomes encode all tRNAs necessary for translation (Maréchal-Drouard et al., 1993). Mitochondrial genomes carry most tRNA genes, but a number of nucleus-encoded tRNAs are also imported from the cytoplasm into the mitochondria (Rubio and Hopper, 2011). Contrary to tRNA genes, plastid and mitochondrial genomes completely lack aaRS genes, and the nuclear genome encodes all organelle-targeted aaRSs as well as cytoplasmic aaRSs (Duchêne et al., 2009). For most nucleus-encoded organelle proteins, N-terminal pre-sequences facilitate the targeting of proteins specifically to either mitochondria or plastids (Schleiff and Becker, 2011). But in *Arabidopsis thaliana*, many organelle aaRSs have ambiguous N-terminal targeting sequences that are recognized by receptors of both mitochondria and plastids, so there are far fewer aaRS genes than there would have been originally (Duchêne et al., 2005; Pujol et al., 2008; Berglund et al., 2009). These dual-targeted aaRS genes have diverse evolutionary origins; one-third appear to be derived from the cyanobacterial endosymbiont (plastid-lineage), while others are closely related to proteobacteria (mitochondrial-lineage) or other bacteria (Brandão and Silva-Filho, 2011).

Ambiguous organelle targeting sequences work in plants because their plastid and mitochondrion have similar characteristics in membrane structure and protein sorting system (Carrie et al., 2009); proteins are post-translationally imported from the cytoplasm into organelles across two envelope membranes. However, in several algal groups, an additional layer of complexity is added by the evolutionary history of their plastids. In these algae, plastids were acquired secondarily by the endosymbiotic uptake of a green or red alga, and their plastids (secondary plastids) are bounded by three or four membranes and reside within the endomembrane system (Archibald, 2009; Keeling, 2010). Thus, their nucleus-encoded plastid proteins are co-translationally transported via endoplasmic reticulum (ER) (Bolte et al., 2009), while mitochondrial proteins are not. This implies that dual targeting to mitochondria and secondary plastids would be more complex than the plant system. Chlorarachniophytes are one group of unicellular marine algae with secondary plastids derived from a green alga. Their plastids are surrounded by four membranes, and further distinguished by retaining a relict nucleus of the endosymbiont, referred to as the nucleomorph, in a residual cytoplasm called the periplastidal compartment (PPC) (McFadden et al., 1994). Thus, chlorarachniophytes have four genomes and four translationally

active compartments: the cytoplasm, mitochondria, plastids and PPC.

All three organelle genomes have been sequenced for the model chlorarachniophyte, *Bigelowiella natans*, and complete or near-complete sets of tRNA genes are found in each; the plastid genome encodes a full set of tRNAs (Rogers et al., 2007), the mitochondrial genome lacks only $tRNA^{Thr}$ (partial sequence GenBank HQ840955), and the nucleomorph genome lacks $tRNA^{Pro}$, $tRNA^{Ala}$, $tRNA^{Arg}$ and $tRNA^{Gly}$ (Gilson et al., 2006). In contrast, aaRS genes are completely absent in the plastid and mitochondrial genomes, and only a single aaRS gene (seryl-tRNA synthetase) is encoded in the nucleomorph genome (Gilson et al., 2006). Organelle aaRSs therefore must be encoded by the nuclear genome and imported from the cytoplasm. In this study, we identified all nucleus-encoded aaRS genes in the draft nuclear genome of *B. natans* CCMP2755 (sequenced by DOE Joint Genome Institute, and recently published by Curtis et al., 2012), and analyzed the subcellular localization and phylogenetic position of aaRSs of the most extremely reduced families, histidyl- and glycyl-tRNA synthetase (HisRS and GlyRS). We found three copies of HisRS and two of GlyRS in the genome. Two HisRSs were targeted to the cytoplasm and PPC, respectively, whereas the third was dually targeted to plastids and mitochondria. In the case of GlyRSs, one localized to the cytoplasm, and the other is dually targeted to the plastids and mitochondria. Dually targeted proteins are expressed from identical mRNAs and site directed mutagenesis shows that alternative translation initiation likely drives alternative targeting. No GlyRS was targeted to the PPC. Since the nucleomorph genome encodes no $tRNA^{Gly}$ gene, we suggest that the PPC imports charged $tRNA^{Gly}$ rather than GlyRS protein. Altogether, this shows that dual targeting to mitochondria and plastids remains a common feature in complex secondary algae, despite the increased complexity of trafficking, but that the dominant strategies used to effect dual targeting in these algae are apparently different from those most common in plants.

Results

Characterization of *B. natans* HisRSs and GlyRSs

We surveyed nucleus-encoded aaRS genes using BlastP against the draft nuclear genome sequence of *B. natans*, and found 63 candidates corresponding to all 20 amino acids (AA). Although four copies of each aaRS are theoretically needed, between two to five copies were found, with half being present in three copies. We focused on HisRS and GlyRS due to their low copy number; three copies of HisRS genes (JGI protein ID No. 77124, No. 85238 and No. 84033) and two of GlyRS genes (No. 87924 and No. 87243). To clarify the N-terminus, and therefore likely the translation start of each aaRSs, we first performed RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE). A single band was obtained for both GlyRS genes and two HisRS genes, but for HisRS No. 85238 two different cDNA products were observed (Fig. 1A); the sequence of the lower band was 471 bases shorter than the upper band in the 5′ region, and the predicted translation start of lower band was at 127 amino acids (AA) downstream of upper product. Three HisRSs (No. 77124, No. 85238 and No. 84033) and one GlyRS (No. 87924) exhibited obvious N-terminal extensions beyond the predicted mature protein.

To predict the subcellular localization of each aaRS protein, we used four prediction servers, TargetP 1.1 (Emanuelsson et al., 2000), SignalP 3.0 (Bendtsen et al., 2004), ChloroP 1.1 (Emanuelsson et al., 1999) and Predotar 1.03 (Small

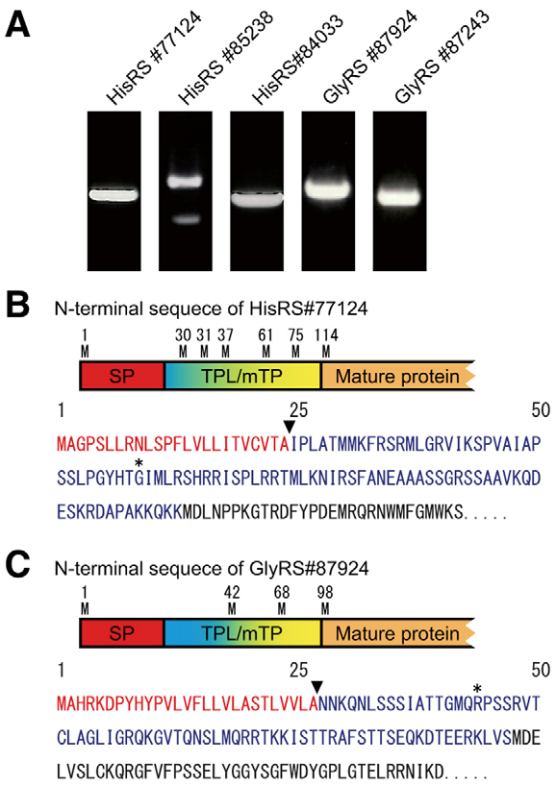

Fig. 1. Amplification of 5′ cDNA ends for *B. natans* HisRS and GlyRS genes. (**A**) Agarose gel electrophoreses of 5′ RLM-RACE products for five aaRS genes. (**B**,**C**) N-terminal amino acid sequences of HisRS No. 77124 and GlyRS No. 87924. M indicates the position of methionine residues. Predicted signal peptide (SP) sequences are shown in red characters (arrowheads show the predicted cleavage sites), and putative transit peptide-like (TPL) and mitochondrial targeting peptide (mTP) sequences are shown in blue. The mature proteins correspond to the conserved sequence regions among homologous aaRSs. Asterisks indicate the positions of the first splice junction of introns.

et al., 2004). Previous studies have demonstrated that chlorarachniophyte plastid-targeted proteins possess an N-terminal bipartite targeting pre-sequence consisting of an endoplasmic reticulum (ER)-targeted signal peptide (SP) followed by a chloroplast transit peptide-like (TPL) sequence that together mediate targeting to plastids via the ER (Rogers et al., 2004; Hirakawa et al., 2009). HisRS No. 77124 and GlyRS No. 87924 were both predicted to encode a SP followed by a TPL (Table 1; Fig. 1B,C), with ChloroP score of 0.575 and 0.522, respectively. This suggests that these two aaRSs are likely targeted to the plastids. Interestingly, however, both proteins were also predicted to be localized to mitochondria if translation initiated at the 2nd possible methionine (Met), as opposed to the first (Table 1). HisRS No. 84033 was predicted to be targeted to mitochondria by TargetP, but its N-terminal extension consisted of a potential bipartite targeting sequence; TMHMM 2.0 (Krogh et al., 2001) predicted a hydrophobic transmembrane helix reminiscent of a SP between amino acids 29 to 51, and a putative TPL sequence was predicted to follow with ChloroP score 0.513. This implies that HisRS No. 84033 might also be targeted to the plastid. The remaining two proteins, HisRS No. 85238 (both upper and lower bands) and GlyRS No. 87243, encoded no identifiable targeting sequence and were predicted to be cytoplasmic (Table 1).

Table 1. Prediction of subcellular localization of *B. natans* aaRSs

Protein ID	Translation start	TargetP		Predotar		Experimental localization
		SP	mTP	SP	mTP	
HisRS No. 77124	1st Met (1 AA)	0.972	0.038	0.98	0.02	Plastids
	2nd Met (30 AA)	0.060	0.879	0.01	0.91	Mitochondria
	7th Met (114 AA)	0.048	0.122	0.01	0.00	Cytoplasm
HisRS No. 85238	Upper band (1 AA)*	0.061	0.056	0.01	0.00	Cytoplasm
	Lower band (127 AA)*	0.041	0.237	0.01	0.01	Cytoplasm
HisRS No. 84033	1st Met (1 AA)	0.156	0.958	0.01	0.04	PPC
GlyRS No. 87924	1st Met (1 AA)	0.940	0.024	0.99	0.00	Plastids
	2nd Met (42 AA)	0.020	0.758	0.01	0.58	Mitochondria
	4th Met (98 AA)	0.097	0.070	0.01	0.00	Cytoplasm
GlyRS No. 87243	1st Met (1 AA)	0.053	0.051	0.01	0.00	Cytoplasm

Each number of columns indicates the estimated probability. The numbers in parentheses show the distance from the first methionines (Met). SP, signal peptide; mTP, mitochondrial targeting peptide; AA, amino acid; PPC, periplastidal compartment.

*Upper and lower bands correspond to the RLM-RACE result shown in Fig. 1A.

In vivo subcellular localization of *B. natans* HisRSs and GlyRSs

To demonstrate the subcellular localization of three HisRSs and two GlyRSs, we used the transient genetic transformation system of the chlorarachniophyte, *Amorphochlora amoebiformis*, with a variety of constructs fused to green fluorescent protein (GFP). First, we expressed the full-length N-terminal extension of HisRS No. 77124 fused to GFP (HisRS No. 77124-1stMet+GFP). In the transformed cells, GFP fluorescence was observed in the plastids and in several small blobs that did not co-localize with chlorophyll-autofluorescence (Fig. 2A). When the 2nd (30 AA position), 3rd (31 AA), and 4th (37 AA) Mets were substituted by alanine residues in the construct, GFP was targeted only to the plastids (Fig. 2B). In contrast, deleting the first 29 AA (HisRS77124-2ndMet+GFP) led to GFP localizing only to the small blobs, and not co-localizing with chlorophyll-autofluorescence (Fig. 2C). We hypothesized that these small blobs correspond to the mitochondria. Unfortunately, MitoTracker failed to stain in transformants, so we used immuno-electron microscopy with an anti-GFP antibody. The conjugated gold particles were observed to accumulate in mitochondria (40 ± 14.2 gold particles/μm^2) but not in other compartments (plastids, 2.4 ± 1.8 gold particles/μm^2; cytoplasm, 2.4 ± 0.9 gold particles/μm^2) (Fig. 2C), confirming this identification. When the mature protein alone was fused with GFP (HisRS77124-7thMet+GFP), GFP fluorescence was observed in the cytoplasm (Fig. 2D). Altogether, these data suggest that HisRS No. 77124 uses two alternative translation starts (the first or second Met codon) in the N-terminal extension, and that the two products are targeted to plastids or mitochondria, respectively.

Next, we examined the localization of HisRS No. 84033 in cells transformed with a fusion protein, and GFP fluorescence was observed in small spots, each of which was adjacent to chlorophyll autofluorescence but clearly distinguished from pyrenoids (Fig. 2E). This localization pattern is identical to the localization observed previously for PPC and nucleomorph proteins (Hirakawa et al., 2009; Hirakawa et al., 2010).

The last HisRS, HisRS No. 85238 had two variant cDNAs (Fig. 1A). Both were fused to GFP (HisRS No. 85238-1stMet+GFP and HisRS No. 85238-5thMet), and both were localized in the cytoplasm of transformed cells (Fig. 2F,G), confirming that HisRS No. 85238 is a cytoplasmic protein.

Because only two GlyRSs were identified, one might expect either that they are both dually targeted or one triple targeted. To test this, we first expressed the N-terminal extension of GlyRS No. 87924 fused to GFP (GlyRS No. 87924-1stMet+GFP), which led to GFP in the plastids and mitochondria-like blobs (see above, Fig. 3A). In construct where the second and third Met codon (42 AA and 68 AA, respectively) were substituted by alanine, GFP was targeted only to the plastids (Fig. 3B). When amino acids 1 to 41 AA of the N-terminal extension were deleted, GFP was targeted exclusively to mitochondrion-like blobs (Fig. 3C), and when the N-terminal extension was removed entirely, GFP localized to the cytoplasm (Fig. 3D). Overall, these data suggest that GlyRS No. 87924 (like HisRS No. 77124) has two translation variants that are imported into plastids and mitochondria, respectively.

When cells expressing GFP fused with the N-terminus of the second GlyRS (GlyRS No. 87243) were observed, GFP florescence was restricted in the cytoplasm (Fig. 3E). Thus, no GlyRS appears to be targeted to the PPC.

Phylogenetic analysis of *B. natans* HisRSs and GlyRSs

To investigate the evolutionary origins of *B. natnas* HisRSs and GlyRSs, we reconstructed phylogenetic trees with diverse organisms including eukaryotes, archaea, and bacteria. In the case of HisRSs, there were two clearly distinct eukaryotic clades. One included a wide range of eukaryotes (clade-E1), while the second clade (clade-E2) mostly consisted of photosynthetic eukaryotes and was further divided into two paralogous groups (Fig. 4A). The predicted subcellular localization of all eukaryotic HisRSs (Supplementary material Tables S1, S2) suggests that clade-E1 consists of cytoplasmic and mitochondrion-targeted sequences, whereas one subgroup of clade-E2 are cytoplasmic sequences and the other are potentially dually targeted to mitochondria and plastids (Fig. 4A). The dual-targeted sequences were closely related to neither proteobacteria nor cyanobacteria. Interestingly, all eukaryotes possessed only one type of cytoplasmic HisRS; either clade-E1 or E2, never both. The three *B. natans* HisRS sequences were found to fall into the three distinct groups; the PPC-targeted HisRS (No. 84033) was sister to clade-E1, whereas the dual-targeted (No. 77124) and cytoplasmic (No. 85238) sequence fell in either of subgroups of clade-E2 (Fig. 4A).

GlyRSs are divided into two distinct and evolutionary unrelated types; a tetramer of $\alpha2\beta2$ or a $\alpha2$ homodimer. Most bacteria and organelle-targeted sequences of land plants are the tetramer type, whereas the homodimer type is found in the remainder of eukaryotes, archaea, and some bacteria (Woese

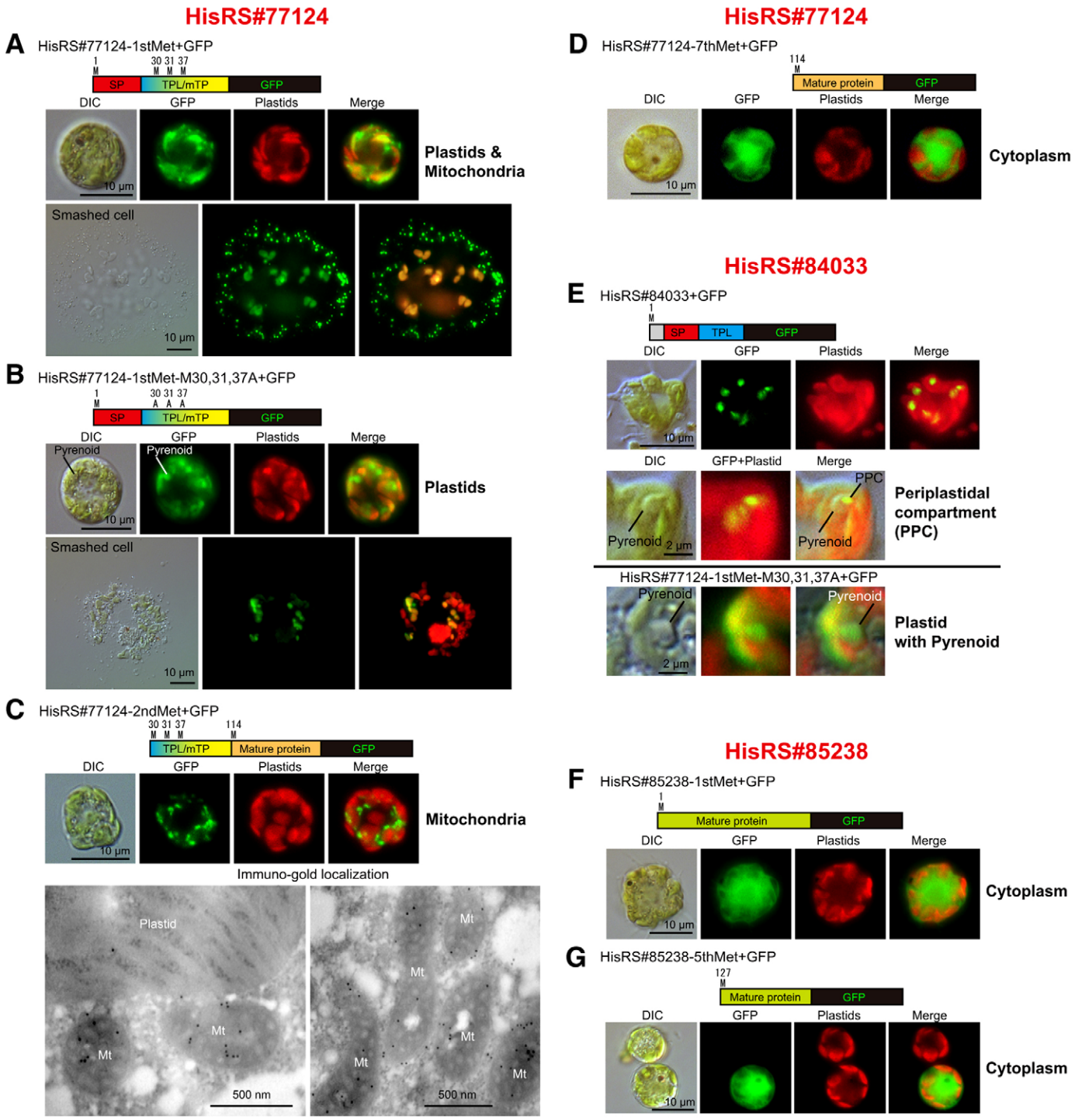

Fig. 2. Subcellular localizations of GFP-tagged HisRS No. 77124, HisRS No. 84033 and HisRS No. 85238 in transformed *A. amoebiformis* cells. Each scheme above the images indicates the expressed GFP fusion protein; M and A shows the position of methionine and substituted alanine residues, respectively. The images labeled GFP and Plastids show GFP fluorescence (green) and chlorophyll autofluorescence (red), respectively. (**A**,**B**) Lower panels show GFP localization in slightly smashed cells. The pyrenoid is a part of plastid stroma that has no chlorophyll autofluorescence (Hirakawa and Ishida, 2010). (**C**) Electron micrographs show the localization of HisRS No. 77124-2ndMet+GFP labeled with 10 nm gold particles. (**D**) GFP localization of HisRS No. 77124-7thMet+GFP in the cytoplasm. (**E**) GFP localization in the periplastidal compartment (PPC) is distinguished from the GFP localization in the plastid with a pyrenoid of a transformant expressing HisRS No. 77124-1stMet-M30,31,37A+GFP (lower panels). (**F**,**G**) GFP localization of both transcriptional variants of HisRS No. 85238 correspond to host cytoplasm, showing that alternative transcription does not play a role in dual targeting of this protein. DIC, differential interference contrast images; SP, signal peptide; TPL, transit peptide-like; mTP, mitochondrial targeting peptide; Mt, mitochondrion.

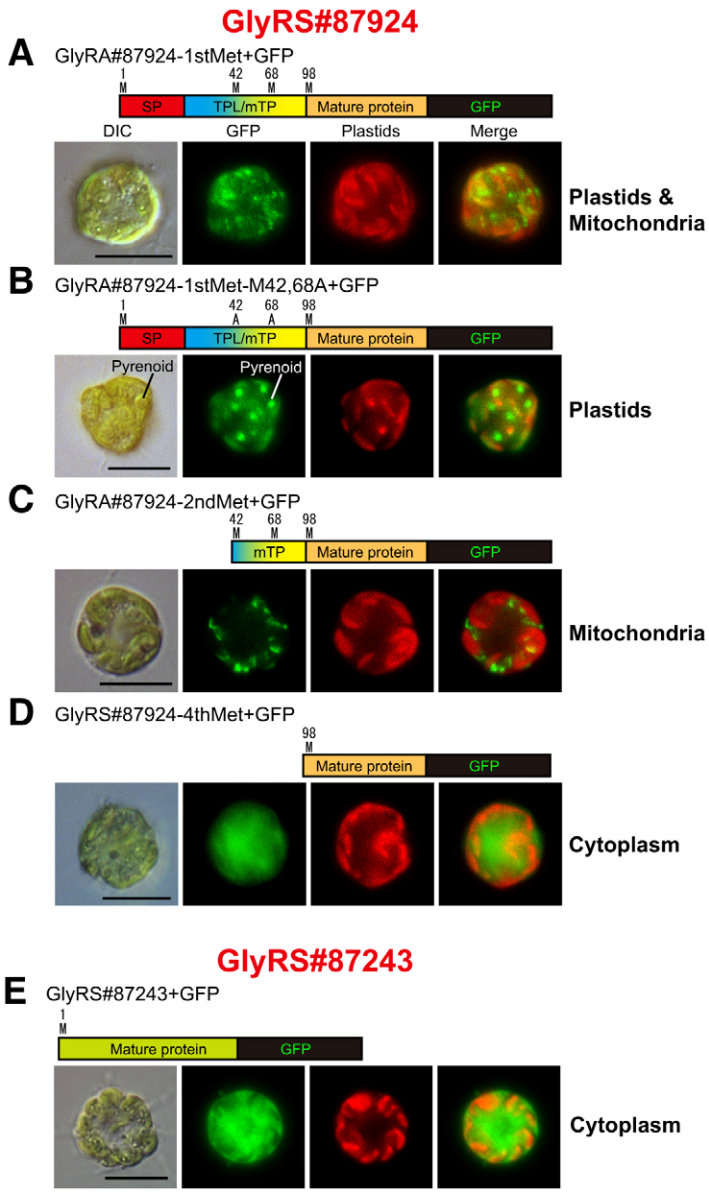

Fig. 3. Subcellular localizations of GFP-tagged GlyRS No. 87924 and GlyRS No. 87243 in transformed *A. amoebiformis* cells. (**A–E**) Each scheme above the images indicates the expressed GFP fusion protein; M and A shows the position of methionine and substituted alanine residues, respectively. The images labeled GFP and Plastids show GFP fluorescence (green) and chlorophyll autofluorescence (red), respectively. DIC, differential interference contrast images; SP, signal peptide; TPL, transit peptide-like; mTP, mitochondrial targeting peptide; Mt, mitochondrion. Scale bars: 10 μm (DIC images).

et al., 2000; Duchêne et al., 2001). All sequences identified here were the homodimer type (Fig. 4B). Reconstructing their phylogeny showed most eukaryotic cytoplasmic and mitochondrial proteins belonged to a single large clade including the cytoplasmic GlyRS (No. 87243) of *B. natans*. In contrast, the dual-targeted GlyRS (No. 87924) branched within bacterial sequences, closely related to four sequences of heterokont algae, two of which were also predicted to be dually targeted into plastids and mitochondria (Fig. 4B; supplementary material Table S3). Three putative mitochondrial GlyRSs were also found in the bacterial clade, but they were not closely related to the dual-targeted sequences (Fig. 4B).

Discussion

Dual targeting to plastids and mitochondria in complex algae

Dual targeting of aaRS proteins to mitochondria and plastids is not uncommon in plants (Peeters and Small, 2001; Duchêne et al., 2005; Berglund et al., 2009), in part because plastid and mitochondrial protein import systems and targeting sequences share similar characteristics (Pujol et al., 2007; Carrie et al., 2009). However, in complex algae derived from secondary endosymbiosis, plastid-targeting systems are more complex and completely different from those of mitochondria. The presence of an ER-targeting signal peptide at the N-terminus of plastid-targeted proteins (Patron and Waller, 2007) means that simply encoding an ambiguous targeting sequence, as in plants, is insufficient because the protein will not be exposed to the mitochondrial import apparatus once they are targeted to the ER. Here, we used a particularly complex alga that also retains a translationally active endosymbiotic cytoplasm to examine protein trafficking complexity of nucleus-encoded aaRSs targeted to the four distinct compartments and found a number of new strategies that may represent general principles for getting around this added complexity.

Sequence analyses and GFP localization experiments of HisRS (No. 77124) and GlyRS (No. 87924) proteins show that dual

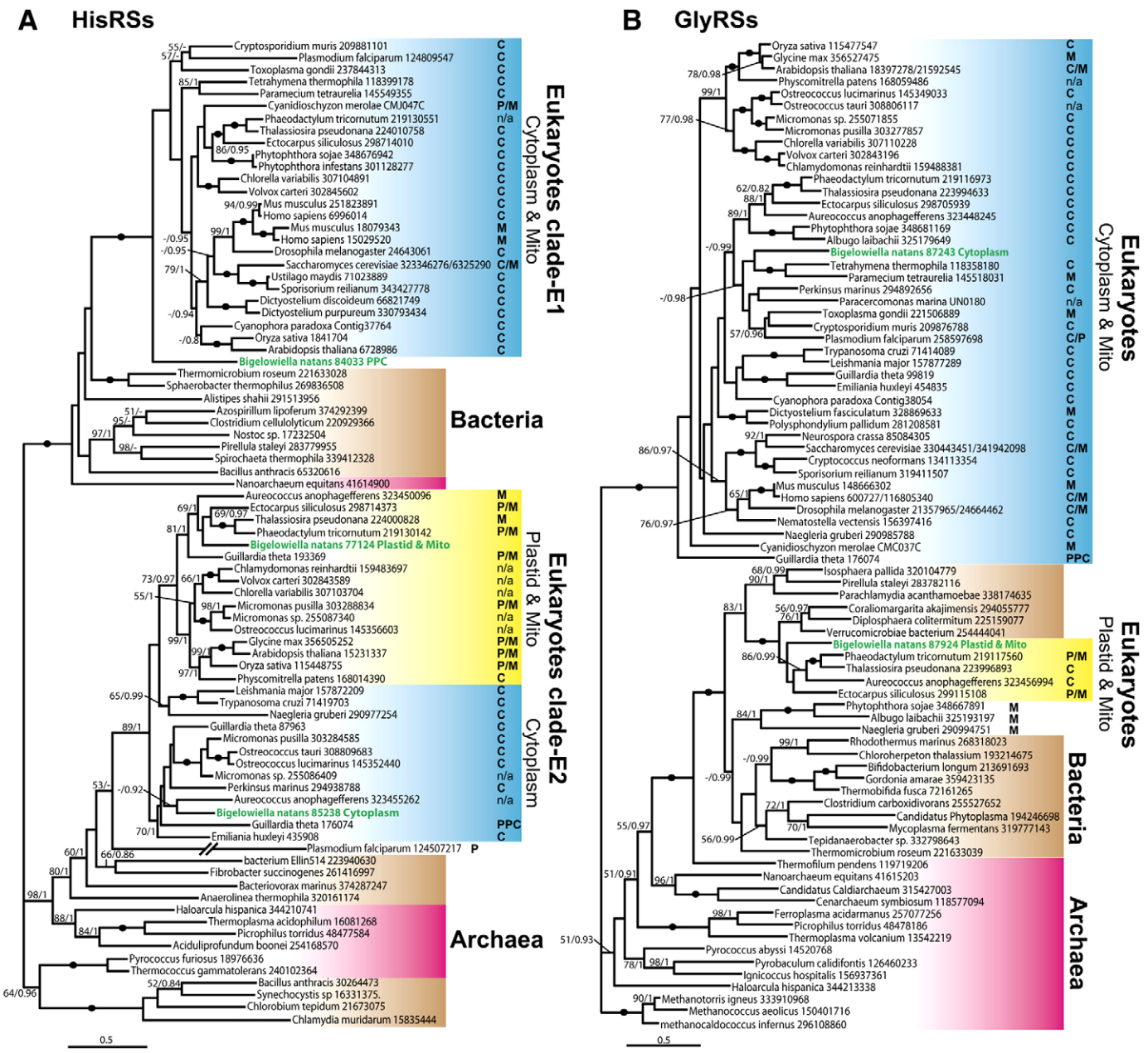

Fig. 4. Unrooted phylogenetic trees of HisRSs and GlyRSs. (A,B) The topologies correspond to the best-scoring ML tree for HisRSs (A) and GlyRSs (B), as obtained with RAxML. The values at nodes indicate the ML bootstrap supports (BS) and Bayesian posterior probabilities (PP) when they are higher than 50% and 0.8, respectively. Black dots correspond to 100% BS and 1.0 PP. Predicted subcellular localizations of eukaryotic sequences are shown on the right side of species names and accession numbers; prediction results are summarized in supplementary material Tables S1–S3. C, cytoplasm; M, mitochondria; P, plastids; PPC, periplastidal compartment; n/a, the sequence was not available for the prediction. The scale bar represents the estimated number of amino acid substitutions per site.

targeting to mitochondria and plastids is also achieved in complex chlorarachniophyte algae using an N-terminal bipartite targeting sequence consisting of a SP followed by an ambiguous TPL. Unlike plants, however, the N-terminal sequence is capable of mitochondrial targeting only if translated from the second Met codon (Figs 2, 3). Since both genes are expressed from a single transcription start (Fig. 1A), we propose that alternatively targeted proteins must be generated by translation initiation variants: one product translated from the first Met codon is co-translationally transported to the plastid via the ER, and another product translated from the second Met codon is post-translationally imported into the mitochondrion. A similar situation is found in pyruvate kinase from the apicomplexan parasite *Toxoplasma gondii* (Saito et al., 2008), which has a non-photosynthetic secondary plastid of red algal origin, suggesting this may be a common solution to dual targeting in complex plastid-bearing organisms. However, ambiguous targeting sequence for mitochondria and secondary plastids have also been identified for *T. gondii* superoxide

dismutase and aconitase (Pino et al., 2007), suggesting a deep pool of possible solutions for dual targeting of organelle proteins, even in complex systems.

Cytoplasmic and PPC-targeted aaRSs

The *B. natans* nuclear genome encodes single cytoplasmic HisRS (No. 85238) and GlyRS (No. 87243) genes. We found no typical targeting sequence at the N-termini of these aaRSs, and GFP localizations confirmed that they accumulated in the cytoplasm. We also identified a periplastidal compartment (PPC)-targeted HisRS (No. 84033) in *B. natans*: the protein encodes an N-terminal extension like other PPC-targeted proteins (Hirakawa et al., 2010; Hirakawa et al., 2011), and GFP localization confirmed that it accumulated in the PPC (Fig. 2E). Half of the aaRS classes were found in three copies, suggesting the situation seen for HisRS, with cytoplasmic, PPC, and dual plastid/mitochondrial copies, may represent the most common outcome to reducing the redundancy of aaRS genes. Indeed, analyzing three other aaRS classes with three completely sequenced copies, LeuRS, TryRS, and CysRS, supports this suggestion as in all three cases leader analysis is consistent with the situation seen for HisRS.

In contrast to HisRS, no PPC-targeted GlyRS was identified in the genome, and neither of the two proteins we did find appear to be PPC-targeted. Although this appears to leave a critical gap in the translational competence of the PPC, an alternative solution exists because the nucleomorph genome also lacks the gene for $tRNA^{Gly}$ (Gilson et al., 2006). Since it is highly unlikely that the amino acid glycine has been lost from the PPC proteome (e.g. replaced by superwobble of another cognate tRNA), and since the PPC does not import GlyRS protein, we propose that aminoacylated $tRNA^{Gly}$ is specifically imported from the cytoplasm, obviating the need for either an aaRS or the cognate tRNA gene. The nucleomorph genome also lacks $tRNA^{Pro}$, $tRNA^{Ala}$ and $tRNA^{Arg}$ genes, and the nuclear genome encodes four, three and two copies of aaRSs corresponding to these tRNAs, respectively. The situation for ArgRS seems to be the same as GlyRS, whereas no conclusion can be made for ProRS and AlaRS because of their incomplete sequences.

Evolutionary history of HisRSs and GlyRSs in eukaryotes

The evolutionary histories of aaRSs have long been known to be complex, being especially prone to horizontal gene transfer and replacement (Brindefalk et al., 2007; Brandão and Silva-Filho, 2011). Eukaryotic HisRS genes form two distinct lineages (clade-E1 and E2), one of which is further subdivided (Fig. 4A). Diverse eukaryotes possess clade-E1 genes for cytoplasmic and/or mitochondrial HisRSs, whereas clade-E2 consists of plastid bearing eukaryotes and a few excavate protists. Probably the clade-E2 organisms acquired an extra HisRS gene from some bacterium (neither proteobacteria nor cyanobacteria) by horizontal gene transfer (HGT) followed by duplication; one copy remained functional in the cytoplasm and the other evolved into dual-targeted proteins for plastids and mitochondria (Fig. 4A). The cytoplasmic genes of clade-E2 were subsequently replaced by clade-E1 genes in several organisms. The cytoplasmic and dual-targeted organelle HisRS genes of *B. natans* appear to share a similar evolutionary history as those of other plastid-bearing eukaryotes, despite their distant phylogenetic relationship (summarized in supplementary material Fig. S1). In contrast, the PPC-targeted HisRS is closely related to clade-E1 (Fig. 4A), so it is unclear whether the gene is derived from the host or endosymbiont. In the case of GlyRSs, the *B. natans* dual-targeted gene was derived from some bacterium by HGT like the dual-targeted HisRS, and dual-targeted GlyRSs from several heterokont algae also likely trace back to this event (Fig. 4B). Therefore, both dual-targeted aaRS sequences of *B. natnas* appear to be derived from neither the plastid nor mitochondrial lineage. These dual-targeted aaRSs have a complex targeting sequence at N-termini, which might have been acquired via exon-shuffing like several plastid targeted proteins (Kilian and Kroth, 2004) since these genes possess an intron in the presequence (Fig. 1B,C).

Overall, our findings shed light on the evolution of dual targeting of organelle proteins in complex cells, and suggest that dual targeting to mitochondrion and plastid is a common feature in algae with secondary plastids as well as land plants, despite their evolutionary divergence. However, how this is accomplished appears to be different due to differences in targeting system to complex plastids: rather than simple ambiguous leaders, our work suggests that multiple endosymbiotic events have increased the complexity of organelle protein trafficking by use of alternative translation.

Materials and Methods

RLM-RACE for HisRS and GlyRS genes

Bigelowiella natans (CCMP621) that is the strain synonym of CCMP2755 were grown at 20℃ under white illumination on a 12:12 hour light:dark cycle in ESM medium (Kasai et al., 2009). Total RNA was extracted from the cells using a TRIzol Reagent (Life Technologies), and full-length cDNA was synthesized by a FirstChoice RLM-RACE Kit (Ambion) according to the manufacturer's protocols. The 5′ end(s) of each target gene was amplified by PCR using Econo Taq DNA polymerase (Lucigen) with specific primer sets (shown in supplementary material Table S4). PCR products were detected by 1% agarose gel electrophoreses with SYBR Safe (Life Technologies), and subsequently cloned and sequenced.

In vivo GFP targeting analysis

To construct plasmids encoding GFP-tagged aaRSs, target cDNA fragments were amplified by PCR with specific primer sets (shown in supplementary material Table S5), and each fragment was inserted between the *Hin*dIII and *Nco*I site of the pLaRGfp+mc vector (Hirakawa et al., 2009). To introduce substitutions into some target fragments, we used PCR-based site-directed mutagenesis technique (Higuchi et al., 1988). To analyze the subcellular localization of GFP fusion proteins, we used the transient transformation system of the chlorarachniophyte, *Amorphochlora amoebiformis* (CCMP2058). The cells were transformed with each plasmid using a Biolistic PDS-1000/He particle delivery system (Bio-Rad), as described previously (Hirakawa et al., 2008). GFP-expressing cells were observed under an Axioplan 2 fluorescent microscope as described previously (Hirakawa et al., 2011).

Immunogold localization

Transient transformed cells with the HisRS77124-2ndMet+GFP construct were fixed for 1 hour at 4℃ in 3% paraformaldehyde, 0.5% glutaraldehyde and 0.25 M sucrose in PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM $MgCl_2$; pH 7.4). A transformed cell identified to have GFP fluorescence was micropipetted onto a poly-L-lysine coated coverslip. The cell was dehydrated by an ethanol series, and embedded with LR White as described previously (Hirakawa et al., 2010). Ultrathin sections of polymerised block were created by a Leica ultramicrotome EM UC6, and collected onto Formvar coated copper grids. Immunogold labeling was performed with anti-GFP primary antibody (JL-8; Clontech), and anti-mouse immunoglobulin G secondary antibody conjugated with 10 nm gold particles (G7777; Sigma) in the condition as described previously (Hirakawa et al., 2010). The ultrathin sections were observed under a Hitachi H7600 transmission electron microscope.

Phylogenetic analysis

Homologs of HisRS and GlyRS sequences were identified by BlastP against GenBank, retrieved and automatically aligned with the L-INS-I method of the MAFFT package (Katoh et al., 2005). Trimal v1.2 (Capella-Gutiérrez et al., 2009) with the –gt 0.8 option was used to automatically and reproducibly eliminate poorly aligned positions; the resulting unambiguously aligned positions were manually inspected with Seaview (Gouy et al., 2010). Maximum Likelihood (ML) analyses were performed using RAxML 7.2.8 (Stamatakis, 2006), with the rapid

hill-climbing algorithm and the LG+γ+F model of evolution (-m PROTGAMMALGF, four discrete rate categories). The best-scoring ML trees were determined in multiple searches using 20 randomized stepwise addition parsimony starting trees. Statistical support was evaluated with non-parametric bootstrapping using 100 replicates. Bayesian analyses using the WAG+γ+F model (four gamma categories) were performed with MrBayes 3.2 (Ronquist et al., 2012). Each inference consisted of two independent runs starting from a random tree and four Metropolis-coupled Markov Chain Monte Carlo (MCMCMC), for 2,000,000 generations with sampling every 100 generations. The average standard deviation of split frequencies was used to assess the convergence of the two runs. Bayesian posterior probabilities were calculated from the majority rule consensus of the tree sampled after the initial burnin period, which corresponded to 25% of the total generations.

Acknowledgements

We thank the US Department of Energy Joint Genome Institute (JGI) for sequencing the genomes of *B. natans* and *G. theta*, and J. M. Archibald, M. W. Gray, G. I. McFadden, C. E. Lane and B. Curtis for project contributions. We thank the Kronstad lab for granting access to the particle bombardment system, and Noriko Okamoto for helping in the electron microscopy. Y.H. is a JSPS Postdoctoral Fellow for Research Abroad, F.B. is a Tula Postdoctoral Fellow, and P.J.K. is a Fellow of the Canada Institute for Advanced Research and a Senior Scholar of the Michael Smith Foundation for Health Research.

Funding

This work was funded by a grant from the Natural Sciences and Engineering Research Council of Canada [grant number 227301]; and by a grant to the Centre for Microbial Diversity and Evolution from the Tula Foundation. The work conducted by the Joint Genome Institute (JGI) is supported by the Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231. Deposited in PMC for release after 12 months.

Supplementary material available online at
http://jcs.biologists.org/lookup/suppl/doi:10.1242/jcs.116533/-/DC1

References

Akashi, K., Grandjean, O. and Small, I. (1998). Potential dual targeting of an Arabidopsis archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS let.* **431**, 39-44.

Archibald, J. M. (2009). The puzzle of plastid evolution. *Curr. Biol.* **19**, R81-R88.

Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**, 783-795.

Berglund, A. K., Pujol, C., Duchêne, A. M. and Glaser, E. (2009). Defining the determinants for dual targeting of amino acyl-tRNA synthetases to mitochondria and chloroplasts. *J. Mol. Biol.* **393**, 803-814.

Bolte, K., Bullmann, L., Hempel, F., Bozarth, A., Zauner, S. and Maier, U. G. (2009). Protein targeting into secondary plastids. *J. Eukaryot. Microbiol.* **56**, 9-15.

Bonnefond, L., Fender, A., Rudinger-Thirion, J., Giegé, R., Florentz, C. and Sissler, M. (2005). Toward the full set of human mitochondrial aminoacyl-tRNA synthetases: characterization of AspRS and TyrRS. *Biochemistry* **44**, 4805-4816.

Brandão, M. M. and Silva-Filho, M. C. (2011). Evolutionary history of Arabidopsis thaliana aminoacyl-tRNA synthetase dual-targeted proteins. *Mol. Biol. Evol.* **28**, 79-85.

Brindefalk, B., Viklund, J., Larsson, D., Thollesson, M. and Andersson, S. G. E. (2007). Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol. Biol. Evol.* **24**, 743-756.

Burki, F., Okamoto, N., Pombert, J. F. and Keeling, P. J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. *Proc. Biol. Sci.* **279**, 2246-2254.

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

Carrie, C., Giraud, E. and Whelan, J. (2009). Protein transport in organelles: Dual targeting of proteins to mitochondria and chloroplasts. *FEBS J.* **276**, 1187-1195.

Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M. C., Ball, S. G., Gile, G. H., Hirakawa, Y. et al**.,** (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59-65.

Duchêne, A. M., Peeters, N., Dietrich, A., Cosset, A., Small, I. D. and Wintz, H. (2001). Overlapping destinations for two dual targeted glycyl-tRNA synthetases in Arabidopsis thaliana and Phaseolus vulgaris. *J. Biol. Chem.* **276**, 15275-15283.

Duchêne, A. M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N. M., Zaepfel, M., Maréchal-Drouard, L. and Small, I. D. (2005). Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* **102**, 16484-16489.

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

Emanuelsson, O., Nielsen, H. and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* **8**, 978-984.

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

Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J. and McFadden, G. I. (2006). Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. *Proc. Natl. Acad. Sci. USA* **103**, 9566-9571.

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221-224.

Higuchi, R., Krummel, B. and Saiki, R. K. (1988). A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351-7367.

Hirakawa, Y. and Ishida, K. (2010). Internal plastid-targeting signal found in a RubisCO small subunit protein of a chlorarachniophyte alga. *Plant J.* **64**, 402-410.

Hirakawa, Y., Kofuji, R. and Ishida, K. (2008). Transient transformation of a chlorarachniophyte alga, Lotharella amoebiformis (chlorarachniophyceae), with uidA and egfp reporter genes. *J. Phycol.* **44**, 814-820.

Hirakawa, Y., Nagamune, K. and Ishida, K. (2009). Protein targeting into secondary plastids of chlorarachniophytes. *Proc. Natl. Acad. Sci. USA* **106**, 12820-12825.

Hirakawa, Y., Gile, G. H., Ota, S., Keeling, P. J. and Ishida, K. (2010). Characterization of periplastidal compartment-targeting signals in chlorarachniophytes. *Mol. Biol. Evol.* **27**, 1538-1545.

Hirakawa, Y., Burki, F. and Keeling, P. J. (2011). Nucleus- and nucleomorph-targeted histone proteins in a chlorarachniophyte alga. *Mol. Microbiol.* **80**, 1439-1449.

Jackson, K. E., Pham, J. S., Kwek, M., De Silva, N. S., Allen, S. M., Goodman, C. D., McFadden, G. I., de Pouplana, L. R. and Ralph, S. A. (2012). Dual targeting of aminoacyl-tRNA synthetases to the apicoplast and cytosol in Plasmodium falciparum. *Int. J. Parasitol.* **42**, 177-186.

Kasai, F., Kawachi, M., Erata, M., Yumoto, K. and Sato, M. (2009). NIES-collection list of strains, 8th edition. *Jpn. J. Phycol. (Sosui)* **57**, 220.

Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511-518.

Keeling, P. J. (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 729-748.

Kilian, O. and Kroth, P. G. (2004). Presequence acquisition during secondary endocytobiosis and the possible role of introns. *J. Mol. Evol.* **58**, 712-721.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580.

Maréchal-Drouard, L., Weil, J. H. and Dietrich, A. (1993). Transfer RNAs and transfer RNA genes in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**, 13-32.

McFadden, G. I., Gilson, P. R., Hofmann, C. J., Adcock, G. J. and Maier, U. G. (1994). Evidence that an amoeba acquired a chloroplast by retaining part of an engulfed eukaryotic alga. *Proc. Natl. Acad. Sci. USA* **91**, 3690-3694.

Morgante, C. V., Rodrigues, R. A., Marbach, P. A., Borgonovi, C. M., Moura, D. S. and Silva-Filho, M. C. (2009). Conservation of dual-targeted proteins in Arabidopsis and rice points to a similar pattern of gene-family evolution. *Mol. Genet. Genomics* **281**, 525-538.

Natsoulis, G., Hilger, F. and Fink, G. (1986). The HTS1 gene encodes both the cytoplasmic and mitochondrial histidine tRNA synthetases of *S. cerevisiae*. *Cell* **46**, 235-243.

Patron, N. J. and Waller, R. F. (2007). Transit peptide diversity and divergence: A global analysis of plastid targeting signals. *Bioessays* **29**, 1048-1058.

Peeters, N. and Small, I. (2001). Dual targeting to mitochondria and chloroplasts. *Biochim. Biophys. Acta* **1541**, 54-63.

Pino, P., Foth, B. J., Kwok, L.-Y., Sheiner, L., Schepers, R., Soldati, T. and Soldati-Favre, D. (2007). Dual targeting of antioxidant and metabolic enzymes to the mitochondrion and the apicoplast of Toxoplasma gondii. *PLoS Pathog.* **3**, e115.

Pujol, C., Maréchal-Drouard, L. and Duchêne, A. M. (2007). How can organellar protein N-terminal sequences be dual targeting signals? In silico analysis and mutagenesis approach. *J. Mol. Biol.* **369**, 356-367.

Pujol, C., Bailly, M., Kern, D., Maréchal-Drouard, L., Becker, H. and Duchêne, A. M. (2008). Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNA^{Gln} synthesis in plants. *Proc. Natl. Acad. Sci. USA* **105**, 6481-6485.

Rogers, M. B., Archibald, J. M., Field, M. A., Li, C., Striepen, B. and Keeling, P. J. (2004). Plastid-targeting peptides from the chlorarachniophyte Bigelowiella natans. *J. Eukaryot. Microbiol.* **51**, 529-535.

Rogers, M. B., Gilson, P. R., Su, V., McFadden, G. I. and Keeling, P. J. (2007). The complete chloroplast genome of the chlorarachniophyte Bigelowiella natans: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. *Mol. Biol. Evol.* **24**, 54-62.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. and Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539-542.

Rubio, M. A. and Hopper, A. K. (2011). Transfer RNA travels from the cytoplasm to organelles. *Wiley Interdiscip. Rev. RNA* **2**, 802-817.

Saito, T., Nishi, M., Lim, M. I., Wu, B., Maeda, T., Hashimoto, H., Takeuchi, T., Roos, D. S. and Asai, T. (2008). A novel GDP-dependent pyruvate kinase isozyme from Toxoplasma gondii localizes to both the apicoplast and the mitochondrion. *J. Biol. Chem.* **283**, 14041-14052.

Schleiff, E. and Becker, T. (2011). Common ground for protein translocation: access control for mitochondria and chloroplasts. *Nat. Rev. Mol. Cell Biol.* **12**, 48-59.

Small, I., Peeters, N., Legeai, F. and Lurin, C. (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* **4**, 1581-1590.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Turner, R. J., Lovato, M. and Schimmel, P. (2000). One of two genes encoding glycyl-tRNA synthetase in *Saccharomyces cerevisiae* provides mitochondrial and cytoplasmic functions. *J. Biol. Chem.* **275**, 27681-27688.

Woese, C. R., Olsen, G. J., Ibba, M. and Söll, D. (2000). Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202-236.