Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes

Daniel B. Sloan, Deborah A. Triant, Martin Wu, and Douglas R. Taylor
1Department of Biology, Colorado State University
2Department of Biology, University of Virginia
*Corresponding author: E-mail: dan.sloan@colostate.edu.
Associate editor: Stephen Wright

Abstract

Many mitochondrial and plastid protein complexes contain subunits that are encoded in different genomes. In animals, nuclear-encoded mitochondrial proteins often exhibit rapid sequence evolution, which has been hypothesized to result from selection for mutations that compensate for changes in interacting subunits encoded in mutation-prone animal mitochondrial DNA. To test this hypothesis, we analyzed nuclear genes encoding cytosolic and organelle ribosomal proteins in flowering plants. The model angiosperm genus Arabidopsis exhibits low organelle mutation rates, typical of most plants. Nevertheless, we found that (nuclear-encoded) subunits of organelle ribosomes in Arabidopsis have higher amino acid sequence polymorphism and divergence than their counterparts in cytosolic ribosomes, suggesting that organelle ribosomes experience relaxed functional constraint. However, the observed difference between organelle and cytosolic ribosomes was smaller than in animals and could be partially attributed to rapid evolution in N-terminal organelle-targeting peptides that are not involved in ribosome function. To test the role of organelle mutation more directly, we used transcriptomic data from an angiosperm genus (Silene) with highly variable rates of organelle genome evolution. We found that Silene species with unusually fast-evolving mitochondrial and plastid DNA exhibited increased amino acid sequence divergence in ribosomal proteins targeted to the organelles but not in those that function in cytosolic ribosomes. Overall, these findings support the hypothesis that rapid organelle genome evolution has selected for compensatory mutations in nuclear-encoded proteins. We conclude that coevolution between interacting subunits encoded in different genomic compartments within the eukaryotic cell is an important determinant of variation in rates of protein sequence evolution.

Key words: Arabidopsis, compensatory evolution, mutation rate, positive selection, ribosomal proteins, Silene.

Introduction

The staggering diversity of eukaryotic life is largely indebted to endosymbiotic events, namely the evolution of mitochondria and plastids from formerly free-living bacteria (Gould et al. 2008; Lane and Martin 2010). One consequence of the endosymbiotic history of eukaryotes is that their genetic material is divided among two or more genomic compartments. Most eukaryotic genes are encoded in the nucleus, as the ancestral genomes of mitochondria and plastids have experienced rampant endosymbiotic gene loss, transfer, and functional replacement (Timmis et al. 2004). However, cytoplasmic organelles retain relics of their ancient bacterial origins, with the few remaining genes encoding proteins and structural RNAs that are essential for mitochondrial and plastid function. As a result, many key enzymatic complexes, including the molecular machinery required for oxidative phosphorylation (OXPHOS), ATP synthesis, and photosynthesis, are consortia of organelle- and nuclear-encoded subunits.

The direct interaction of gene products from different genomic compartments raises fascinating questions about how nuclear and organelle genomes coevolve, especially in light of the fundamental differences in mutation rate, inheritance, recombination, and effective population size that can exist between the nucleus and organelles (Rand et al. 2004). In the face of high mutation rates and an effectively haploid and asexual mode of transmission, organelle genomes are prone to rapid accumulation of deleterious mutations by genetic drift (Lynch and Blanchard 1998; Neiman and Taylor 2009). The effect of organelle mutation accumulation on nuclear genome evolution has been the subject of tremendous interest (Blanchard and Lynch 2000; Rand et al. 2004; Turelli and Moyle 2007; Brandvain and Wade 2009; Burton and Barreto 2012), and a number of mitochondrial-targeted proteins in animals have been found to exhibit a history of positive selection and/or accelerated protein sequence evolution (Grossman et al. 2004; Willett and Burton 2004; Mishmar et al. 2006; Werren et al. 2010; Gagnaire et al. 2012; Osada and Akashi 2012; Barreto and Burton 2013). One explanation for the rapid evolution of mitochondrial-targeted proteins is the rapid evolution of the mitochondrial genome itself, which may create selection for compensatory changes in nuclear-encoded subunits that interact with mitochondrial genes. This idea has proven difficult to test. Although it is well established that cytonuclear incompatibilities accumulate between isolated populations and species, and in some cases these have been associated with specific nucleotide
substitutions (McKenzie et al. 2003; Schmitz-Linneweber et al. 2005; Harrison and Burton 2006; Lee et al. 2008; Meiklejohn et al. 2013), it is not clear whether mutation accumulation in organelle genomes and the resulting selection for compensatory changes are a major determinant of evolutionary rates in the nucleus.

Some of the strongest evidence for the occurrence of compensatory substitutions during cytonuclear coevolution was found in the cytochrome oxidase complex in primates by Osada and Akashi (2012), who showed that amino acid substitutions in mitochondrial-encoded proteins were often followed by changes at interacting sites in nuclear-encoded subunits. In addition, Barreto and Burton (2013) recently addressed the question of how cytonuclear interactions affect evolutionary rates by taking advantage of the fact that organelles contain their own ribosomes, which are distinct from the cytosolic ribosomes that translate nuclear transcripts. Ribosomes contain both protein and RNA subunits, and many of the proteins found in mitochondrial ribosomes are encoded in the nucleus. Barreto and Burton compared nuclear genes that code for mitochondrial ribosomal proteins (mRPs) versus cytosolic ribosomal proteins (cRPs) and found that the former evolve much more rapidly in both arthropods and yeast, two groups in which the mitochondrial mutation rate is much higher than in the nucleus (Haag-Liautard et al. 2008; Lynch et al. 2008; Oliveira et al. 2008; Willett 2012). This pattern confirms observations from early electrophoretic studies (Matthews et al. 1978; Pietromonaco et al. 1986) and is consistent with the interpretation that nuclear genes are under selection to “keep up” with the rapid evolution in mitochondrial-encoded ribosomal RNAs (rRNAs). However, it is difficult to disentangle confounding factors. For example, cRPs could simply be subject to more intense purifying selection because they are more highly expressed and functionally constrained (Pietromonaco et al. 1986; Subramanian and Kumar 2004; Drummond et al. 2005; Barreto and Burton 2013). Similar arguments were used to explain the observation that highly expressed mitochondrial-encoded components of the major OXPHOS complexes exhibit stronger evidence of purifying selection than their nuclear-encoded counterparts despite the reduced efficiency of selection in mitochondrial genomes (Nabholz et al. 2013; Zhang and Broughton 2013).

Here, we utilize genomic and transcriptomic data from plants to test the idea that compensatory mutations and cytonuclear coevolution are important determinants of molecular evolutionary rates in eukaryotes. Plants are particularly well suited to the study of cytonuclear interactions for a number of reasons (Moison et al. 2010; Greiner and Bock 2013). First, in addition to mitochondria, they contain plastids, an independently derived organelle of bacterial origin that also retains its own genome and translational machinery. Second, mutation rates in plant organelle genomes are typically much lower than in the nucleus (fig. 1 and table 1; Wolfe et al. 1987; Palmer and Herbon 1988; Drouin et al. 2008), offering a contrast to animals and other eukaryotic groups in which the opposite is true. For example, consistent
with the expectation that accumulation of deleterious mutations should slow under low mutation pressure, Lynch (1997) found that mitochondrial tRNAs are significantly less stable than their nuclear counterparts in animals and fungi but that no such difference exists in plants. Finally, although low rates of sequence evolution in organelle genomes are the rule in plants, there are scattered angiosperm lineages that have experienced massive accelerations, in some cases on very recent evolutionary timescales (Jansen et al. 2007; Mower et al. 2007).

A striking example of recent accelerations in mitochondrial mutation rate is found within the genus *Silene* (Caryophyllaceae). A subset of species in this genus, including *S. noctiflora* and *S. conica* have experienced accelerated organelle genome evolution in the last 5–10 My, resulting in dramatic rate differences among closely related lineages (Mower et al. 2007; Sloan et al. 2009; Sloan, Alverson, Chuckalovcak, et al. 2012; Sloan, Alverson, Wu, et al. 2012). Although both mitochondrial and plastid genomes exhibit abnormal evolution within this genus, the changes appear to be driven by different mechanisms. In *Silene* mitochondria, there is evidence of enormous mutation rate variation on a genome-wide scale, with approximately 100-fold differences in synonymous substitution rates across the genus (Sloan, Alverson, Chuckalovcak, et al. 2012). Although the high mitochondrial substitution rates in some *Silene* species rival those observed in animals, other *Silene* species have maintained slower rates that are typical of plants (Mower et al. 2007; Sloan et al. 2009). In contrast to the observed pattern in mitochondrial DNA, plastid genomes in some *Silene* species have experienced altered selection pressures with disproportionate increases in nonsynonymous substitution rates in a subset of genes, including plastid-encoded RPs (Sloan, Alverson, Wu, et al. 2012). The existence of closely related *Silene* species with highly divergent rates of organelle genome evolution provides an opportunity to investigate the effect of cytonuclear interactions and compensatory mutations on the evolution of mitochondrial- and plastid-targeted genes in the nucleus.

By analyzing dozens of nuclear-encoded proteins in each of the three types of ribosomes found in plant cells (cytosolic, mitochondrial, and plastid), we show that the organelle-targeted subunits evolve faster even in the absence of high mutation pressure in organelle genomes. However, we also find that the rate of evolution in organelle-targeted RPs is even higher in plant species with elevated rates of organelle genome evolution. In contrast, these species did not exhibit accelerations in cRPs, providing strong evidence that cytonuclear interactions and compensatory coevolution have a substantial impact on rates of amino acid substitution.

**Results**

*Arabidopsis* Organelle RPs Evolve Faster Than Their Cytosolic Counterparts

Genome-wide comparisons of nuclear-encoded RPs in *Arabidopsis* showed that cRPs have experienced a significantly lower ratio of nonsynonymous to synonymous substitutions ($d_N/d_S$) than either mRPs or plastid RPs (pRPs) (fig. 2). This pattern is qualitatively similar (albeit less extreme) to observations in animals and fungi (Barreto and Burton 2013) despite the much slower rate of organelle genome evolution in plants (fig. 1 and table 1). The differences in $d_N/d_S$ ratios are driven by variation in amino acid sequence divergence between organelle and cytosolic ribosomes, as $d_S$ values did not differ significantly among the three cellular compartments in *Arabidopsis* (table 2; $P > 0.05$ for all pairwise comparisons; Mann–Whitney U tests). This suggests that the analyzed gene sets all experience similar underlying mutation rates, which is consistent with the fact that they are all encoded in the same (nuclear) genome and that the respective loci are scattered across the *Arabidopsis* chromosomes, minimizing the possibility for systematic regional differences in mutation rate (Barakat et al. 2001; Bonen and Calixte 2006; Hu et al. 2011). Although selection of biased codon usage (which is especially strong in highly expressed genes like RPs) can create variation in $d_S$ even under constant mutation rates (Sharp and Li 1987), the absence of significant differences in $d_S$ suggests that variation in codon usage bias among RP genes has only a modest effect on synonymous substitution rates. As expected given the higher expression levels of cRPs, we found that codon usage bias is more extreme in these genes. The mean effective number of codons ($N_C$) from cRP genes (46.7) was lower than $N_C$ values from either mRPs (53.2) or pRPs (52.0). These
differences were both highly significant ($P < 0.0001$; t tests), whereas the difference between mRPs and pRPs was not ($t = 1.14$; df = 94; $P = 0.25$). If anything, the stronger selection on codon usage bias in cRP genes would be expected to disproportionately reduce their synonymous substitution rate and thereby inflate $d_N/d_S$ ratios. This is opposite the pattern observed in our data set in which cRP genes exhibited significantly lower $d_N/d_S$ ratios. We, therefore, conclude that the observed variation in $d_N/d_S$ was primarily caused by differential selection on amino acid substitutions rather than on synonymous substitutions.

Some of the observed difference between proteins targeted to the cytosol versus organelles can be attributed to the presence of an N-terminal signal peptide in mRPs and pRPs. These cleaved targeting peptides exhibited higher presence of an N-terminal signal peptide in mRPs and pRPs. To the cytosol versus organelles can be attributed to the

differential selection on amino acid substitutions rather than on synonymous substitutions.

Intraspecific Polymorphism in Arabidopsis thaliana Does Not Exhibit Evidence of a History of Positive Selection on Organelle Ribosomes

Elevated $d_N/d_S$ ratios can reflect positive selection, relaxed purifying selection, or a combination of the two. Patterns of intra- and interspecific sequence variation offer a useful tool to distinguish among these alternative possibilities. In particular, the rapid fixation of nonsynonymous substitutions under positive selection is expected to create elevated ratios of nonsynonymous to synonymous (N/S) changes for interspecific divergence relative to intraspecific polymorphism (McDonald and Kreitman 1991). However, analysis of polymorphism data within A. thaliana and divergence data between A. thaliana and A. lyrata did not reveal evidence of positive selection on proteins in any of the three ribosomes (table 3). Although pRPs did not differ significantly from the null expectation of equal N/S ratios for polymorphism and divergence, mRP and cRP gene sets both exhibited a significant excess of nonsynonymous polymorphisms. The polymorphism and divergence data were also used to calculate the mean direction of selection (DoS) and a modified version of the Neutrality Index (NITG) (Rand and Kann 1996; Stoletzki and Eyre-Walker 2011). All three ribosomes had NITG values greater than 1 and negative mean DoS values, which also indicate an excess of nonsynonymous polymorphisms (Stoletzki and Eyre-Walker 2011). These data suggest differences between polymorphism and divergence in Arabidopsis predominantly result from weakly deleterious alleles that segregate within the species but rarely spread to fixation—a pattern that is broadly representative of the A. thaliana genome and may reflect the demographic history and breeding system of this species (Bustamante et al. 2002). The similar elevation of N/S ratios for both polymorphism and divergence in mRPs relative to cRPs (table 3) suggests that relaxed purifying selection rather than positive selection explains the higher rate of protein sequence evolution in mRPs (table 2).

Table 2. Sequence Divergence in Nuclear-Encoded RP Genes.

<table>
<thead>
<tr>
<th>Species Comparison</th>
<th>Organelle Rate</th>
<th>Ribosome</th>
<th>Genes</th>
<th>$d_N$</th>
<th>$d_S$</th>
<th>$d_N/d_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. thaliana–A. lyrata</td>
<td>Low</td>
<td>Cytosolic</td>
<td>204</td>
<td>0.111</td>
<td>0.135</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mitochondrial</td>
<td>49</td>
<td>0.019</td>
<td>0.145</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plastid</td>
<td>40</td>
<td>0.013</td>
<td>0.153</td>
<td>0.087</td>
</tr>
<tr>
<td>S. latifolia–S. vulgaris</td>
<td>Low</td>
<td>Cytosolic</td>
<td>140</td>
<td>0.008</td>
<td>0.154</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mitochondrial</td>
<td>43</td>
<td>0.048</td>
<td>0.180</td>
<td>0.408</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plastid</td>
<td>34</td>
<td>0.010</td>
<td>0.156</td>
<td>0.088</td>
</tr>
<tr>
<td>S. conica–S. noctiflora</td>
<td>High</td>
<td>Cytosolic</td>
<td>140</td>
<td>0.005</td>
<td>0.149</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mitochondrial</td>
<td>43</td>
<td>0.069</td>
<td>0.148</td>
<td>0.647</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plastid</td>
<td>34</td>
<td>0.062</td>
<td>0.144</td>
<td>0.454</td>
</tr>
</tbody>
</table>

Note.—Reported values are means from analyses of individual genes performed after removal of predicted N-terminal targeting sequences.

Table 3. Nucleotide Polymorphism and Divergence in Arabidopsis Nuclear-Encoded RP Genes.

<table>
<thead>
<tr>
<th>Ribosome</th>
<th>Data Type</th>
<th>Nonsynonymous</th>
<th>Synonymous</th>
<th>N/S</th>
<th>MK</th>
<th>NITG</th>
<th>DoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosolic</td>
<td>Polymorphism</td>
<td>586</td>
<td>1,532</td>
<td>0.38</td>
<td>$P &lt; 0.001$</td>
<td>1.67</td>
<td>−0.06</td>
</tr>
<tr>
<td></td>
<td>Divergence</td>
<td>852</td>
<td>3,680</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mitochondrial</td>
<td>Polymorphism</td>
<td>210</td>
<td>290</td>
<td>0.72</td>
<td>$P &lt; 0.001$</td>
<td>1.67</td>
<td>−0.17</td>
</tr>
<tr>
<td></td>
<td>Divergence</td>
<td>399</td>
<td>824</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plastid</td>
<td>Polymorphism</td>
<td>64</td>
<td>275</td>
<td>0.23</td>
<td>$P &gt; 0.5$</td>
<td>1.14</td>
<td>−0.04</td>
</tr>
<tr>
<td></td>
<td>Divergence</td>
<td>163</td>
<td>671</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.—Data do not include sites in predicted N-terminal targeting sequences.
Silene Species with Rapidly Evolving Organelle Genomes Exhibit Accelerated Sequence Change in Nuclear-Encoded Proteins in Organelle Ribosomes but Not in Cytosolic Ribosomes

Similar to observations in Arabidopsis, cRPs in Silene exhibited evidence of strong purifying selection (fig. 3). Amino acid sequence divergence in cRPs was very low in pairwise comparisons between S. conica and S. noctiflora and only marginally higher in comparisons between S. latifolia and S. vulgaris (table 2; \( Z = 2,416.5; P = 0.015 \); Wilcoxon-signed rank test). In both species comparisons, \( d_{s}/d_{s} \) ratios were significantly elevated in mRPs and pRPs (relative to cRPs) even after trimming sequences that encode targeting peptides (\( P < 0.05 \) for all comparisons; Mann–Whitney \( U \) tests), but the difference was much more pronounced in the pair of species with accelerated organelle genome evolution (fig. 3 and table 2). The \( d_{s}/d_{s} \) ratios for mRPs and pRPs were both significantly higher in S. conica/S. noctiflora than in S. latifolia/S. vulgaris (table 3; \( P < 0.0001 \) for both mRPs and pRPs; Wilcoxon signed rank tests). The differences in \( d_{s}/d_{s} \) values in Silene reflected variation in amino acid substitution rates, as \( d_{s} \) values did not differ significantly among gene sets (table 3; \( P > 0.05 \) for all comparisons; Mann–Whitney \( U \) tests).

![Fig. 3. Rapid evolution of nuclear-encoded RPs in Silene species with elevated rates of organelle genome evolution. The black circles, red triangles, and green squares represent cRPs, mRPs, and pRPs, respectively. The triangle highlighted in black represents the cytosolic-like RPS15A gene that has functionally replaced the mitochondrial rps8 gene in angiosperms (Adams, Daley, et al. 2002). The grouping of this gene with other mRP genes indicates that evolutionary rates are more a reflection of a protein’s functional role rather than any fundamental difference between RPs of bacterial versus archaean/eukaryotic ancestry. The diagonal is a 1:1 line, showing where \( d_{s}/d_{s} \) ratios are equal for species with fast (S. conica/S. noctiflora) and slow (S. latifolia/S. vulgaris) evolving organelle genomes. Values were calculated after removal of predicted N-terminal targeting sequences. Note that three outliers (all mitochondrial) with \( d_{s}/d_{s} \) values of 2.0 or greater were not plotted to improve readability.

Nuclear-Encoded Succinate Dehydrogenase (Complex II) Proteins Do Not Exhibit Elevated \( d_{s}/d_{s} \) Ratios in Silene Species with High Mitochondrial Mutation Rates

The proteins that make up OXPHOS complex II (succinate dehydrogenase [SDH]) represent a classic tool for decoupling the effects of mitochondrial targeting from the effects of interactions with mitochondrial-encoded gene products (e.g., Ellison and Burton 2006). Unlike the other OXPHOS and ATP synthase complexes, SDH is composed solely of nuclear-encoded proteins in most eukaryotes (some angiosperms retain the SDH genes sdh3 and sdh4 in their mitochondrial genomes, but this is not the case in Silene; Sloan, Alverson, Chuckalovcak, et al. 2012). Therefore, SDH genes are involved in core mitochondrial function, but they should experience little or no effects of cytonuclear coevolution resulting from direct interactions with mitochondrial-encoded gene products.

Comparisons of sequence divergence in seven SDH protein genes did not find evidence of elevated \( d_{s}/d_{s} \) values in Silene species with high mitochondrial mutation rates. For three of the seven genes, pairwise \( d_{s}/d_{s} \) values were higher for S. conica/S. noctiflora than for S. latifolia/S. vulgaris, but the opposite was true for the other four genes. The \( d_{s}/d_{s} \) estimates from a concatenation of all seven genes were statistically indistinguishable between the two species comparisons (S. conica/S. noctiflora = 0.164; S. latifolia/S. vulgaris = 0.142; \( D = 0.80; P = 0.37; \) likelihood ratio test).

Discussion

It has long been hypothesized that accelerated rates of amino acid substitution in nuclear-encoded mitochondrial proteins result from selection to compensate for rapid evolutionary change in the mitochondrial genome. This idea has emerged predominantly from work on bilaterian animals, in which mitochondrial mutation rates are consistently higher than in the nucleus (Grossman et al. 2004; Willett and Burton 2004; Mishmar et al. 2006; Werren et al. 2010; Gagnaire et al. 2012; Osada and Akashi 2012; Barreto and Burton 2013). However, it is still unclear how common elevated mitochondrial mutation rates are outside of animals (Sloan and Taylor 2012), and rapid sequence evolution in organelle genomes is certainly not the norm in plants.

Our analysis of plant RP evolution across three different cellular compartments revealed that species with high rates of organelle genome evolution exhibited rapid change in RPs targeted to organelles, providing compelling evidence that cytonuclear interactions and compensatory coadaptation can have a large effect on the rate of sequence evolution in the nucleus. Alternative explanations for this pattern, such as a genome-wide reduction in the efficacy of selection in these species, can be rejected because there is no evidence of elevated \( d_{s}/d_{s} \) values in the proteins that make up cytosolic ribosomes. In addition, two lines of evidence indicate that it is unlikely that these species have experienced a general relaxation of selection on mitochondrial function. First, protein genes encoded in the mitochondrial genome itself...
show no sign of elevated \(d_\mu/d_S\) values despite the apparent increases in the underlying mitochondrial mutation rates in these species (Sloan, Alverson, Chuckalovcak, et al. 2012). Second, nuclear-encoded SDH genes, which function in mitochondrial OXPHOS reactions but do not form complexes with mitochondrial-encoded gene products, also fail to exhibit elevated \(d_\mu/d_S\) values. Our findings generate the additional testable prediction that, as a result of positive selection, organelle-targeted RP genes in high-rate \textit{Silene} species will exhibit an excess of nonsynonymous divergence relative to intraspecific polymorphism (McDonald and Kreitman 1991). This would require polymorphism data in \textit{Silene} for dozens of RP genes, which unfortunately are not currently available. Our results, therefore, highlight the potential value of generating population genomic data to disentangle the causes and consequences of cytonuclear evolution.

Interestingly, we also found higher rates of protein evolution in mRPs and pRPs (relative to cRPs) in the absence of strong organelle mutation pressure. Therefore, cRPs appear to be under more functional constraint than their organelle counterparts, and the rapid amino acid sequence evolution in eukaryotic mRPs likely represents a combination of weaker purifying selection and positive selection for compensatory mutations.

\textbf{Cytonuclear Coevolution and the Rate of Protein Sequence Evolution}

Data from species with divergent rates of organelle genome evolution can be used to approximate the contribution of compensatory adaptation to amino acid substitution rates in nuclear-encoded proteins. If we assume, unrealistically but conservatively (discussed later), that all else is equal and that none of the substitutions observed in \textit{Silene} species with low rates of organelle genome evolution are positively selected compensatory changes, the difference in \(d_\mu/d_S\) ratios between \textit{Silene} species comparisons can be used to estimate the effect of compensatory evolution. The mean \(d_\mu/d_S\) ratio for mRPs is 37% lower in \textit{S. latifolia}/\textit{S. vulgaris} than in \textit{S. conica}/\textit{S. noctiflora}. We, therefore, estimate that selection for compensatory changes is directly or indirectly responsible for at least one-third of the amino acid substitutions in the rapidly evolving \textit{S. noctiflora} and \textit{S. conica} mRPs. Not all these substitutions, however, were necessarily the target of adaptive evolution because positive selection reduces the efficiency of purifying selection at closely linked sites, and can lead to an increase in the rate of fixation of weakly deleterious mutations (Hill and Robertson 1966; Gillespie 2000). Therefore, some of the substitution rate increase is likely an indirect byproduct of selection for compensatory mutations. Although a single statistical contrast between high rate and low rate \textit{Silene} species offers an admittedly rough way to estimate the magnitude of compensatory coevolution, the extensive variation in mitochondrial substitution rates in independent eukaryotic lineages (Mower et al. 2007; Nabholz et al. 2008, 2009) should provide the means to more robustly model the relationship between mitochondrial mutation rates and the fraction of amino acid substitutions that are driven by cytonuclear interactions.

The rate variation among \textit{Silene} species in pRPs is much greater than in mRPs, but interpreting this result is more complex. In contrast to \textit{Silene} mitochondrial genomes, in which selection acting on mitochondrial-encoded genes appears to be largely unaltered despite massive changes in mutational input (Sloan, Alverson, Chuckalovcak, et al. 2012), the rapid evolution of plastid-encoded translational machinery in \textit{S. conica} and \textit{S. noctiflora} suggests significant changes in selection pressures (Sloan, Alverson, Wu, et al. 2012). Therefore, the accelerated evolution of nuclear-encoded pRPs could reflect compensatory changes in response to plastid genome evolution, but it could also be a direct cause or consequence of the altered selection pressures that have increased substitution rates in plastid-encoded genes. Disentangling these alternative explanations may require a more mechanistic understanding of the altered selection pressures acting on plastid translational machinery in \textit{S. conica} and \textit{S. noctiflora}.

One of the central premises of this study is that the low rates of nucleotide substitution that typically occur in plant organelle genomes will significantly alter the dynamics of cytonuclear interactions relative to other eukaryotic lineages with much more mutation-prone organelle genomes (e.g., animals). It is, therefore, worth considering sources of genomic instability other than point mutations that could affect cytonuclear coevolution in plants. For example, the process of mitochondrial and plastid gene loss/transfer/replacement is still active in many angiosperm lineages and has had an especially large effect on organelle translational machinery (Millen et al. 2001; Adams, Qiu, et al. 2002). Although evidence is limited for positive selection on organelle genes that have recently been transferred to the nucleus (Liu et al. 2009), the broader effects of gene transfer on compensatory coevolution between organelle- and nuclear-encoded genes should be investigated. It is also noteworthy that, although most plant mitochondrial genomes experience exceptionally low rates of nucleotide substitution, they are subject to frequent structural mutations including indels (table 1; Palmer and Herbon 1988).

The peculiarities of plant organelle genome evolution are particularly relevant to cytonuclear coevolution in \textit{Silene}, a group in which organelle genome structure is highly labile (Sloan, Alverson, Chuckalovcak, et al. 2012; Sloan, Alverson, Wu, et al. 2012; Sloan, Muller, et al. 2012). \textit{Silene} and related angiosperm genera have experienced especially rapid loss of genes encoding mitochondrial translation machinery, a process that is often associated with functional transfer to the nucleus (Adams, Qiu, et al. 2002; Sloan et al. 2010; Sloan, Alverson, Chuckalovcak, et al. 2012). The mitochondrial-to-nuclear transfer process appears to have been caught in the act in \textit{S. vulgaris}, with two mRP genes still encoded in the mitochondrial genome in some individuals but not in others (Sloan, Muller, et al. 2012). Notably, even \textit{Silene} species with relatively slow mitochondrial genome evolution have experienced rapid divergence in the mitochondrial 5S rRNA gene and replacement of a portion of the mitochondrial 18S rRNA gene with the homologous sequence from the plastid 16S rRNA gene (Sloan et al. 2010). For all these reasons, some
degree of selection for compensatory mutations in nuclear genes should be expected even in lineages with “slowly” evolving organelle genomes. Notably, \( d_N/d_S \) values for mRPs were substantially higher in \( S. \) latifolia/\( S. \) vulgaris than in \( A. \) thaliana (table 2), which could be caused by some of the unusual patterns of mitochondrial genome evolution that are observed throughout \( S. \) silene.

**Relaxed Functional Constraint in Organelle Ribosomes**

We found that plant cRPs are subject to intense purifying selection, which is not surprising given that they are among the most highly expressed proteins in eukaryotes (Pal et al. 2001). A key implication of our study is that mRPs experience faster evolutionary rates, in part, because of relaxed functional constraint. This finding has some intuitive appeal, as mitochondrial ribosomes are responsible for translating only the small number of genes in the mitochondrial genome, whereas cytosolic ribosomes must synthesize tens of thousands of nuclear-encoded proteins. Interestingly, pRPs exhibited substantially lower substitution rates than mRPs, suggesting that they evolve under intermediate levels of functional constraint (table 2). This may reflect the fact that plastids are highly protein rich (Shikanai et al. 2001), and, although cytosolic ribosomes synthesize much of this protein content, many of the most abundant plant proteins, such as the large subunit of RubisCO, are translated by plastid ribosomes. The functional differences between mitochondria and plastids make plants a particularly attractive system to tease apart the complex effects of mutation and selection on cytonuclear coevolution.

**Materials and Methods**

**rRNA Sequence Analysis**

To compare rates of sequence evolution in nuclear, mitochondrial, and plastid genomes, we analyzed large and small subunit rRNA genes from a diverse sample of angiosperms. Sequences were collected from published sources (supplementary table S1, Supplementary Material online). In addition, we obtained consensus rRNA sequences for \( A. \) lyrata by mapping publicly available short-read data (NCBI SRX1583737) against \( A. \) thaliana reference sequences for nuclear, mitochondrial, and plastid rRNA genes using SOAP v2.21 with gap (g) and mismatch (v) parameters set to 3 and 5, respectively (Li et al. 2009). Recursive mappings were performed to determine \( A. \) lyrata sequences in more divergent gene regions. Sequences were aligned with MUSCLE v3.7 (Edgar 2004) and phylogenetic branch lengths were estimated with the baseml package in PAML v4.7 (Yang 2007), using a GTR model of evolution and a constrained topology.

**Arabidopsis RPs**

Nuclear-encoded RPs in \( A. \) thaliana were identified from the literature and the Plant Proteome Database (Yamaguchi and Subramanian 2000; Yamaguchi et al. 2000; Barakat et al. 2001; Bonen and Calixte 2006; Sun et al. 2009), and the corresponding coding sequences were obtained from Phytozone v9.1 (Goodstein et al. 2012). Three genes with complex or ambiguous subcellular targeting were excluded from the data set: the cytosolic-like RPS15Ab that may be targeted to the mitochondria but has not been confirmed experimentally (At2g19720), the plastid-like RPS13 gene that has functionally replaced its mitochondrial counterpart (At1g77750), and the cytosolic-like RPL10 gene that is dual-targeted to both organelles (At3g12370) (Adams, Daley, et al. 2002; Mower and Bonen 2009; Kubo and Arimura 2010). To estimate the extent of codon usage bias in RP genes, CodonW v1.4.4 was used to calculate \( Nc \) for each \( A. \) thaliana sequence (Wright 1990; Peden 2000).

To identify orthologs of \( A. \) thaliana RPs in \( A. \) lyrata, we performed reciprocal searches with NCBI-BlastN v2.2.24 + against all annotated \( A. \) lyrata coding sequences (Hu et al. 2011), which were obtained from Phytozone. For a small number of RPs, there appeared to have been recent gene duplication in either \( A. \) thaliana or \( A. \) lyrata since the divergence of these two species. In such cases, one of the two duplicate copies was randomly chosen for the analysis. To compare levels of interspecific polymorphism with interspecific divergence in RPs, single nucleotide polymorphisms with a quality score of at least 25 and a minimum of five supporting reads were obtained from 80 \( A. \) thaliana genomes from the POLYMORPH database (Cao et al. 2011).

**Silene Transcriptome Sequencing and Assembly**

We selected four species in the genus \( S. \) silene with extreme differences in rates of organelle genome evolution, including two (\( S. \) noctiflora and \( S. \) conica) with very recent accelerations and two (\( S. \) latifolia and \( S. \) vulgaris) with much slower rates that are more typical of plants (Sloan, Alverson, Chuckalovcak, et al. 2012; Sloan, Alverson, Wu, et al. 2012). Plants were grown under greenhouse conditions with supplemental lighting and regular watering and fertilization. Total cellular RNA was extracted from leaf tissue from a single individual of each species using the Qiagen Plant RNAeasy Kit. Two rounds of selection with Sera-Mag Oligo(dT) Magnetic Particles (Thermo Scientific) were used to enrich for mRNAs. These samples were then used for RNA-Seq library construction following standard Illumina protocols with the NEBNext mRNA Library Prep Reagent Set (New England Biolabs) and a multiplexed barcoding strategy developed at the University of California Davis Genome Center. To normalize transcript abundance, each of the libraries was treated with Duplex-Specific Nuclease (Evrogen) according to manufacturer’s instructions with hybridization performed for 4 h at 70°C. Normalized samples were then reamplified with 15 PCR cycles. The resulting libraries were sequenced in multiplexed lanes on an Illumina HiSeq 2000 (\( 2 \times 100 \) bp) at the University of Minnesota Genomics Center and the Yale Center for Genome Analysis. The resulting sequence data were demultiplexed using the FASTX-Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/, last accessed September 24, 2013) and custom Perl scripts and assembled using Trinity r2012-06-08 (Grabherr et al. 2011). Illumina reads were deposited in the NCBI Sequence Read Archive (SRX353031, SRX353047, SRX353048, and SRX353050).
Silene Ribosomal and SDH Proteins

Reciprocal blast searches (TBLastN and BlastP) between Arabidopsis protein sequences and the Silene transcriptome assemblies were used to identify mRPs and pRPs. Identified sequences were manually curated to exclude spurious hits and to merge transcripts that had been fragmented into multiple contigs. Plant genomes typically contain two to four paralogs for each cRP, complicating the identification of orthologous gene sets. We found that the presence of paralogs resulted in many fragmented transcripts and chimeric artifacts in the Silene transcriptome assemblies. Therefore, we reassembled the transcriptome data using SOAPdenovo-Trans v1.02 (Xie et al. 2013) with a relatively large k-mer value (51) to reduce the number of assembly chimeras (Yang and Smith 2013). SOAPdenovo-Trans was run with default assembly parameters and the gap-filling option enabled. cRP gene families were identified by NCBI-TBLastN search against the new Silene transcriptome assemblies and subjected to extensive manual curation. Fragmented transcripts were merged with the aid of individual sequencing reads and contigs from the original Trinity assemblies. Families of related cRP genes were aligned and neighbor-joining trees were constructed with MEGA v5.01 (Tamura et al. 2011) to distinguish orthologs and paralogs within the four Silene species. Genes were excluded from further analysis if a copy was not recovered from all four Silene species or if the orthology/paralogy relationships were unclear.

To identify SDH genes in the Silene transcriptome data sets, the corresponding A. thaliana protein sequences were obtained from the Arabidopsis Mitochondrial Protein Database (Heazlewood et al. 2004) and used for reciprocal blast searches. SDH genes were extracted and aligned using the same approach described for cRP genes. Although eight different types of SDH subunits have been identified in A. thaliana, SDH8 has not been found in other angiosperms examined to date (Millar et al. 2004; Huang and Millar 2013). This was also the case in Silene, as we only identified homologs of first seven SDH subunits.

Identification of N-Terminal Targeting Sequences

Nuclear-encoded proteins that are targeted to the mitochondria or plastids generally contain an N-terminal targeting peptide that is cleaved upon import into the organelles. The presence of these additional sequences in mRPs and pRPs represents a potentially confounding difference in comparisons with cRPs. Therefore, we generated modified alignments by removing putative targeting peptides that were identified using the cleavage site prediction feature on the TargetP 1.1 Server (Emanuelsson et al. 2007). Predicted targeting peptides were also removed from Silene SDH genes prior to data analysis.

Sequence and Statistical Analysis of Protein Genes

Amino acid sequences from orthologous RP genes were aligned with MUSCLE and then converted back to nucleotide sequences to produce in-frame nucleotide alignments. Pairwise nonsynonymous (dN) and synonymous (dS) divergences per site were estimated for each RP gene using codeml in PAML with an F1x4 model of codon frequencies. Analyses were conducted on three different species pairs: A. thaliana versus A. lyrata, S. latifolia versus S. vulgaris, and S. noctiflora versus S. conica. Differences in divergence values and dN/dS ratios were analyzed with nonparametric tests to avoid normality assumptions about distributions across genes or species comparisons. Differences between genomic compartments were analyzed with Mann–Whitney U tests. Wilcoxon-signed rank tests were also performed to compare rates across different species pairs and between sets of mRPs and pRPs before and after removal of their N-terminal targeting sequences. Differences in N among cRPs, mRPs, and pRPs were analyzed with t-tests. Analyses were run using the wilcox.test and t.test functions implemented in R v2.15.2.

Sequence divergence in dN/dS ratios in Silene SDH genes were estimated in the same fashion. However, because of the smaller sample size (seven genes), statistical comparisons between the Silene species were conducted differently. The seven SDH genes were concatenated into a single alignment, and a likelihood ratio test was used to compare two PAML models: one in which pairwise dN/dS ratios were independently estimated for S. conica/S. noctiflora and S. latifolia/S. vulgaris and a second in which both species pairs were constrained to have identical dN/dS ratios.

Polymorphism and divergence data in Arabidopsis were analyzed with MK tests (McDonald and Kreitman 1991). For each gene set, a Fisher’s exact test was used to test for a significant difference in the ratios of nonsynonymous to synonymous changes in intraspecific (within A. thaliana) vs. interspecific data (between A. thaliana and A. lyrata). Polymorphism analyses were performed after removing sequences coding for predicted targeting peptides. RP, SDH, and rRNA sequence alignments used in this study are provided as supplementary material (supplementary file S1, Supplementary Material online), and gene names and aligned sequence lengths are summarized in (supplementary table S2, Supplementary Material online).

Supplementary Material

Supplementary file S1 and tables S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Laura Bergner for laboratory assistance with transcriptome library construction and Peter Fields for sharing transcriptome assembly data and for helpful discussion regarding assembly methods. They also thank the Associate Editor and three anonymous reviewers for their comments on an earlier version of this manuscript. This work was supported by the National Science Foundation (MCB-1022128). Computational resources were provided by the IStE Cray HPC System at Colorado State University, which is supported by the National Science Foundation (CNS-0923386).
References


