

Aminoacyl-tRNA Synthetase Evolution within the Dynamic Tripartite Translation System of Plant Cells

Daniel B. Sloan ^{1,*}, Rachael A. DeTar¹, and Jessica M. Warren²

¹Department of Biology, Colorado State University, Fort Collins

²Center for Mechanisms of Evolution, Biodesign Institute and School of Life Sciences, Arizona State University, Tempe

*Corresponding author: E-mail: dan.sloan@colostate.edu.

Accepted: 09 March 2023

Abstract

Eukaryotes maintain separate protein translation systems for nuclear and organellar genes, including distinct sets of tRNAs and aminoacyl-tRNA synthetases (aaRSs). In animals, mitochondrial-targeted aaRSs are expressed at lower levels and are less conserved in sequence than cytosolic aaRSs involved in translation of nuclear mRNAs, likely reflecting lower translational demands in mitochondria. In plants, translation is further complicated by the presence of plastids, which share most aaRSs with mitochondria. In addition, plant mitochondrial tRNA pools have a dynamic history of gene loss and functional replacement by tRNAs from other compartments. To investigate the consequences of these distinctive features of translation in plants, we analyzed sequence evolution in angiosperm aaRSs. In contrast to previously studied eukaryotic systems, we found that plant organellar and cytosolic aaRSs exhibit only a small difference in expression levels, and organellar aaRSs are slightly *more* conserved than cytosolic aaRSs. We hypothesize that these patterns result from high translational demands associated with photosynthesis in mature chloroplasts. We also investigated aaRS evolution in *Sileneae*, an angiosperm lineage with extensive mitochondrial tRNA replacement and aaRS retargeting. We predicted positive selection for changes in aaRS sequence resulting from these recent changes in subcellular localization and tRNA substrates but found little evidence for accelerated sequence divergence. Overall, the complex tripartite translation system in plant cells appears to have imposed more constraints on the long-term evolutionary rates of organellar aaRSs compared with other eukaryotic lineages, and plant aaRS protein sequences appear largely robust to more recent perturbations in subcellular localization and tRNA interactions.

Key words: aminoacyl-tRNA synthetase, cytonuclear coevolution, flowering plant, subcellular targeting, *Sileneae*.

Significance

In eukaryotes, separate protein translation systems exist for nuclear and organellar (mitochondrial and chloroplast) genomes. The molecular machinery involved in these systems can evolve at different rates. In this study, we investigated aminoacyl-tRNA synthetases—the enzymes responsible for recognizing tRNAs and charging them with the correct amino acid. By taking advantage of the distinct features of protein translation in plant cells, we found evidence that differences in expression level and functional importance are primarily responsible for the contrasting evolutionary rates between different subcellular translation systems for these enzymes. In contrast, we found little or no evidence that coevolution between aminoacyl-tRNA synthetases and their tRNA substrates contributes to rate differences.

Introduction

Eukaryotic cells contain multiple genomic compartments, reflecting their endosymbiotic origins. Accordingly, separate translation systems exist for production of proteins encoded in nuclear, mitochondrial (mt), and plastid (pt) genomes. Organellar translation systems retain bacterial-like ribosomes (Harris et al. 1994; Smits et al. 2007) and often rely on ancestral mt-tRNA genes that still reside in the mt genome (Salinas-Giegé et al. 2015). However, much of the organellar translation machinery is now encoded in the nuclear genome (Timmis et al. 2004; Giannakis et al. 2022). In most eukaryotes, this includes all of the aminoacyl-tRNA synthetases (aaRSs), which are responsible for recognizing tRNAs and charging them with the correct amino acid. Most eukaryotes express two largely distinct sets of aaRSs—one that is responsible for translation of nuclear mRNAs in the cytosol and a second organellar set that functions in mitochondria (and plastids in lineages such as plants) (Brindefalk et al. 2007; Duchêne et al. 2009; Salinas-Giegé et al. 2015). The direct functional interactions between gene products encoded in different genomes within eukaryotic cells can lead to cytonuclear coadaptation (Rand et al. 2004; Hill 2015; Sloan et al. 2018). The rapid sequence divergence observed in many mt genomes may select for coevolutionary responses in the nucleus (Osada and Akashi 2012; Havird et al. 2015; Barreto et al. 2018), and incompatibilities between mt-tRNAs and nuclear-encoded aaRSs can have severe fitness consequences (Meiklejohn et al. 2013).

In animals, mt aaRSs exhibit faster sequence divergence than their cytosolic counterparts (Pett and Lavrov 2015; Adrion et al. 2016; Barreto et al. 2018). One possible explanation for these higher evolutionary rates is selection on nuclear-encoded aaRSs to respond to changes in mt-tRNA genes. However, mt aaRSs also have lower expression levels than cytosolic aaRSs, which may indicate that they are simply under weaker purifying selection because translational demands are less intense in mitochondria than in the cytosol (Sloan et al. 2014; Pett and Lavrov 2015). Indeed, when accounting for gene expression levels, there is often no detectable difference in evolutionary rates between mt and cytosolic aaRSs in animals (Adrion et al. 2016; Barreto et al. 2018). Therefore, the evidence that mitonuclear coevolution is a substantial contributor to rates of mt aaRS sequence evolution remains very limited.

In contrast to the detailed work in animal systems, evolutionary rates for organellar and cytosolic aaRSs have not been studied extensively in plants. Yet, there are reasons to expect that the presence of plastids may place distinct constraints on plant organellar aaRS evolution. Translation in plastids is responsible for massive levels of protein production. The chloroplasts themselves house ~80% of the total protein content in leaf mesophyll cells (Heinemann et al. 2021). Although this

total also includes the large number of nuclear-encoded proteins that are imported into chloroplasts from the cytosol, pt genes produce the majority of mRNA transcripts in the cell (Forsythe et al. 2022), and their protein products such as PsbA and Rbcl have exceptionally high rates of translation (Chotewutmontri and Barkan 2018). There are two largely separate classes of aaRSs in plants: cytosolic and organellar, with the latter being dual targeted to mitochondria and plastids (Duchêne et al. 2005; Duchêne et al. 2009). Therefore, translational demands in the plastid are expected to affect conservation and rates of evolution of aaRSs with shared function in the mitochondria.

The dynamic mixture of tRNAs that function in plant mitochondria adds another potential complexity to the evolution of plant aaRSs. Plant mt genomes contain tRNA genes from multiple origins, including a portion of the ancestral mt gene set as well as horizontal transfers from the pt genome, other bacterial genomes, and the mt genomes of other plants (Small et al. 1999; Warren and Sloan 2020). However, even this heterogeneous set of tRNA genes is insufficient for translation of all codons, and plant mitochondria import additional nuclear-encoded tRNAs from the cytosol (Michaud et al. 2011). The extent of cytosolic import varies among plant species as there have been varying degrees of mt-tRNA gene loss among lineages (Warren and Sloan 2020). Most angiosperm mt genomes encode 16–20 different types (i.e., anticodons) of tRNAs (Richardson et al. 2013). However, some lineages have experienced much more extensive mt-tRNA gene loss (Guo 2014; Petersen et al. 2015; Skippington et al. 2015; Sanchez-Puerta et al. 2017). The angiosperm tribe *Sileneae* is a striking example of recent and ongoing loss, with some species retaining as many as 14 mt-tRNA genes while other close relatives have only two or three (Warren et al. 2021).

The history of plant mt-tRNA gene loss raises questions about the effects on aaRS function. Eukaryotic/nuclear tRNAs are essentially unrecognizable in primary sequence when compared with bacterial-like mt-tRNAs. Therefore, it is not clear if and how organellar aaRSs are able to charge these newly imported tRNAs in the mitochondria. Results from *in silico* analysis of putative targeting peptides and fluorescence microscopy assays of subcellular localization in *Sileneae* have identified two alternative pathways (Warren et al. 2022). In some cases, changes in tRNA import have been accompanied by retargeting of the corresponding cytosolic aaRS such that the ancestral pairing between tRNA and aaRS is maintained and simply relocated to a second compartment. In these cases, there is also the potential for concomitant loss of mt targeting in the organellar aaRSs such that ancestral dual-targeted enzymes now function exclusively in plastids. In other cases of mt-tRNA gene loss and functional replacement, there is no evidence of aaRS retargeting, implying that the ancestral organellar

aaRS has evolved to charge a newly imported cytosolic tRNA. These two alternative responses to mt-tRNA gene loss and functional replacement are also observed in more ancient examples in plant evolution (Duchêne et al. 2009).

Such changes in targeting or tRNA substrates represent potentially radical perturbations in aaRS function, and we hypothesize that they could result in positive selection and accelerated evolution in aaRS sequence for at least three different reasons. First, aaRSs evolving to charge new tRNA substrates may require changes to key regions involved in recognition of tRNA “identity elements” (Giegé et al. 1998; Igloi 2021, 2022). Second, mitochondria differ from the cytosol in numerous respects (e.g., osmotic conditions) that could potentially affect protein folding and tRNA interactions and, thus, create selection for changes in protein sequence. Third, organellar aaRSs that ancestrally functioned in both the mitochondria and plastids but lose mt targeting may experience altered selection pressures in specializing exclusively on pt function. All these mechanisms would be predicted to accelerate evolution and produce signatures of positive selection for amino acid substitutions in lineages that have experienced recent changes in mt-tRNA gene content.

In this study, we test predictions about how the distinctive features of plant organellar translation have shaped rates of aaRS evolution. We take a phylogenetic approach to analyze sequence evolution in plant aaRS families, including long-term patterns across divergent angiosperms and the more recent history of divergence among *Sileneae* species that differ greatly in mt-tRNA gene content and cytosolic import.

Results

Angiosperm Organellar aaRSs Exhibit a High Degree of Sequence Conservation

To characterize long-term rates of aaRS amino acid sequence divergence in angiosperms, we generated trees for each aaRS with representatives of eudicots (*Arabidopsis thaliana* and *Vitis vinifera*) and monocots (*Oryza sativa* and *Spirodela polyrhiza*). Total sequence divergence (substitutions per site) was determined based on summed branch lengths within the tree (using an averaging approach in case of duplicated gene copies). We found that organellar aaRS sequences evolved slightly (but significantly) slower than their cytosolic counterparts across these four deeply divergent angiosperm lineages ($P = 0.0178$; paired t -test; fig. 1). Although the magnitude of the difference between these two groups was small (25% lower rates for organellar aaRSs on average), this result presents a striking contrast with previous work in animals, which has shown that mt aaRSs evolve much more rapidly than cytosolic aaRSs (Pett and Lavrov 2015; Adrion et al. 2016; Barreto et al. 2018).

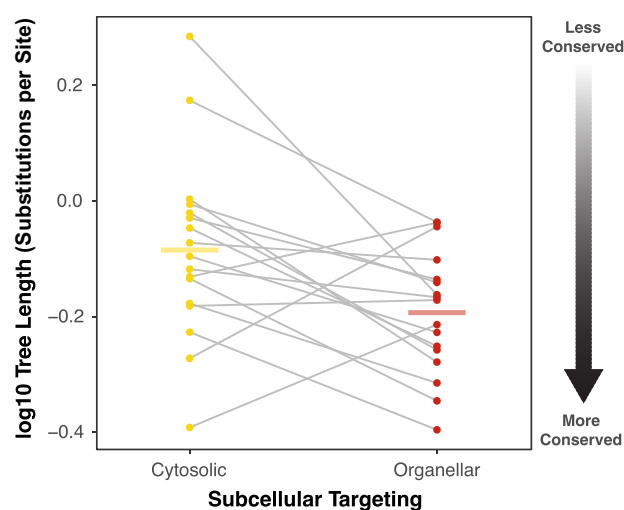


Fig. 1.—Lower rates of amino acid substitutions in angiosperm organellar aaRSs than cytosolic aaRSs. Sequence divergence was quantified by summing branch lengths (measured as amino acid substitutions per site) from phylogenetic trees for each aaRS, containing orthologs from *A. thaliana*, *V. vinifera*, *O. sativa*, and *S. polyrhiza*. Points represent log10-transformed values for the total tree length for each cytosolic (left, yellow) and organellar (right, red) aaRS, and gray lines connect pairs which connect counterparts associated with the same amino acid. Horizontal bars represent the mean for each group. Organellar rates are significantly lower than cytosolic rates based on a paired t -test on log-transformed values ($P = 0.0178$).

Angiosperm Organellar aaRSs Are Expressed at Only Slightly Lower Levels Than Their Cytosolic Counterparts

One hypothesis to explain why mt aaRS sequences evolve rapidly in animals is that they are expressed at the lower levels than cytosolic aaRSs (Adrion et al. 2016; Barreto et al. 2018), and expression tends to be negatively correlated with rates of protein evolution—the so-called E–R anticorrelation (Zhang and Yang 2015; Bédard et al. 2022). Indeed, mRNA transcript abundance for mt aaRS has been shown to be approximately 5-fold lower than for cytosolic aaRSs in multiple animal systems (Adrion et al. 2016). Thus, we reasoned that plants may not show the same imbalance in expression level that is observed in animals. Using RNA-seq data from the EMBL-EBI Expression Atlas (Moreno et al. 2022) from multiple *A. thaliana* tissue types, we found that plants exhibited lower expression levels for their organellar aaRS genes relative to their cytosolic counterparts (fig. 2). This difference is in the same direction as observed in animals. Therefore, the expression level alone cannot fully explain the inverted relationship for aaRS substitution rates in organellar and cytosolic aaRSs for plants versus animals. However, the differences in expression levels that we observed in *A. thaliana* (2.2-fold in flowers, 2.0-fold in leaves, and 1.5-fold in seedlings) are substantially smaller than those previously found in animal systems (Adrion et al. 2016). This contrast suggests that plants

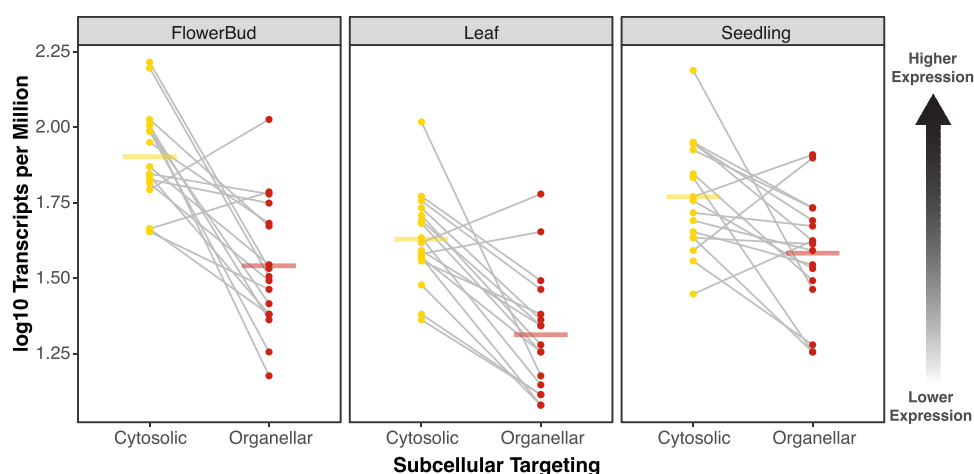


FIG. 2.—Lower levels of expression (mRNA transcript abundance) for organellar aaRS genes than cytosolic aaRS genes in three different *A. thaliana* tissue types based on data from the EMBL-EBI Expression Atlas. Points represent cytosolic (left, yellow) and organellar (right, red) aaRSs, and gray lines connect pairs that connect counterparts associated with the same amino acid. Horizontal bars represent the mean for each group. A three-way ANOVA found significant effects of aaRS targeting ($P = 1.2 \times 10^{-13}$), tissue ($P = 1.5 \times 10^{-8}$), and aaRS amino acid ($P = 4.6 \times 10^{-4}$) on the expression level as measured by log-transformed TPM.

maintain greater demands on organellar aaRS function than animals, which may contribute to their high degree of sequence conservation (fig. 2).

Angiosperm Organellar aaRSs Are Present in Lower Gene Copy Numbers Than Their Cytosolic Counterparts

Gene duplication is pervasive in plants (Panchy et al. 2016), and the presence of duplicates potentially alters the selection pressures that can affect levels of sequence conservation (Lynch and Conery 2000). Therefore, we considered the possibility that organellar and cytosolic aaRS gene families systematically differ in copy number. Using a previously generated phylogenomic sampling of 20 angiosperm species (Forsythe et al. 2021), we compared the size of aaRS gene families and found that cytosolic families were 31% larger on average ($P = 0.0006$; paired t -test; fig. 3). Very similar results were obtained when the analysis was restricted to the 13 species for which whole genomes (and not just transcriptome assemblies) were available, with cytosolic families being 30% larger than organellar families ($P = 0.0026$; paired t -test). Therefore, the rate of gene duplication and/or retention appears to be higher for cytosolic aaRSs, potentially relaxing selection pressures on these gene copies. The larger number of copies for cytosolic aaRSs also parallels the very high level of paralogy for cytosolic ribosomal proteins relative to mt and pt ribosomal proteins (Yamaguchi et al. 2000; Yamaguchi and Subramanian 2000; Barakat et al. 2001; Bonen and Calixte 2006).

Rates of aaRS Sequence Evolution in *Sileneae*

To assess whether recent changes in subcellular targeting or tRNA substrates for aaRSs in *Sileneae* had altered selection pressures and rates of sequence evolution, we partitioned

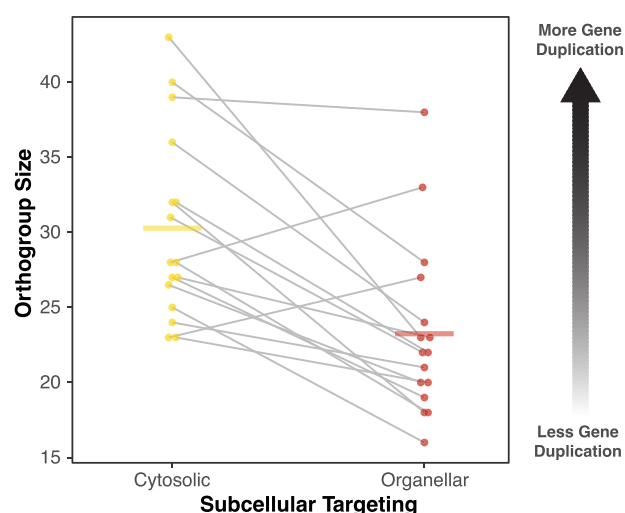


FIG. 3.—Higher rates of gene duplication/retention for cytosolic aaRSs than organellar aaRSs in angiosperms. The gene family size was assessed based on the number of genes in orthogroups from a previous analysis of angiosperm phylogenomic diversity (Forsythe et al. 2021). Points represent cytosolic (left, yellow) and organellar (right, red) aaRSs, and gray lines connect pairs which connect counterparts associated with the same amino acid. Horizontal bars represent the mean for each group. The cytosolic aaRS orthogroup size was significantly larger than the organellar orthogroup size ($P = 0.0006$; paired t -test).

aaRS gene trees to assign “foreground” and “background” branches. We defined foreground branches as those with an aaRS in one of the following categories: 1) an ancestral organellar aaRS inferred to charge a novel cytosolic tRNA substrate that is now imported into the mitochondria, 2) an ancestral cytosolic aaRS inferred to have gained mt targeting, or 3) an ancestral dual-targeted organellar aaRS inferred to have lost mt targeting and specialized on pt function.

We first tested the hypothesis that these foreground branches had experienced positive selection and accelerated evolution in the form of an increased ratio of nonsynonymous to synonymous substitutions (d_N/d_S) by performing branch tests on each tree. These tests generally failed to detect evidence that foreground branches had experienced faster protein sequence evolution (table 1). In two cases (organellar AspRS and cytosolic TyrRS), the nominal P value for increased rates on foreground branches was less than 0.05, but these comparisons did not remain significant after correction for multiple tests.

Because branch tests average d_N/d_S ratios across the entire length of a gene may dilute the effects of positive selection acting on a small subset of codons within a gene, we also applied a suite of branch site tests, which aim to detect positive selection even if it has favored amino acid substitutions at only a small number of sites: aBSREL (adaptive Branch-Site Random Effects Likelihood; Smith et al. 2015), BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification; Murrell et al. 2015), and MEME (Mixed Effects Model of Evolution; Murrell et al. 2012). However, these tests also yielded little evidence of positive selection or accelerated evolution on foreground branches. aBSREL tests each foreground branch for evidence that at least some proportion of codons have evolved under positive selection and only identified a single branch in a single tree (*Silene vulgaris* 49961 in the cytosolic TyrRS tree) as significant, but the raw P value of 0.0297 did not remain significant after correction for multiple tests (table 2). BUSTED performs a similar but more sensitive analysis, assessing whether there is evidence for positive selection at any site on any foreground branch. This analysis also identified the cytosolic TyrRS tree as only one with evidence of any positive selection (table 2). Finally, we used MEME to identify specific codons that may have experienced episodic positive selection on foreground branches, using a permissive (raw) significance threshold of $P < 0.10$. This analysis identified a relatively small number of codons as candidates for positive selection, again finding the most evidence of positive selection in the cytosolic TyrRS tree (table 2). Therefore, cytosolic TyrRS clearly emerged as the strongest candidate for positive selection, but the overall evidence for accelerated aaRS evolution in response to changes in subcellular targeting or tRNA substrates was limited.

Discussion

The Translational Demands of Photosynthesis and Their Effects on Rates of Organellar aaRS Sequence Evolution

An overall picture emerging from our analysis is that plant organellar aaRSs evolve under more intense constraints relative to cytosolic aaRSs than their counterparts in animals (Pett and Lavrov 2015; Adrion et al. 2016; Barreto et al. 2018). We hypothesize that this difference reflects the

fact that plant organellar aaRSs must also function in plastids (Duchêne et al. 2005) and that photosynthesis creates massive demands on pt gene expression and translational systems (Chotewutmontri and Barkan 2018; Heinemann et al. 2021; Forsythe et al. 2022). Therefore, it is likely that organellar aaRSs are under unusually strong purifying selection in plants to function efficiently.

An alternative hypothesis is that plant aaRSs experience less positive selection than animal mt aaRSs. The mt genomes of bilaterian animals evolve much more rapidly than organellar genomes in plants (Wolfe et al. 1987) and encode tRNAs that are unusually divergent in structure (Salinas-Giegé et al. 2015). Thus, it is possible that rapid evolution in animal mt aaRSs results from selection for co-evolutionary responses to changes in their cognate mt-tRNAs, whereas such selection pressures would be largely absent for plant organellar aaRSs because of the slow sequence evolution that is typical of plant mt and pt tRNAs. However, support for this hypothesis is limited. If individual mutations in animal mt-tRNAs are a driver of co-evolutionary responses and rapid evolution in mt aaRSs, then functionally replacing a mt-tRNA with an anciently divergent cytosolic counterpart would be expected to select for extensive changes in aaRS sequence. However, our analysis of the *Sileneae* organellar aaRS that appear to have adapted to charge newly imported cytosolic tRNAs found little or no evidence for accelerated aaRS evolution (tables 1 and 2). Likewise, previous studies have cast doubt on the hypothesis that mitonuclear coevolution is primarily responsible for the rapid evolution of animal mt aaRSs; in particular, the low expression levels of mt aaRS genes may be a more important contributor to relaxed constraints on sequence evolution (Pett and Lavrov 2015; Adrion et al. 2016; Barreto et al. 2018). Our analysis of transcript abundance for plant aaRS genes provided partial support for this role of gene expression. Although we found that plants did have lower expression levels for organellar aaRSs than for cytosolic aaRSs (fig. 2), the gap was substantially smaller than observed in animals, which may contribute to why plant organellar aaRSs do not evolve more slowly than their cytosolic counterparts (fig. 1).

Because most plant organellar aaRSs are dual-targeted (Duchêne et al. 2005), it is difficult to decouple the effects of mt and pt function on constraining their sequence evolution. However, some clues may come from comparing other components of organellar translation machinery that are distinct between mitochondria and plastids. For example, unlike organellar aaRSs, there are separate sets of ribosomal proteins for mt and pt translation rather than a single set of dual-targeted proteins. In accordance with the hypothesis that pt translational demands impose more functional constraint than mt translation, *Arabidopsis* pt ribosomal protein genes exhibit rates of nonsynonymous sequence divergence that are much lower than those in mt ribosomal protein genes

Table 1Branch Tests for Increased d_N/d_S Ratios in Response to Inferred Changes in Subcellular Localization or tRNA Substrate

aaRS	Ancestral Targeting	Hypothesized Mechanism	Foreground d_N/d_S	Background d_N/d_S	<i>P</i> value (raw)
AsnRS	Organellar	New substrate	0.14	0.08	0.1153
AspRS	Cytosolic	Mito import	0.16	0.26	0.4166
AspRS	Organellar	New substrate	0.28	0.17	0.0253
CysRS	Organellar	New substrate	0.15	0.15	0.8875
GlnRS	Cytosolic	Mito import	0.14	0.11	0.2301
GluRS	Organellar	New substrate	0.31	0.16	0.0853
HisRS	Organellar	New substrate	0.12	0.13	0.7518
LysRS	Cytosolic	Mito import	0.13	0.09	0.1473
MetRS	Organellar	Plastid specialization	0.27	0.26	1.0000
PheRS	Organellar	New substrate ^a	0.14	0.12	0.6390
ProRS	Cytosolic	Mito import	0.12	0.08	0.0719
TrpRS	Organellar	Plastid specialization	0.27	0.14	0.0979
TrpRS	Cytosolic	Mito import	0.10	0.17	0.1492
TyrRS	Cytosolic	Mito import	0.34	0.23	0.0088

NOTE.—Foreground branches for each test were assigned as described in the Materials and Methods section.

^aOrganellar PheRS could also be a potential case of pt specialization because *Sileneae* contains duplicate copies with one apparently functioning in the mitochondria and another apparently functioning in the plastids (Warren, et al. 2022), but this analysis was run with the putative mt copy set as the foreground branches.**Table 2**Summary of Branch Site Tests for Increased d_N/d_S Ratios in Response to Inferred Changes in Subcellular Localization or tRNA Substrate

aaRS	Ancestral Targeting	aBSREL Branches (<i>P</i> value)	BUSTED <i>P</i> value	MEME Codon Position(s)
AsnRS	Organellar	.	0.4165	.
AspRS	Cytosolic	.	0.0997	88;89
AspRS	Organellar	.	0.3811	7;141;370;439;582
CysRS	Organellar	.	0.5000	15;424
GlnRS	Cytosolic	.	0.1150	226;752;776
GluRS	Organellar	.	0.5000	83;311;385
HisRS	Organellar	.	0.5000	.
LysRS	Cytosolic	.	0.4042	242;539;581
MetRS	Organellar	.	0.3507	181;516;543
PheRS	Organellar	.	0.3601	175;225;363;373
ProRS	Cytosolic	.	0.2282	346;433
TrpRS	Organellar	.	0.5000	.
TrpRS	Cytosolic	.	0.5000	.
TyrRS	Cytosolic	<i>S. vulgaris</i> 49961 (0.0297)	0.0006	28;57;95;138;197;361;479;667;731;738

NOTE.—The trees and hypothesized mechanisms are the same as described in table 1. Foreground branches for each test were assigned as described in the Materials and Methods section. All reported *P* values are raw (i.e., uncorrected for multiple tests). Reported codon from MEME analyses is based on a raw significance threshold of $P < 0.1$.

but statistically indistinguishable from those in cytosolic ribosomal protein genes (after excluding the rapidly evolving transit peptide sequences from mt- and pt-targeted proteins) (Sloan et al. 2014).

One testable prediction of the hypothesis that translational demands associated with photosynthesis are the primary source of the selection constraining plant organellar aaRS sequence evolution is that heterotrophic plants that perform little or no photosynthesis (Wicke et al. 2013) will show faster evolutionary rates for organellar aaRSs than for cytosolic aaRSs—more akin to observations in animals. We are currently investigating the evolution of aaRS sequence evolution and subcellular targeting in parasitic plants and their photoautotrophic relatives to test this prediction. More generally, an

expanded sampling of plant taxa could test how different modes of photosynthesis (e.g., C3, C4, and CAM) and other features of plant physiology, metabolism, and life history affect the evolution organellar translation machinery. Our rate analysis consisted of only four species. Although these taxa included both monocots and eudicots and captured deep angiosperm divergence, the small number of species precludes testing for associations between biological traits and rates of aaRS evolution. One challenge that will need to be overcome in analyzing rates of aaRS evolution across large samples of species is to disentangle the history of recurring gene duplication in plants that makes it difficult to distinguish orthology/paralogy relationships and estimate lineage-specific rates.

Minimal Effects of Changes in Subcellular Targeting and tRNA Substrates on the Rate of aaRS Sequence Evolution

Despite the recent and major changes in subcellular targeting of tRNAs and aaRSs in *Sileneae* (Warren et al. 2021, 2022), we found little evidence that this rewiring of tRNA interaction networks has created positive selection for changes in aaRS protein sequence (tables 1 and 2). One possible explanation for this apparently limited effect is that *Sileneae* organellar aaRS enzymes that have evolved to charge a newly imported cytosolic tRNA substrate were already preadapted to successfully recognize these tRNAs and, thus, required few changes to enzyme sequence. For example, we previously showed that mt and cytosolic tRNAs already shared key identity elements in most cases where the ancestral organellar aaRS apparently retained mt function upon import of a cytosolic tRNA (Warren et al. 2022). We also found very limited evidence of positive selection in response to correlated retargeting of cytosolic tRNAs and aaRSs to the mitochondria. This may simply mean that the change in subcellular environment (cytosol vs. mitochondria) does not create strong selection for change in aaRS protein sequence because the ancestral aaRS–tRNA charging relationship is unchanged in these cases.

We should also note that some forms of positive selection may be difficult to detect with existing methods. Many of the current approaches to detect site-level positive selection were devised in the context of antagonistic coevolution, such as host–pathogen interactions, where selection for recurring amino acid substitutions at the same position(s) might be expected. However, in other forms of positive selection, a single substitution may be sufficient to improve and stabilize molecular function (Hughes 2007). Therefore, future work could more directly test whether recent changes in *Sileneae* aaRS sequences have altered their tRNA specificity by using *in vitro* charging assays. Likewise, genome editing approaches could be used to assess whether the ancestral cytosolic aaRS enzyme bodies are interchangeable with those of the *Sileneae* aaRSs that have since been retargeted to the mitochondria. Such approaches may be able to detect more subtle fine-tuning that is required to maintain function in response to perturbations to subcellular localization or tRNA substrates.

Materials and Methods

Angiosperm aaRS Sequence Curation, Alignment, Trimming, and Phylogenetic Rate Analysis

Gene identifiers and subcellular targeting data for *A. thaliana* aaRS sequences were taken from Duchêne et al. (2005) and Warren and Sloan (2020). Three types of aaRSs were excluded for the following reasons, which prevented a clean comparison between organellar and cytosolic aaRSs. 1) The only cytosolic AlaRS is also targeted to the organelles, 2) the only organellar ArgRS is also targeted the cytosol, and 3) there is no organellar GlnRS because plant mitochondria and plastids

typically use a bacterial-like indirect charging pathway that involves GluRS (Duchêne et al. 2005; Pujol et al. 2008). Four aaRSs (GlyRS, LeuRS, ThrRS, and ValRS) previously shown in *A. thaliana* to function in the cytosol but also having mt (but not pt) localization were classified as cytosolic for the purposes of this analysis, and the LeuRS with pt (but not mt) localization was classified as organellar (Duchêne et al. 2005). Amino acid sequences for aaRSs from *A. thaliana* and three other distantly related angiosperms (*O. sativa*, *S. polyrhiza*, and *V. vinifera*) were taken from “orthogroups” produced with OrthoFinder (Emms and Kelly 2015) in a previous study of 20 angiosperm species (Forsythe et al. 2021). See below for description of the full species sampling in this previous study. The relevant orthogroups were selected based on the presence of the *A. thaliana* gene identifiers described above.

Sequences for each aaRS gene family were aligned with MAFFT v7.453 using the --auto option. The resulting alignments were manually curated in Geneious (Kearse et al. 2012) to remove partial-length sequences and previously identified pseudogenes (Duchêne et al. 2005) and to merge fragmented gene models. Sequences were then realigned with MAFFT. The N-terminus of alignments for organellar aaRSs was trimmed to eliminate predicted transit peptides because these are known to evolve rapidly and are cleaved from the rest of the functional enzyme during the mt and/or pt import process. Trimming was performed at the alignment position corresponding to the cleavage site predicted for *A. thaliana* by TargetP v2.0 (Armenteros et al. 2019). After transit-peptide trimming, all alignments were manually inspected to trim poorly aligned regions at the N- and C-termini. In addition, poorly aligned internal regions in two sequences (*O. sativa* cytosolic TyrRS LOC_Os08g09260.1 and *V. vinifera* organellar PheRS GSVIVT01020339001) were also trimmed.

The resulting trimmed alignments were used for phylogenetic analysis with RAxML v8.2.12 (Stamatakis 2014) and the PROTGAMMALG model of sequence evolution. To compare evolutionary rates among aaRSs, the total tree length was calculated by summing individual branch lengths after averaging terminal branches for any paralogs to avoid inflating tree lengths for aaRSs with a history of gene duplication. Cytosolic PheRS is encoded as two separate subunits, so the total tree lengths for these subunits were averaged. Tree lengths for cytosolic versus organellar aaRS were compared with a paired *t*-test on log-transformed values in R v4.0.5.

aaRS Gene Copy Number Analysis

To test whether cytosolic and organellar aaRSs differed in gene copy number, we used the same orthogroup data set described above (Forsythe et al. 2021). The full data set consists of 20 diverse angiosperm species: *Acacia ligulata*, *Amborella trichopoda*, *A. thaliana*, *Cucumis sativus*, *Eucalyptus grandis*, *Geranium maderense*, *Gossypium raimondii*, *Helianthus annuus*, *Liriodendron chinense*, *Lobelia*

siphilitica, *Musa acuminata*, *Oenothera biennis*, *O. sativa*, *Plantago maritima*, *Populus trichocarpa*, *Prunus persica*, *Silene noctiflora*, *Solanum lycopersicum*, *S. polyrhiza*, and *V. vinifera*. We used the total number of sequences assigned to each orthogroup from all these species as a proxy for the relative gene copy number. AlaRS, ArgRS, and GlnRS were excluded from this comparison for the same reasons as described above for the rate analysis. In addition, CysRS was excluded because the organellar and cytosolic targeting classes were assigned to the same orthogroup by OrthoFinder. Copy number values for the two cytosolic PheRS subunits were averaged. Copy numbers for cytosolic versus organellar aaRS were compared with a paired *t*-test in R.

Arabidopsis thaliana Gene Expression Analysis

To analyze levels of aaRS gene expression (transcript abundance), data were obtained from the EMBL-EBI Expression Atlas (E-CURD-1: Araport 11—RNA-seq of *A. thaliana* Col-0 plants under different growth conditions from multiple studies) (Papatheodorou et al. 2018). Three different tissue types were chosen from this data set to be representative of diverse developmental stages:

- E-GEOD-30795: petal differentiation and expansion stage, long-day length regimen, and floral bud
- E-GEOD-44635: adult, long-day length regimen, and leaf
- E-MTAB-4242: seedling, long-day length regimen, and aerial part

Expression data were analyzed as transcripts per million (TPM). These TPM values were summed for paralogs. TPM values for the two cytosolic PheRS subunits were averaged. Log-transformed TPM data were analyzed with a three-way ANOVA using the *aov* function in R, with targeting (cytosolic or organellar), tissue (flower bud, leaf, or seedling), and aaRS (AsnRS, AspRS, CysRS, etc.) as independent variables.

Sileneae aaRS Positive Selection Analysis

Alignments of trimmed *Sileneae* aaRS nucleotide sequences and phylogenetic trees were previously generated with full-length cDNA (PacBio Iso-Seq) data from five species: *Agrostemma githago*, *Silene conica*, *Silene latifolia*, *S. noctiflora*, and *S. vulgaris* (Warren et al. 2022). The *A. thaliana* sequences that were originally used as outgroups were removed from the alignments. Coding sequences were then manually edited to be in frame and then aligned by codon (i.e., aligned with MAFFT as translated amino acid sequences and then back-translated to nucleotide sequences). Putative transit peptides for organellar aaRSs were trimmed based on predicted cleavage sites from TargetP.

To perform phylogenetic tests for selection on *Sileneae* aaRSs, *A. thaliana* genes were pruned from the original trees (Warren et al. 2022), using the *drop.tip* function in the ape

v5.4-1 package in R (Paradis and Schliep 2018). The phylo-tree.js tool (Shank et al. 2018) was then used to label branches as “Foreground” if they were inferred to meet one of the criteria described in the main text or table 1 (i.e., new tRNA substrate, gain of mt import by a cytosolic aaRS, or specialization on pt function). The cytosolic GluRS was excluded from subsequent analyses because the gain of mt targeting for this enzyme appears to be based on the addition of a transit peptide by alternative splicing. Therefore, there is no change in the gene body sequence relative to the enzyme that retains function in the cytosol.

Each aaRS sequence alignment and associated tree were used to perform branch site tests for positive selection with three different tools in HyPhy v2.5.31 (Kosakovsky Pond et al. 2020): aBSREL (Smith et al. 2015), BUSTED (Murrell et al. 2015), and MEME (Murrell et al. 2012). Each analysis was performed twice—both with and without the “--branches Foreground” option. In addition to the MEME selection test, the output from this analysis was also used to perform a simple branch test because the MEME runs provide maximum likelihood values for when the entire tree is constrained to a single d_N/d_S value and for when the foreground and background are assigned two different d_N/d_S . These values were compared with a likelihood ratio test to determine whether allowing different d_N/d_S ratios for the two tree partitions results in a significantly improved fit to the data.

Acknowledgments

We thank two anonymous reviewers for valuable comments on an earlier version of this manuscript. This work was supported by a grant from the National Science Foundation (MCB-2048407).

Data Availability

All data sets including expression TPM values, orthogroups sizes, and alignments and trees for rate analyses are available via https://github.com/dbsloan/aaRS_rates.

Literature Cited

- Adrion JR, White PS, Montooth KL. 2016. The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. *Mol Biol Evol.* 33:152.
- Armenteros JJA, et al. 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Sci Alliance.* 2:e201900429.
- Barakat A, et al. 2001. The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.* 127:398–415.
- Barreto F, et al. 2018. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat Ecol Evol.* 2:1250–1257.
- Bédard C, Cisneros AF, Jordan D, Landry CR. 2022. Correlation between protein abundance and sequence conservation: what do recent experiments say? *Curr Opin Genet Dev.* 77:101984.

- Bonen L, Calixte S. 2006. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. *Mol Biol Evol.* 23:701–712.
- Brindefalk B, Viklund J, Larsson D, Thollessen M, Andersson SG. 2007. Origin and evolution of the mitochondrial aminoacyl-tRNA synthetases. *Mol Biol Evol.* 24:743–756.
- Chotewutmontri P, Barkan A. 2018. Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. *PLoS Genet.* 14:e1007555.
- Duchêne A-M, et al. 2005. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 102:16484–16489.
- Duchêne A-M, Pujol C, Maréchal-Drouard L. 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Curr Genet.* 55:1–18.
- Emms DM, Kelly S. 2015. Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:1.
- Forsythe ES, et al. 2022. Organellar transcripts dominate the cellular mRNA pool across plants of varying ploidy levels. *Proc Natl Acad Sci U S A.* 119:e2204187119.
- Forsythe ES, Williams AM, Sloan DB. 2021. Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. *Plant Cell* 33:980–997.
- Giannakis K, et al. 2022. Evolutionary inference across eukaryotes identifies universal features shaping organelle gene retention. *Cell Syst.* 13:874–884.
- Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26:5017–5035.
- Guo W. 2014. Evolution of organellar genome architecture in seed plants: the role of intracellular gene transfer, recombination and mutation. Lincoln, Nebraska: University of Nebraska.
- Harris EH, Boynton JE, Gillham NW. 1994. Chloroplast ribosomes and protein synthesis. *Microbiol Rev.* 58:700–754.
- Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* 69:3069–3081.
- Heinemann B, Künzler P, Eubel H, Braun H-P, Hildebrandt TM. 2021. Estimating the number of protein molecules in a plant cell: protein and amino acid homeostasis during drought. *Plant Physiol.* 185:385–404.
- Hill GE. 2015. Mitonuclear ecology. *Mol Biol Evol.* 32:1927–1937.
- Hughes AL. 2007. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity (Edinb)* 99:364–373.
- Igloi GL. 2021. The evolutionary fate of mitochondrial aminoacyl-tRNA synthetases in amitochondrial organisms. *J Mol Evol.* 89:484–493.
- Igloi GL. 2022. Evolutionary adjustment of tRNA identity rules in *Bacillariophyta* for recognition by an aminoacyl-tRNA synthetase adds a facet to the origin of diatoms. *J Mol Evol.* 90:215–226.
- Kearse M, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Kosakovsky Pond SL, et al. 2020. Hyphy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol.* 37:295–299.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Meiklejohn CD, et al. 2013. An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLoS Genet.* 9:e1003238.
- Michaud M, Cognat V, Duchêne AM, Maréchal-Drouard L. 2011. A global picture of tRNA genes in plant genomes. *Plant J.* 66:80–93.
- Moreno P, et al. 2022. Expression Atlas update: gene and protein expression in multiple species. *Nucleic Acids Res.* 50:D129–D140.
- Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764.
- Murrell B, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol.* 32:1365–1371.
- Osada N, Akashi H. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol Biol Evol.* 29:337.
- Panchy N, Lehti-Shiu M, Shiu S. 2016. Evolution of gene duplication in plants. *Plant Physiol.* 171:2294–2316.
- Papatheodorou I, et al. 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* 46:D246–D251.
- Paradis E, Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528.
- Petersen G, Cuenca A, Möller I, Seberg O. 2015. Massive gene loss in mistletoe (*Viscum. Viscaceae*) mitochondria. *Sci Rep.* 5:17588.
- Pett W, Lavrov DV. 2015. Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol Evol.* 7:2089–2101.
- Pujol C, et al. 2008. Dual-targeted tRNA-dependent amidotransferase ensures both mitochondrial and chloroplastic Gln-tRNA_{Gln} synthesis in plants. *Proc Natl Acad Sci U S A.* 105:6481–6485.
- Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol (Amst).* 19:645–653.
- Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 11:29.
- Salinas-Giegé J, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. *Int J Mol Sci.* 16:4518–4559.
- Sanchez-Puerta MV, García LE, Wohlfeiler J, Ceriotti LF. 2017. Unparalleled replacement of native mitochondrial genes by foreign homologs in a holoparasitic plant. *New Phytol.* 214:376–387.
- Shank SD, Weaver S, Kosakovsky Pond SL. 2018. Phylotree.js—a JavaScript library for application development and interactive data visualization in phylogenetics. *BMC Bioinformatics* 19:276.
- Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A.* 112:E3515–E3524.
- Sloan DB, et al. 2018. Cytonuclear integration and co-evolution. *Nat Rev Genet.* 19:635–648.
- Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol.* 31:673–682.
- Small I, et al. 1999. The strange evolutionary history of plant mitochondrial tRNAs and their aminoacyl-tRNA synthetases. *J Hered.* 90:333–337.
- Smith MD, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 32:1342–1353.
- Smits P, Smeitink JA, van den Heuvel LP, Huynen MA, Ettema TJ. 2007. Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Res.* 35:4686–4703.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet.* 5:123–135.
- Warren JM, et al. 2021. Rapid shifts in mitochondrial tRNA import in a plant lineage with extensive mitochondrial tRNA gene loss. *Mol Biol Evol.* 38:5735–5751.

- Warren JM, et al. 2022. Rewiring of aminoacyl-tRNA synthetase localization and interactions in plants with extensive mitochondrial tRNA gene loss. *bioRxiv* 2022.01.27.478071v2.
- Warren JM, Sloan DB. 2020. Interchangeable parts: the evolutionarily dynamic tRNA population in plant mitochondria. *Mitochondrion* 52:144–156.
- Wicke S, et al. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25:3711–3725.
- Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84:9054–9058.
- Yamaguchi K, Subramanian AR. 2000. The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *J Biol Chem*. 275:28466–28482.
- Yamaguchi K, von Knoblauch K, Subramanian AR. 2000. The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J Biol Chem*. 275:28455–28465.
- Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. 16:409–420.

Associate editor: John Archibald