

Causes and Consequences of Rapidly Evolving mtDNA in a Plant Lineage

Justin C. Havird^{1,*}, Paul Trapp¹, Christopher M. Miller¹, Ioannis Bazos², and Daniel B. Sloan^{1,*}

¹Department of Biology, Colorado State University, Fort Collins, CO

²Department of Ecology and Systematics, National and Kapodistrian University of Athens, Panepistimiopolis, Greece

*Corresponding authors: E-mails: justin.havird@colostate.edu; dbsloan@rams.colostate.edu.

Accepted: January 30, 2017

Data deposition: Newly generated RNA-seq data and transcriptome assemblies were deposited into the NCBI SRA and TSA databases (see [supplementary table S3](#), [Supplementary Material](#) online, for accession numbers). Sequence alignments are provided via Figshare at <https://figshare.com/s/64e28b2ed14177b1ab14>.

Abstract

Understanding mechanisms of coevolution between nuclear and mitochondrial (mt) genomes is a defining challenge in eukaryotic genetics. The angiosperm genus *Silene* is a natural system to investigate the causes and consequences of mt mutation rate variation because closely related species have highly divergent rates. In *Silene* species with fast-evolving mtDNA, nuclear genes that encode mitochondrially targeted proteins (N-mt genes) are also fast-evolving. This correlation could indicate positive selection to compensate for mt mutations, but might also result from a recent relaxation of selection. To differentiate between these interpretations, we used phylogenetic and population-genetic methods to test for positive and relaxed selection in three classes of N-mt genes (oxidative phosphorylation genes, ribosomal genes, and “RRR” genes involved in mtDNA recombination, replication, and repair). In all three classes, we found that species with fast-evolving mtDNA had: 1) elevated d_N/d_S , 2) an excess of nonsynonymous divergence relative to levels of intraspecific polymorphism, which is a signature of positive selection, and 3) no clear signals of relaxed selection. “Control” genes exhibited comparatively few signs of positive selection. These results suggest that high mt mutation rates can create selection on N-mt genes and that relaxed selection is an unlikely cause of recent accelerations in the evolution of N-mt genes. Because mt-RRR genes were found to be under positive selection, it is unlikely that elevated mt mutation rates in *Silene* were caused by inactivation of these mt-RRR genes. Therefore, the causes of extreme increases in angiosperm mt mutation rates remain uncertain.

Key words: mitonuclear interactions, cytonuclear interactions, organelle DNA, mitochondrial mutations, nuclear compensation hypothesis, McDonald–Kreitman test.

Introduction

With one recent and notable exception (Karnkowska et al. 2016), all eukaryotes maintain mitochondria or a mitochondrial(mt)-like organelle, and those with a functional electron transport chain maintain a mt genome (Allen 2015; Hjort et al. 2010). The evolutionary forces shaping the mt genome are quite different than those that affect the nuclear genome. The mt genome has a lower effective population size than the nuclear genome, owing mainly to its uniparental transmission and effectively haploid nature, which limits the ability of selection in the mt genome to purge deleterious mutations (Lynch and Blanchard 1998; Neiman and Taylor 2009). These effects can also be magnified due to a high mt mutation rate in many eukaryotic lineages (Brown et al. 1979; Denver

et al. 2000; Lynch et al. 2008; Lynch 1997; Lavrov and Pett 2016).

Despite the disparate evolutionary forces acting on them, gene products encoded by the mt and nuclear genomes must interact in a precise and intimate fashion to meet the energetic demands of eukaryotic cells. When these interactions are compromised due to mt mutations or the generation of maladapted mt and nuclear genomes via hybridization, the consequences for organismal health and fitness can be severe (Burton et al. 2013; Dowling 2014; Wallace 2005). Therefore, it is thought that mt and nuclear genomes coevolve in order to maintain proper mitonuclear interactions. In particular, rapidly evolving mtDNA may create selection for corresponding changes in

© The Author(s) 2017. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

nuclear-encoded proteins that interact with mt-encoded gene products (Barreto and Burton 2013; Beck et al. 2015; Gong et al. 2014; Havird et al. 2015b; Osada and Akashi 2012; Sloan et al. 2014). We refer to the nuclear genes that encode these mitochondrially targeted proteins as N-mt genes.

The angiosperm genus *Silene* is a model system to examine the effects of mitochondrial mutations on the nucleus. Like most angiosperms (Wolfe et al. 1987), some *Silene* species have slowly evolving mtDNA, but others have experienced a recent (i.e., within ~5 Myr) acceleration in mtDNA evolution. This acceleration appears to result from a genome-wide increase in the underlying mt mutation rate, though the specific cause of the elevated mutation rates remains mysterious. The resulting mt mutation rates in some fast *Silene* species are higher than those for nuclear genes, creating a more mammalian-like mitonuclear mutation balance (Mower et al. 2007; Sloan et al. 2012a). If N-mt genes must coevolve in response to mt mutations, then “fast” *Silene* (those with elevated mt mutation rates) should show evidence of positive selection in N-mt genes compared with their “slow” relatives. Previous work has supported this prediction, as the ratio of nonsynonymous to synonymous substitutions (d_N/d_S) is higher in N-mt genes in fast vs. slow *Silene* for OXPHOS and ribosomal proteins (Havird et al. 2015b; Sloan et al. 2014). Elevated d_N/d_S ratios are consistent with positive selection on N-mt genes in response to elevated mt mutation rates. However, there are important alternative interpretations to consider when evaluating evidence for positive selection and mitonuclear coevolution (Adrion et al. 2015; Nabholz et al. 2013; Pett and Lavrov 2015; Popadin et al. 2013; Sloan et al. 2014; Zhang and Broughton 2013). In particular, elevated d_N/d_S ratios could alternatively imply a recent relaxation of purifying selection on N-mt genes in fast *Silene* species.

Disentangling the effects of positive and relaxed purifying selection on N-mt genes requires a diversity of approaches. One underutilized technique in the study of mitonuclear interactions is to test for positive selection on N-mt genes in a population genetic framework. For example, the McDonald-Kreitman (MK) test checks for an excess of nonsynonymous divergence between species relative to intraspecific polymorphism within species, which is a strong sign of positive selection (McDonald and Kreitman 1991). However, previous studies using this approach to test for positive selection on N-mt genes have yielded mixed results (Willett and Burton 2004; Gagnaire et al. 2012; Parmakelis et al. 2013) and in some cases may have suffered from analyzing a very limited number of N-mt genes (i.e., only one or two).

Another class of population genetic tests examines the site frequency spectrum (SFS) to determine whether rare alleles are present at higher or lower frequencies than would be expected under neutrality (e.g., Tajima’s D; (Tajima 1989). Tajima’s D and related tests are complementary to MK tests and d_N/d_S analyses because Tajima’s D provides a more

current snapshot of ongoing evolution within a species, whereas d_N/d_S ratios and MK tests detect older bouts of selection based on divergence among species. New methods have also been developed to explicitly test for relaxed selection in a phylogenetic framework by examining the distribution of nonsynonymous and synonymous mutations among taxa of interest (Wertheim et al. 2014). To date, however, tests based on the SFS and tools that investigate relaxed selection are rarely applied to N-mt genes to study mitonuclear coevolution.

It is also unclear why mt mutation rates have undergone accelerations within *Silene* and independently in other angiosperm lineages (Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007). One obvious reason could stem from altered repair, recombination, and replication (RRR) of mtDNA in angiosperm lineages with accelerated mt mutation rates. Supporting this possibility, fast *Silene* species show altered frequencies of mtDNA recombination and gene conversion (Sloan et al. 2012a). Mutated RRR genes can often lead to genome instability (Marechal and Brisson 2010), and fast *Silene* species also tend to show massive expansions in mt genome size and altered mt genome structures (Sloan et al. 2012a). With the exception of DNA polymerases involved in replication of linear plasmids in some plant mitochondria (Warren et al. 2016), mt-RRR machinery is encoded in the nucleus, raising the possibility that changes in nuclear mt-RRR genes are responsible for elevated mtDNA mutation rates. One possibility is that mt-RRR genes have been pseudogenized or lost outright in fast *Silene* species. Alternatively, these genes may have remained intact but been subject to reduced functional constraint or inefficient selection, resulting in the accumulation of mutations that alter their function. If this is the case, we would predict elevated d_N/d_S in mt-RRR genes in fast *Silene* species compared to their slow relatives but with signatures of relaxed, rather than positive, selection. Although positive and relaxed selection have not been examined in mt-RRR genes from angiosperms with elevated mt mutation rates, Zhang et al. (2016) recently reported an analysis of plastid-targeted RRR genes in the Geraniaceae, echoing some of the above hypotheses about the role of RRR genes in accelerated mutation rates in organelle genomes.

Here, we used transcriptomic data from 12 *Silene* species with variable mt mutation rates (Fig. 1), including 19 populations of *S. conica*, to investigate the roles of positive and relaxed purifying selection on N-mt OXPHOS, ribosomal, and RRR genes. Overall, we find evidence of positive selection in these nuclear genes, which we argue is likely due to increased mtDNA mutation rates. Moreover, we find little evidence for relaxed selection, suggesting positive selection is the predominant driver of accelerated evolution in N-mt genes of fast *Silene* species. We discuss this finding in light of mitonuclear interactions and possible causes of elevated mitochondrial mutation rates in *Silene*.

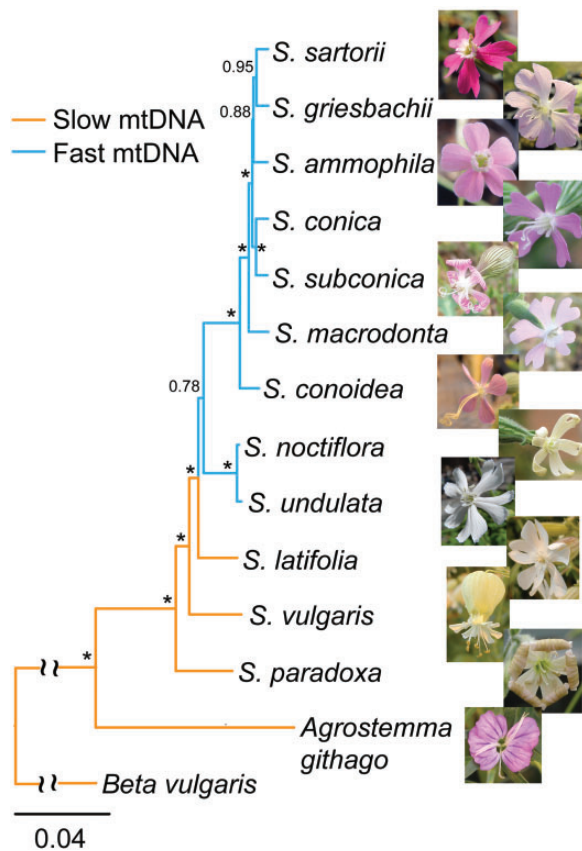


FIG. 1.—The relationships between taxa utilized in this study, as shown by a maximum likelihood phylogeny based on a concatenation of 44 nuclear genes (60,579 bp). Species with fast-evolving mtDNA are highlighted with blue branches, while those with slow-evolving mtDNA are orange. Note that *S. paradoxa* was encoded as a “slow” species here and in RELAX analyses because it does not have a mt genome-wide acceleration in d_s , but does show elevated rates in a subset of mt genes (Sloan et al. 2009). Moreover, there may have been independent accelerations in the lineage leading to *S. noctiflora*/*S. undulata* and in the lineage leading to *S. conoidea*, etc. This is evidenced by low bootstrap support (0.78) for the sister relationship between these clades, and because a sister relationship was not recovered in previous analyses with more taxa, but fewer loci (Sloan et al. 2009; Rautenberg et al. 2012). This phylogeny was obtained by concatenating the 44 DNA sequences from the random nuclear genes that were used in RELAX analyses (Fig. 5) and inputting the concatenation into RAxML version 8.2.6 using the “GTRGAMMAI” model of evolution (Stamatakis 2014). Support values are from 1,000 “rapid” bootstrap replicates (Stamatakis et al. 2008) and the scale bar represents substitutions/site. Asterisks indicate 100% support. Photos are from greenhouse-reared individuals except for *S. subconica*.

Materials and Methods

New *Silene* Transcriptomic Resources

To develop new transcriptomic resources for *Silene*, we obtained seeds and/or leaf tissue from *S. ammophila*, *S. conoidea*, *S. griesbachii*, *S. sartorii*, *S. subconica*, and *S. undulata* (Fig. 1;

supplementary table S1, Supplementary Material online). For *S. subconica*, 4–5 rosette leaves were preserved in RNAlater (Sigma) from a single individual and frozen at -20°C for ~ 1 week prior to transport to Colorado State University, where they were frozen at -80°C for ~ 4 months prior to RNA extraction via an RNeasy Plant Mini Kit (Qiagen). For the other species, seeds were germinated on soil (Fafard 2SV mix supplemented with vermiculite and perlite) and grown under a 16-hr/8-hr light/dark cycle with regular watering and fertilizer treatments in greenhouse facilities at Colorado State University. After growing for 5 months, RNA was extracted from 4–5 rosette leaves from a single individual of each species as described above.

Resulting RNA was sent to the Yale Center for Genome Analysis for cDNA library preparation and Illumina sequencing. A Ribo-Zero Plant Leaf rRNA Removal Kit (Illumina) was used during library preparation in order to capture organellar transcripts, which are not consistently poly-adenylated, for future studies. The resulting six strand-specific cDNA libraries were then sequenced on a single Illumina HiSeq 2500 lane. Paired-end, 151 bp reads were delivered in FASTQ format and are available via the sequence read archive (SRA) at the National Center for Biotechnology Information (NCBI) (supplementary table S2, Supplementary Material online). Reads were used as delivered (i.e., no quality filtering was performed because filtering can reduce assembly quality; MacManes 2014) to assemble strand-specific transcriptomes *de novo* for each species using Trinity r20140717 (Grabherr et al. 2011), with default parameters and *in silico* normalization of reads. Resulting contigs were deposited in NCBI’s transcriptome shotgun assembly (TSA) database (supplementary table S2, Supplementary Material online).

In addition to these new resources, we made use of previously generated *Silene* transcriptomes. For *S. macrodonta*, *S. latifolia*, *S. paradoxa*, *S. noctiflora*, *S. vulgaris*, and the outgroup *Agrostemma githago* (Fig. 1), we used previously described sequence resources (Rockenbach et al. 2016), based on assembled sets of contigs generated *de novo* using Trinity r20140717. We also used transcriptomes for 19 geographically diverse populations of *S. conica* that were recently described in Rockenbach et al. (2016). In some analyses, we also made use of the *Beta vulgaris* genome (Dohm et al. 2013) as an additional outgroup by downloading publicly available coding sequences from The *Beta vulgaris* Resource (<http://bvseq.molgen.mpg.de/index.shtml>; last accessed June 1, 2016). In total, we examined 14 species (Fig. 1), nine of which were *Silene* species with fast-evolving mtDNA, while the remaining five (*A. githago*, *B. vulgaris*, *S. latifolia*, *S. paradoxa*, and *S. vulgaris*) have slowly evolving mtDNA typical of most other angiosperms.

Curation of Orthologous OXPHOS, RRR, and Ribosomal Sequences

We were interested in three classes of nuclear-encoded genes whose products interact with mtDNA-encoded proteins or

mtDNA itself: 1) OXPHOS, 2) ribosomal, and 3) mt-RRR genes. N-mt OXPHOS proteins form multisubunit complexes with mtDNA-encoded proteins and are responsible for electron transport and ATP generation. Proteins that make up complex II (succinate dehydrogenase; SDH) serve as a control for mitochondrial interactions in this gene class, as all complex II proteins are nuclear-encoded in *Silene* (Sloan et al. 2012a). BLAST query sequences (see below) were obtained from *Silene* OXPHOS sequences reported in Havird et al. (2015b). N-mt ribosomal proteins form multisubunit complexes with mtDNA-encoded ribosomal proteins and rRNAs and are responsible for mt translation. We used the *Silene* ribosomal protein sequences from Sloan et al. (2014) as BLAST queries. Mt-RRR gene sequences from *Arabidopsis* as reported in [supplementary file S1](#) of Zhang et al. (2016), Table 1 of Cupp and Nielsen (2014), and Boesch et al. (2009) were used as queries for BLAST searches ([supplementary table S3](#), [Supplementary Material](#) online). We focused on RRR genes that were targeted solely to the mitochondria, or those that were dual-targeted to both mitochondria and plastids.

BLAST 2.2.29+ (Camacho et al. 2009) was used with tblastn (with a cutoff expect value of $1e-5$) and the query sequences mentioned above to extract orthologous sequences from the *Silene* and outgroup transcriptomes, as well as transcriptomes from the 19 populations of *S. conica*. Top BLAST hits were extracted using custom Perl scripts and BioPerl modules (Stajich et al. 2002), and were aligned to the reference query sequence using MUSCLE 3.8.31 (Edgar 2004) as implemented in MEGA 6 (Tamura et al. 2013). After manual curation and preliminary tree building in MEGA to ensure all sequences were orthologous, the alignment was trimmed according to the longest ORF. Sequences that were < 50% complete were excluded. After manual curation, final alignments were re-aligned by codon using MUSCLE and any stop codons were removed. Finally, TargetP (Emanuelsson et al. 2007) was used to predict and remove signal peptides for mt-targeted proteins before analyses.

McDonald–Kreitman and Other Population Genetic Tests

The MK test uses counts of synonymous and nonsynonymous substitutions both within species (polymorphism) and between species (divergence) to test for positive selection (McDonald and Kreitman 1991). Specifically, the neutrality index (NI) (Rand and Kann 1996) is calculated by dividing the ratio of nonsynonymous to synonymous polymorphisms (P_r/P_s) by the ratio of nonsynonymous to synonymous divergence (D_r/D_s), with NI values less than one indicative of an excess of nonsynonymous divergence and positive selection. Note that the D and P terms in these expressions represent simple counts. We distinguish these from the d_N and d_S values described previously, which are defined on a per site basis. We implemented MK tests using the PopGenome package in R (Pfeifer et al. 2014) on datasets for each of the three classes of

genes, consisting of the 19 *S. conica* populations and using either *S. macrodonta* or *S. latifolia* as outgroups. These two outgroup species were chosen to represent two different scales of interspecific divergence: within (*S. macrodonta*) and outside (*S. latifolia*) the section *Conoimorpha*. NI values were assessed for statistical significance using Fisher's Exact Tests. Direction of Selection (DoS) values were calculated following Stoletzki and Eyre-Walker (2011), with positive DoS values indicative of positive selection. We also report MK-associated statistics from a set of ribosomal proteins targeted to the cytosol, which act as a specific control for the mt-targeted ribosomal proteins. Fifty randomly chosen, single-copy nuclear genes were also investigated as described previously (Rockenbach et al. 2016) to serve as a general control for mitochondrial interactions. To categorize positive selection on an entire class of genes, we summed counts for genes belonging to each gene class. Because summing counts across contingency tables can introduce statistical bias, we also calculated NI_{TG} , which is an unbiased estimator of NI (Stoletzki and Eyre-Walker 2011).

In addition to MK tests, we calculated several other population genetic metrics to assess if and how selection was currently acting on these genes. Tajima's D (Tajima 1989) and Fu and Li's D and F (Fu and Li 1993) (with associated simulation tests for statistical significance) were calculated using the batch mode and the outgroup options in DnaSP 5.10.01 (Librado and Rozas 2009). These tests do not explicitly test for positive selection but rather for a deviation from neutrality by taking the SFS into account. Negative values indicate positive selection and/or population expansion after a bottleneck or selective sweep, while positive values indicate balancing selection and/or a decrease in population size. Using the DH program with default settings (Zeng et al. 2007), we also calculated Fay and Wu's H (Fay and Wu 2000), which utilizes an outgroup to explicitly test for positive selection, and DHEW, which is a similar test that combines Tajima's D , Fay and Wu's H , and the Ewens-Watterson test of neutrality (Watterson 1978). Negative H and DHEW values are indicative of positive selection. For these statistics (Tajima's D , Fu and Li's D , Fu and Li's F , Fay and Wu's H , and DHEW), we generated test statistics and P -values for each gene as well as concatenated alignments for the gene categories described above. These tests were performed twice for each gene and dataset: once for the sequences used in MK analyses when *S. latifolia* served as an outgroup, and again for the sequences where *S. macrodonta* was used as an outgroup. Note that even though many of these tests do not employ an outgroup, the results for both gene sets are presented as a comparison with MK tests. A smaller gene set was used for the *S. macrodonta* analyses because homologous sequences could not be extracted from *S. macrodonta* for several genes (most likely owing to lower coverage of nuclear transcripts because RiboZero was used in cDNA library preparation). In the main text we report uncorrected P values, although all statistics, including P values corrected for multiple comparisons via Bonferroni

correction, are reported in [supplementary data file S1, Supplementary Material](#) online.

Measuring d_N/d_S in *Silene* RRR Genes with Slow and Fast-Evolving mtDNA

d_N/d_S values in *Silene* species with fast-evolving mtDNA are larger than in species with slow-evolving mtDNA for both N-mt OXPHOS and ribosomal genes (Sloan et al. 2014; Havird et al. 2015b). To investigate if this applies to N-mt RRR genes, we estimated d_N/d_S for mt-RRR genes between two species with fast-evolving mtDNA (*S. conica* and *S. noctiflora*) and two species with slowly evolving mtDNA (*S. latifolia* and *S. vulgaris*), using the same strategy as we did previously (Havird et al. 2015b). Briefly, all mt-RRR genes were concatenated from each species and pairwise d_N/d_S values between either slow or fast species were estimated using codeml in PAML 4.8 (runmode = -2; codon frequency = $F1 \times 4$) (Yang 2007). To test if the fast group had a significantly different d_N/d_S than the slow group, the slow group's d_N/d_S was constrained to that of the fast group and a likelihood ratio test between constrained and unconstrained models was performed.

We also performed the same analyses on four other previously described datasets: OXPHOS genes, mt-targeted ribosomal proteins, cytosolic-targeted ribosomal proteins, and the set of 50 randomly chosen, single-copy nuclear genes. For the OXPHOS genes, complex II genes were analyzed separately from the genes in the rest of the complexes. Although we have previously examined d_N/d_S in these datasets (Sloan et al. 2014; Havird et al. 2015b; Rockenbach et al. 2016), our methods here were slightly different than in previous studies (e.g., different taxa, concatenation schemes, or models in PAML were used), so we reanalyzed these datasets so they would all be processed with the exact same methods.

Testing for Relaxed Selection

Because elevated d_N/d_S in N-mt genes could be due to relaxed, rather than positive selection, we tested for signatures of relaxed selection in N-mt genes using the RELAX program (Wertheim et al. 2014) as implemented in HyPhy (Pond et al. 2005). RELAX detects relaxed selection by fitting different distributions of d_N/d_S on reference and test branches in a phylogeny, and testing whether this model fits the data significantly better than a single distribution across all branches. RELAX distributes sites in an alignment among three classes of d_N/d_S : those presumed to be under purifying selection ($d_N/d_S < 1$), neutrality ($d_N/d_S \approx 1$), and positive selection ($d_N/d_S > 1$). Under relaxed selection, test branches should have a d_N/d_S distribution skewed towards the neutrality class when compared to the reference branches (i.e., for test branches, the d_N/d_S value in the purifying selection class will increase and the d_N/d_S value in the positive selection class will decrease). Under intensified selection test branches will have a d_N/d_S distribution skewed away from neutrality. By raising each

d_N/d_S class of the reference branches to the exponent k , the corresponding d_N/d_S class of the test branches is obtained. Therefore, $k > 1$ and $k < 1$ indicate intensified and relaxed selection, respectively.

For each gene, custom Perl scripts were used to extract orthologs from the 14 species we examined (12 *Silene* species, *Agrostemma githago*, and *Beta vulgaris*). Briefly, sequences were extracted and aligned from *Silene* and outgroup transcriptomes as above. Phylogenies were then created for each alignment/gene using PhyML v3.1 (Guindon et al. 2010). Reference branches were then selected to include all terminal branches leading to species with slowly evolving mtDNA (*Beta*, *Agrostemma*, *S. latifolia*, *S. vulgaris*, and *S. paradoxa*), while test branches were set to include terminal branches leading to species with fast-evolving mtDNA (*S. ammophila*, *S. conica*, *S. conoidea*, *S. grisebachii*, *S. macrodonta*, *S. noctiflora*, *S. sartorii*, *S. subconica*, and *S. undulata*). Any internal branches that led exclusively to one type of branch were designed as that type, while internal branches giving rise to both types of branches were designated as reference branches (e.g., Fig. 1). In addition to individual gene analyses, RELAX analyses were performed after concatenating sequences within each gene class as above.

Data Availability

Newly generated sequence reads and transcriptome assemblies have been deposited into NCBI's SRA and transcriptome shotgun assembly (TSA) database, respectively (see [supplementary table S2, Supplementary Material](#) online, for accession numbers). All statistics generated for population genetic and RELAX analyses are available in [supplementary data file S1, Supplementary Material](#) online. Alignments used for MK, PAML, and RELAX analyses are available via FigShare at <https://figshare.com/s/64e28b2ed14177b1ab14>.

Results

Elevated d_N/d_S in N-Mt Genes in Fast *Silene* Species

In previous studies, we documented elevated d_N/d_S in N-mt genes for *Silene* with fast-evolving mtDNA compared to close relatives with slowly evolving mtDNA, specifically in N-mt ribosomal genes (Sloan et al. 2014) and N-mt OXPHOS genes encoding proteins in complexes I and III-V (Havird et al. 2015b). However, this pattern was not found in genes encoding cytosolic-targeted ribosomal proteins (Sloan et al. 2014), OXPHOS complex II proteins (Havird et al. 2015b), or a random set of 50 proteins not targeted to mitochondria or plastids (Rockenbach et al. 2016). Here, we used slightly different methods to reanalyze these gene sets in a standardized fashion, and the results confirm our previous analyses. Briefly, N-mt genes showed elevated d_N/d_S ratios in fast compared to slow *Silene* species, while control genes showed either no difference between fast and slow species or the opposite trend (Fig. 2).

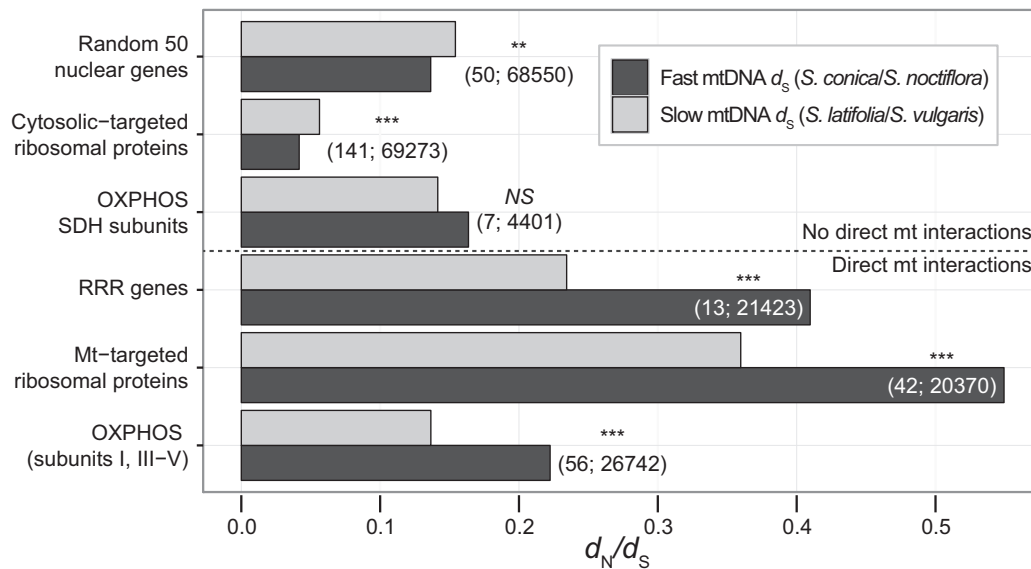


Fig. 2.— d_N/d_S for nuclear-encoded genes (N-mt genes) that code for proteins targeted to the mitochondria, except cytosolic-targeted ribosomal proteins and a set of 50 randomly chosen nuclear genes, which act as controls for mitonuclear interactions. SDH subunits in OXPHOS complex II act as another type of control, as they are targeted to the mitochondria, but do not interact directly with any mt-encoded subunits. Comparisons between *Silene* species with fast (*S. conica* and *S. noctiflora*) or slowly evolving mtDNA (*S. latifolia* and *S. vulgaris*) are shown, along with results of a likelihood ratio test between the two groups (NS not significant; ** $P < 0.01$; *** $P < 0.001$). Numbers in parentheses indicate the number of genes and nucleotides analyzed in each gene set.

Silene mt-RRR genes were identified and analyzed here for the first time (supplementary table S3, Supplementary Material online). They showed a similar pattern as the other N-mt genes (Fig. 2). Mt-RRR genes in *Silene* with fast-evolving mtDNA had significantly higher d_N/d_S than in species with slowly evolving mtDNA ($P < 0.001$, likelihood ratio test). Importantly, this trend was not driven by outlier genes, because when genes were analyzed individually, all mt-RRR genes showed the same pattern (supplementary fig. S1, Supplementary Material online). The difference in d_N/d_S between the fast and slow species was the largest for *TWINKLE* and *POL1*, which are involved in replication, and *RECA2*, *RECA3*, and *RECG*, which are involved in recombination (supplementary fig. S1, Supplementary Material online).

Population Genetic Signatures of Positive Selection in N-Mt Genes

When performing MK tests using the 19 populations of *S. conica* with either *S. latifolia* or *S. macrodonta* as an outgroup, a significant excess of nonsynonymous divergence was found in several datasets when summing counts across all genes (Fig. 3, supplementary fig. S2, Supplementary Material online, Table 1). This pattern is a signature of positive selection (McDonald and Kreitman 1991). For mt-interacting OXPHOS subunits (i.e., complexes I, III-V), there was a significant excess of nonsynonymous divergence between *S. conica* and *S. latifolia*. Similar, although nonsignificant, results were obtained when the closer relative *S. macrodonta* was used as an

outgroup. Similar results were also found in the complex II (SDH) OXPHOS subunits even though these subunits do not interact directly with any mt-encoded gene products and did not show elevated d_N/d_S values in fast *Silene* species (Fig. 2). Mt-targeted and cytosolic-targeted ribosomal proteins both showed a significant excess of nonsynonymous divergence when *S. latifolia* was used as the outgroup, but to a greater extent in the mt-targeted gene set (Fig. 3, Table 1). For mt- but not cytosolic-targeted ribosomal proteins, a significant excess of nonsynonymous divergence was also found when using *S. macrodonta* as an outgroup (Fig. 3, Table 1). For mt-RRR genes, a significant excess of nonsynonymous divergence was found when using either *S. latifolia* or *S. macrodonta* as an outgroup (Fig. 3, Table 1). As described previously (Rockenbach et al. 2016), the set of 50 randomly chosen nuclear-encoded genes shows neither an excess nor paucity of nonsynonymous divergence (Fig. 3, Table 1).

We also calculated several metrics that examine the SFS. When performing Tajima's D tests, none of the concatenated sets of genes in any category had significantly more or less variation than would be expected under neutrality (i.e., $D = 0$), although there was a general trend towards an excess of rare alleles in all datasets (i.e. $D < 0$), which may be indicative of a population expansion after a recent bottleneck (Fig. 4). Fu and Li's D and F statistics, Fay and Wu's H, and the related DHEW showed similar, nonsignificant results as in Tajima's D (supplementary figs. S3–S6, Supplementary Material online).

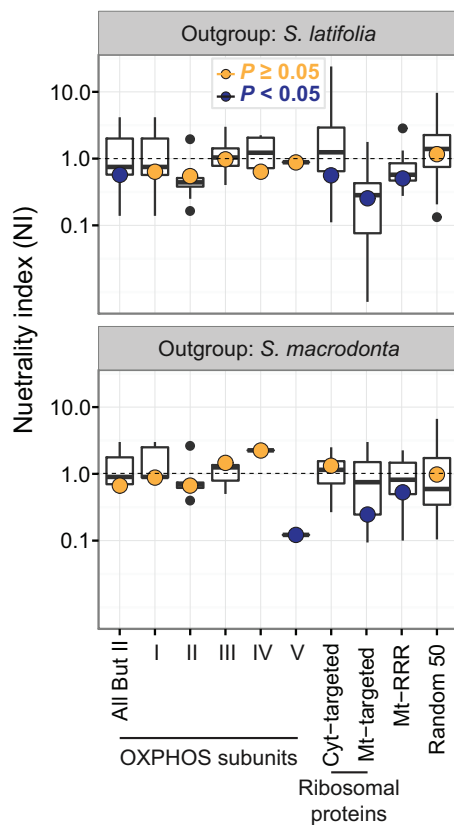


Fig. 3.—Results of McDonald–Kreitman tests based on 19 populations of *Silene conica*, using either *S. latifolia* or the closer relative *S. macrodonta* as outgroups. All gene sets include nuclear-encoded genes (N-mt genes) that code for proteins targeted to the mitochondria, except cytosolic-targeted ribosomal proteins and a set of 50 randomly chosen nuclear genes, which act as controls for mitonuclear interactions. Subunits in OXPHOS complex II act as another type of control, as they are targeted to the mitochondria, but do not interact directly with any mt-encoded subunits. Neutrality indices (NI) above or below 1 (indicated by dashed lines) suggest purifying or positive selection, respectively. Larger points indicate normalized NI_{IG} values for each gene set, colored by significance of the summed NI value (via Fisher’s Exact Tests). Upper and lower “hinges” of boxplots correspond to the first and third quartiles of the data, whiskers extend to the most extreme values within 1.5-fold of the inter-quartile range, and values beyond these are plotted as points. Note that the wide variation in mt-targeted ribosomal proteins is due to a single gene (RPS11) with an abnormally low NI (0.007), see [supplementary data file S1, Supplementary Material](#) online.

Similar Signatures of Relaxed Selection in N-Mt and Control Genes

OXPHOS subunits that interact with mt-encoded subunits, mt-targeted ribosomal proteins, and mt-targeted RRR genes exhibited k values in RELAX analyses that were generally clustered around 1, and genes that were individually significant were roughly evenly split between k values greater than and less than 1 (Fig. 5). In addition, the distribution of k values was

similar between N-mt genes and the set of 50 random nuclear genes (Fig. 5). These results indicate that neither relaxed nor intensified selection is acting in any consistent way on these sets of genes in fast compared with slow *Silene* species.

When all genes within a dataset were concatenated, there were significant signatures of relaxed selection in fast *Silene* species for some gene sets, but overall these effects were limited (Fig. 5). For mt-interacting OXPHOS subunits, a significant signature of relaxed selection was obtained ($k=0.73$, $P<0.001$), although it was similar in magnitude to the OXPHOS subunits from complex II ($k=0.70$). For mt-targeted ribosomal proteins, there was no indication whatsoever of relaxed selection ($k=1.02$, $P=0.72$). For mt-RRR genes, a significant signature of relaxed selection was obtained ($k=0.58$, $P<0.001$), although, perplexingly, individual mt-RRR genes all had higher k values (mean $k=1.68$). To investigate this discrepancy between individual mt-RRR genes and the concatenated set, we examined the distribution of d_N/d_S values estimated by RELAX for all the concatenated gene sets ([supplementary fig. S7, Supplementary Material](#) online). For mt-RRR genes, the d_N/d_S ratio for sites under positive selection in the reference branches (species with slowly evolving mtDNA) was estimated as 9999.9999, which is an obvious artifact and likely the upper limit on d_N/d_S ratios estimated by RELAX. This high d_N/d_S ratio, which was not observed in the other concatenated analyses, is responsible for producing an unexpectedly low k value for the concatenated mt-RRR analysis. There are several potential reasons why this artifact may have been produced (see Discussion).

Presence of Functional RRR Genes in *Silene*

One possibility for why some *Silene* species have undergone an acceleration in mtDNA mutation rate and structural rearrangements in the mitochondrial genome is that mt-RRR genes are absent or have altered functionality in these species. Of the mt-RRR genes that have been characterized in other angiosperms ([supplementary table S3, Supplementary Material](#) online), we were able to identify homologs for all of them in *Silene* species with both fast and slowly evolving mtDNA. None of these genes had obvious frameshift mutations, premature stop codons, or other hallmarks of pseudogenization or loss of function. Two mt-DNA polymerases (*POL1A* and *POL1B*) are present in *Arabidopsis thaliana*, but only a single gene was found in *Silene*. After examining genomic and transcriptomic resources from *Nicotiana tabacum* and *Oryza sativa*, we found that these species also have two or more mt-DNA polymerases, but that they each represent independent duplications and are not orthologous to *POL1A* and *POL1B* of *Arabidopsis* ([supplementary fig. S8, Supplementary Material](#) online). Therefore, the difference in mt-DNA polymerase copy number appears to be the result of independent duplications in other lineages rather than a loss in *Silene*. Similarly, two endonucleases (*NTH1*, plastid-targeted;

Table 1

Summary Statistics from Population Genetics Tests Using 19 Populations of *Silene conica* to Estimate Polymorphism Counts (P_n and P_s) Relative to Divergence Counts (D_n and D_s) for Concatenated Sets of Genes, with *Silene latifolia* Used as an Outgroup

| Gene Set | Num Genes | Alignment Length (nt) | P_n | P_s | D_n | D_s | NI | NI _{TC} | DoS | Tajima's D | Fu and Li's D | Fu and Li's F | Fay and Wu's H | DHEW P |
|-----------------|-----------|-----------------------|-------|-------|-------|-------|---------|------------------|-------|------------|---------------|---------------|----------------|--------|
| OXPHOS I | 27 | 12096 | 29 | 83 | 164 | 298 | 0.63 | 0.68 | 0.10 | -0.44 | -0.40 | -0.50 | -0.57 | 1.00 |
| OXPHOS II | 7 | 4701 | 17 | 49 | 64 | 101 | 0.55 | 0.46 | 0.13 | 0.20 | 0.43 | 0.40 | 0.51 | 1.00 |
| OXPHOS III | 5 | 3264 | 4 | 6 | 57 | 83 | 0.97 | 1.12 | 0.01 | -0.19 | -1.15 | -1.13 | 0.56 | 0.32 |
| OXPHOS IV | 5 | 1524 | 10 | 18 | 35 | 40 | 0.63 | 0.56 | 0.11 | 0.05 | 0.03 | 0.04 | 0.29 | 0.20 |
| OXPHOS V | 6 | 3123 | 5 | 4 | 129 | 91 | 0.88 | 1.34 | 0.03 | -0.72 | -0.25 | -0.40 | -1.10 | 0.09 |
| OXPHOS I, III-V | 43 | 20007 | 48 | 111 | 385 | 512 | 0.58** | 0.72 | 0.13 | -0.35 | -0.39 | -0.47 | -0.39 | 1.00 |
| Mt. ribos. | 32 | 16617 | 69 | 102 | 873 | 329 | 0.25*** | 0.26 | 0.32 | -0.60 | -0.75 | -0.85 | -0.94 | 1.00 |
| Cyt. ribos. | 45 | 17585 | 22 | 174 | 137 | 609 | 0.56* | 0.57 | 0.07 | -0.21 | -0.04 | -0.11 | -1.38 | 1.00 |
| Mt-RRR | 16 | 26865 | 89 | 135 | 626 | 477 | 0.50*** | 0.55 | 0.17 | -0.54 | -0.63 | -0.75 | -1.20 | 1.00 |
| Random 50 | 49 | 50163 | 187 | 311 | 628 | 1219 | 1.17 | 1.14 | -0.04 | -0.26 | -0.27 | -0.34 | -1.44 | 1.00 |

NOTE.—The neutrality index (NI and its unbiased estimator NI_{TC}) and direction of selection (DoS) are reported for McDonald-Kreitman tests, along with statistics from several tests that make use of the SFS.

* $P < 0.05$,

** $P < 0.01$,

*** $P < 0.001$.

NTH2, mt-targeted) in *Arabidopsis* appear to be *Arabidopsis*-specific duplicates, and only a single, mt-targeted *NTH* gene was found in *Silene* and analyzed here (supplementary fig. S8, Supplementary Material online).

Some mt-RRR genes were excluded from phylogenetic and population genetic analyses because of difficulties in their assemblies. Specifically, only fragmented assemblies were generated for *RDR* (RNA-dependent DNA polymerase) in many species, which prevented its use in our analyses. However, the detection of transcript fragments indicates that *RDR* is present in *Silene*, albeit expressed at low levels or suffering from assembly problems. Four organelle-targeted, single-stranded DNA binding proteins are found in *Arabidopsis*: *OSB1*, *OSB3*, and *OSB4* are mt-targeted, while *OSB2* is solely plastid-targeted. When investigating closely related *OSB* sequences from *Arabidopsis*, *Nicotiana*, *Oryza*, and *Silene*, we found three *OSB*-like groups of sequences: one corresponding to *Arabidopsis OSB1*, one corresponding to *Arabidopsis OSB2/3/4* (which likely represent *Arabidopsis*-specific duplications), and one corresponding to a plastid-targeted *OSB* that is present in many angiosperms but does not appear to have an ortholog in *Arabidopsis* and was not analyzed here. The *Silene OSB2/3/4*-like sequences are all predicted to be plastid-targeted and were not analyzed here. *Silene* appears to have undergone a lineage-specific duplication for *OSB1* (supplementary fig. S9, Supplementary Material online), leading to multiple distinct *OSB1*-like sequences, the assembly of which is further complicated by large tandem duplications of the PDF domain (Zaegel et al. 2006). Many of these sequences contained predicted mt-targeting peptides. These *OSB1* duplicates were found in both slow and fast *Silene*, but were not analyzed here because of paralogy and assembly concerns among the *OSB1*-like duplicates.

Overall, fast *Silene* appear to have a complete, functional contingent of mt-RRR genes and show no obvious deviations in gene content from slow *Silene*.

Discussion

Positive, Not Relaxed, Selection Shapes the Evolution of N-Mt Genes in Silene

Mitochondrial interactions have been theorized to have important implications for human health, diversification of eukaryotes, and major transitions in the evolution of life (Burton and Barreto 2012; Dowling 2014; Havird et al. 2015a; Hill 2015; Sloan et al. 2015). One of the underlying assumptions of these arguments is that genes encoded by the nuclear genome are under positive selection to compensate for slightly deleterious mitochondrial mutations, which may escape selection owing to the uniparental and haploid nature of mtDNA (Lynch 1996; Lynch and Blanchard 1998). While several independent studies support this scenario (Osada and Akashi 2012; Barreto and Burton 2013; Gong et al. 2014; Sloan et al. 2014; Beck et al. 2015; Havird et al. 2015b), two specific predictions can be made regarding sequence evolution under nuclear compensation: 1) nuclear genes should show molecular signatures of positive selection, especially in species with rapidly evolving mtDNA, and, 2) such signatures should be overrepresented in N-mt genes vs. those that encode proteins targeted to elsewhere in the cell. Here, we present three lines of evidence that positive selection is playing a significant role in shaping the evolution of N-mt genes in *Silene* with fast mtDNA evolution.

First, we confirmed previous patterns of elevated d_N/d_S in N-mt genes for fast *Silene* species (Fig. 2). This pattern could be caused by positive or relaxed selection. However, if relaxed selection was due to a population bottleneck in fast *Silene* that

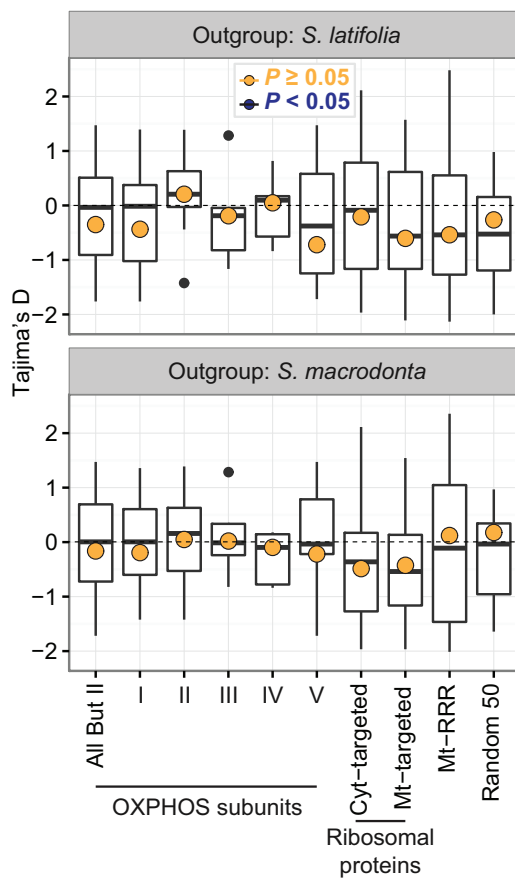


FIG. 4.—Results of Tajima’s D tests for neutrality for 19 populations of *Silene conica*, for genes that included using *S. latifolia* or the closer relative *S. macrodonta* as outgroups (as in Fig. 3). Note that outgroup choice does not matter for Tajima’s D, but that the two analyses used different sets of genes and are presented here for comparison with the MK results. Values above or below 0 (indicated by dashed lines) suggest balancing selection/population contraction or a recent selective sweep/population expansion, respectively. Colors and boxplots as in Figure 3, except larger points represent values for concatenated sequences within a gene set.

reduced effective population size, then d_N/d_S should be elevated across the entire nuclear genome. Because this pattern was not observed in a set of randomly chosen nuclear genes that do not interact with mitochondria or plastids (Fig. 2), we can exclude the possibility that a recent population bottleneck or other genome-wide phenomenon in fast *Silene* is responsible for elevated d_N/d_S . Although increased d_N/d_S in N-mt genes could be caused by a change in selection on overall mt function, normal d_N/d_S values in SDH and mt CLP genes make this unlikely (Havird et al. 2015b; Rockenbach et al. 2016).

Second, we used population genetic data to show an abundance of nonsynonymous divergence compared to polymorphism in *S. conica* N-mt genes (Fig. 3), which is a hallmark of positive selection (McDonald and Kreitman 1991). This was

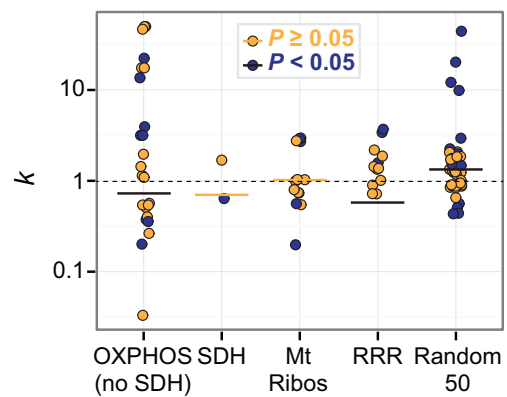


FIG. 5.—Results of RELAX tests for relaxed selection on *Silene* with fast-evolving mtDNA. All gene sets are for nuclear-encoded genes (N-mt genes) that code for proteins targeted to the mitochondria, except a set of 50 randomly chosen nuclear genes, which act as a control for mitonuclear interactions. SDH subunits in OXPHOS complex II act as another type of control, as they are targeted to the mitochondria, but do not interact directly with any mt-encoded subunits. k values above or below 1 (indicated by a dashed line) suggest intensified or relaxed selection, respectively. Points indicate individual loci, while horizontal lines indicate values for concatenated sequences in each gene set. Colors of lines/points indicate significance (as assessed by likelihood ratio tests).

not observed for the set of 50 random nuclear genes, again excluding any demographic cause of this pattern. Moreover, these effects remained significant in many of the N-mt gene sets examined whether using *S. latifolia* or *S. macrodonta* as outgroups. This implies positive selection on N-mt genes took place across at least two time scales (because *S. conica* is ~5.7 Myr divergent from *S. latifolia* and ~1.8 Myr divergent from *S. macrodonta*; (Rautenberg et al. 2012), and provides further evidence against the possibility of demographic effects causing this pattern. Moreover, because a much lower number of homologous sequences were identified in *S. macrodonta* and there is a lower amount of sequence divergence between *S. conica* and *S. macrodonta*, *S. macrodonta* serves as a more conservative comparison.

Third, we explicitly tested for relaxed selection by implementing a phylogeny-based analysis of the distribution of d_N/d_S across sites and branches using the program RELAX (Wertheim et al. 2014). There was no more evidence for relaxed selection in N-mt genes in fast *Silene* than in the set of 50 random nuclear genes. However, our interpretation of the RELAX data is complicated by the fact that we found a potential artifact when analyzing the concatenated mt-RRR dataset (which exhibited the opposite direction of effect than each gene analyzed individually). It is not clear what caused this anomalous result. One potential explanation is that large amounts of missing data in the concatenated mt-RRR alignment (i.e., no gene was present in all 14 species) led to an erroneously large d_N/d_S estimate for sites under positive

selection in the reference taxa. Another possibility is that incomplete lineage sorting or introgression of mt-RRR genes may have also influenced results of concatenation analyses where all genes were forced into a single phylogeny. A relatively low number of species included in our analyses (14, whereas the median number used on the RELAX webserver is 19) may have also reduced the power to detect changes in selection using RELAX.

Overall, N-mt genes showed evidence of positive selection. Although this pattern was largely consistent with our predictions under the nuclear compensation hypothesis, there were some unexpected results. For example, SDH subunits in complex II had similar NI values to other complexes and OXPHOS genes (Fig. 3), even though SDH subunits are completely nuclear-encoded and do not show elevated d_N/d_S in fast *Silene* species (Fig. 2). The lack of a difference in d_N/d_S between fast and slow *Silene* suggests that mt mutation rate may not have driven this apparent selection on SDH genes. Nevertheless, it is possible that even though complex II does not contain any mt-encoded subunits, it may experience indirect effects associated with having to interact with mitochondrial import and assembly proteins, which may be under positive selection themselves.

A similar result was observed when comparing mt- and cytosolic-targeted ribosomal proteins, as MK tests produced evidence of positive selection for both. While there may have been a history of positive selection on cytosolic-targeted ribosomal proteins, it is important to note that positive selection appears to be much stronger on mt-targeted ribosomal proteins. Importantly, across eukaryotes mt-targeted ribosomal proteins are likely less functionally constrained than their cytosolic-targeted counterparts because of the smaller number of transcripts that mt ribosomes must translate (Sloan et al. 2014; Adrion et al. 2015; Pett and Lavrov 2015). Our analyses are not at odds with this conclusion, as we simply found that there is no evidence that mt-targeted ribosomal proteins have experienced a *further* relaxation of selection in fast *Silene* compared to slow *Silene*. Instead, the accelerated evolution of N-mt genes in species such as *S. conica* appears to be associated with signatures of positive, not relaxed, selection.

Although MK tests indicated a history of positive selection in N-mt genes, Tajima's D and related tests suggest N-mt genes may not *currently* be under positive selection, because the SFS in *S. conica* populations did not significantly deviate from neutral expectations. In *Silene* and other angiosperms with fast-evolving mtDNA, it is unclear if elevated mtDNA mutation rates remain uniformly high in species such as *S. conica*, or whether elevated mitochondrial mutation rates occur in episodic bursts and then subside. Specifically, mt mutation rates appear to have undergone accelerations and subsequent decelerations in the Geraniaceae (Parkinson et al. 2005). Some data in *S. noctiflora* (e.g., long internal branches vs. short terminal branches; low levels of mt polymorphisms) support a scenario in which ancient accelerations in mt

mutation rates have since reverted (Sloan et al. 2009; Sloan et al. 2012b; Wu et al. 2015). If we assume N-mt genes are responding to mitochondrial mutations, then our results from the MK and Tajima's D-like tests support this scenario, because we recover evidence for older episodes of positive selection on N-mt genes (between species), but not evidence for ongoing positive selection (within species). Moreover, it is unclear if the species analyzed here represent a single episode of mtDNA acceleration or two independent events (in section *Conoimorpha* and *S. noctiflora/S. undulata*; Fig. 1) (Sloan et al. 2009). Because the D-like statistics were generally low across most gene sets examined, including the random 50 nuclear genes, there may have been a population expansion after a recent bottleneck in *S. conica*. This is supported by phylogeographic studies in closely related *Silene* species, which indicate a history of post-glacial population expansion into Northern Europe from Mediterranean refugia (Taylor and Keller 2007).

Why are N-mt genes under positive selection in *Silene* with fast-evolving mtDNA? This specific result supports a more general trend across eukaryotes, showing that lineages with relatively high rates of mt and plastid genome evolution also tend to have elevated rates of evolution in nuclear genes that are targeted to those organelles (Sloan et al. 2014; Havird et al. 2015b; Zhang et al. 2015; Havird and Sloan 2016; Weng et al. 2016). Importantly, nuclear compensation and coevolution with organelle genomes need not necessarily be invoked to explain these trends. For example, N-mt genes and mtDNA may be responding simultaneously to the same external selective pressure(s) or N-mt changes may have predated mtDNA changes. More generally, altered mtDNA evolution and genome structure may create a new adaptive landscape for many N-mt genes encoding proteins involved in mitochondria-related pathways, not just those that must directly interact with mt-encoded products. However, changes in N-mt OXPHOS proteins tend to occur at protein residues that physically interact with mutated mtDNA residues in *Silene* and primates (Havird et al. 2015b; Osada and Akashi 2012). Moreover, altered mtDNA residues were equally likely to occur at or away from N-mt residues, and mt-encoded OXPHOS genes do not show elevated d_N/d_S in *Silene* with fast-evolving mtDNA (Havird et al. 2015b). This suggests the mtDNA itself is not under positive selection in fast *Silene* and provides some support that nuclear genes are responding to mitochondrial changes, and not vice versa. Even if original mtDNA changes are beneficial in some cases, these results suggest coevolution between the genomes if not outright nuclear compensation. Regardless of the exact evolutionary interplay between mt and N-mt genes in *Silene*, the pervasive signals of positive selection in N-mt genes from fast *Silene* largely refute the possibility of relaxed functional constraints on mitochondrial function in these lineages.

The seeds collected here represent multiple species and populations of *Silene*, which will allow for inter- and intra-species crossing designs. Such hybrids often expose

mitonuclear incompatibilities and provide some of the strongest evidence for the nuclear compensation hypothesis (Burton and Barreto 2012; Burton et al. 2013; Gershoni et al. 2009; Hill 2015). Coupled with previous studies and the population genetic work presented here, creating *Silene* hybrids would further test the nuclear compensation hypothesis and the consequences of elevated mitochondrial mutation rate.

RRR Genes and Elevated Mutation Rates in Silene

In most bilaterian animals, mutation rates in the mtDNA outstrip those in the nucleus (Lavrov and Pett 2016). In most angiosperms, the opposite is true (Wolfe et al. 1987), possibly owing to RRR mechanisms that are found in plant mtDNA but absent in animals (Christensen 2014). However, in several independent angiosperm lineages, mutation rates in mitochondrial genomes have undergone an acceleration, shifting the mitonuclear mutation balance to a more animal-like pattern (Mower et al. 2007). The cause of these rapid increases is still unknown, but alterations in mt-RRR genes would be an obvious candidate mechanism to explain changes in mtDNA mutation rate, as mutated mt-RRR genes can lead to organelle genome instability and dysfunction (Marechal and Brisson 2010). Zhang et al. (2016) recently examined plastid RRR genes in species within the Geraniaceae, which have elevated plastid substitution rates and genomic instability (Guisinger et al. 2008, 2011). They found that rates of plastid evolution (d_N) across their phylogeny significantly correlated with rates of evolution in three plastid-targeted RRR genes. Similar to our findings in *Silene*, these results support the hypothesis that changes in organelle-targeted RRR genes may be related to altered evolution in organelle genomes (although the strongest candidate gene identified in that study, *UVRB/C*, may not actually play a role in plastid RRR, as it has recently been reclassified as *CLPF*, a component of a plastid protease complex, and it is not found in the nucleoid fraction from *Arabidopsis* plastids; Huang et al. 2013).

We predicted mt-RRR genes in *Silene* with fast-evolving mtDNA may be under relaxed selection or even lost/pseudogenized, leading to increased mtDNA mutation rates. While mt-RRR genes did show increased d_N/d_S , the results presented lead to the surprising conclusion that this is most likely due to positive, not relaxed, selection. Moreover, there was no evidence of pseudogenization or loss of mt-RRR genes in *Silene* with fast-evolving mtDNA, ruling out one obvious mechanism for runaway mitochondrial mutation rates. These results largely contradict the hypothesis that deleterious mutations in mt-RRR genes have escaped selection, compromised RRR function, and caused elevated mitochondrial mutation rates in *Silene*. Of course, caution is in order because it is likely that many of the genes involved in plant mt-RRR machinery are yet to be identified. Furthermore, even in the known genes we investigated, single changes with large effects on mt-RRR

function could easily have been overlooked in our statistical analyses.

The finding that elevated rates of evolution in mt-RRR genes are likely due to positive selection was surprising and opens the door to new deliberation on the role of RRR in angiosperms with elevated mt mutation rates. One possibility is that positive selection on mt-RRR genes is *caused by*, rather than the *cause of*, elevated mtDNA mutation rates. Mt-RRR genes may be under selection to coevolve with mtDNA mutations as in other N-mt genes in order to maintain a functional “match” between the mtDNA and RRR genes. Interpopulation hybrids of the copepod *Tigriopus californicus* show altered mtDNA transcription, suggesting mtDNA regulatory sites may not be recognized by nuclear-encoded mt RNA polymerases from other lineages (Ellison and Burton 2008, 2010). A similar process could occur with mtDNA and mt-RRR genes. Another possibility is that, when mitochondrial mutation rates increased, there was selection on RRR genes to offset what would otherwise lead to (even more) genome instability. This scenario would also be consistent with ancient deleterious changes at specific sites in mt-RRR genes that caused accelerated mt mutation rates and then underwent reversions or compensatory changes to restore functionality. This would not cause selection on specific, mtDNA-interacting sites in the mt-RRR genes, but would rather lead to a change in overall RRR function. In this scenario, the mtDNA sequence changes themselves are not the selective force. Instead, the cause would be overall changes in mtDNA stability, which could have been caused by forces such as an increase in reactive oxygen species (ROS) production, increased UV exposure, or a change in physiology.

It is striking that, in many respects, many fast and slow *Silene* species maintain similar geographic distributions and ecology. Interestingly, however, members of the fast clades tend to be hermaphroditic annuals that readily self-fertilize, whereas much of the genus is characterized by other life histories and breeding systems (Desfeux et al. 1996; Davis and Delph 2005). For example, *S. latifolia* and *S. vulgaris*, two of the slower species investigated here, are outcrossing perennials with dioecious and gynodioecious breeding systems, respectively. Therefore, it is conceivable that shorter generation times or increased levels of linkage disequilibrium and hitchhiking associated with selfing in fast *Silene* species have played a role in acceleration of mt mutation rates. However, we consider this a highly speculative argument, especially given that there are many examples of selfing annuals with slowly evolving mtDNA, including other species within *Silene* (Desfeux et al. 1996; Sloan et al. 2009).

Another possibility is that the changes in mt-RRR genes are, in fact, responsible for the elevated mitochondrial mutation rates of *Silene* but that they are adaptive rather than deleterious. This could be because abnormal mutation rates and structural evolution are themselves adaptive or because mt-RRR genes play other roles that are under positive selection, and the altered mtDNA phenotype is a pleiotropic effect. Of

course, interpretations of population genetic and phylogenetic evidence for positive selection should be made with caution. MK tests, although generally robust, can suffer from false positives (Nielsen 2001, 2005), and the observed effect sizes here were generally small for RRR genes. Examining the spatial context and timing of changes in mtDNA and RRR genes with a similar approach as employed by Osada and Akashi (2012) may help distinguish between the possibilities suggested here. Moreover, examining expression of mt-RRR genes could reveal important regulatory differences between fast and slow *Silene* species. Therefore, the ultimate cause(s) of the rapid accelerations in mitochondrial mutation rates in independent angiosperm lineages remains an open area for exploration.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Arne Strid for providing distribution maps of *Silene* in Greece (now available in his *Atlas of the Aegean Flora*), which greatly aided our collection efforts. We also thank Greece's Ministry of the Environment for permission to collect seeds from wild *Silene* populations in Greece. We thank Andrea Berardi, Rolland Douzet, Peter Fields, Michael Hood, Andreas König, Arne Saatkamp, the Kew Millennium Seed Bank, the Ornamental Plant Germplasm Center, and the Ville de Nantes Jardin Botanique for collecting/providing seeds. Bengt Oxelman suggested investigating *S. undulata* as a close relative of *S. noctiflora*. Joel Wertheim provided helpful comments on RELAX. Cody Kalous and Jessica Hurley performed RNA extractions for *S. ammophila* and *S. conoidea*. We also thank members of the labs of Rachel Mueller and DBS, Dennis Lavrov, and two anonymous reviewers for valuable comments on an earlier version of this manuscript. This work was supported by a Division of Molecular and Cellular Biosciences grant at the National Science Foundation (MCB 1412260) and a National Institutes of Health Ruth L. Kirschstein National Research Service Award (F32GM116361).

Literature Cited

- Adrión JR, White PS, Montooth KL. 2015. The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. *Mol Biol Evol.* 33:152–161.
- Allen JF. 2015. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocalization for redox regulation of gene expression. *Proc Natl Acad Sci U S A.* 112:10231–10238.
- Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Mol Biol Evol.* 30:310–314.
- Beck EA, Thompson AC, Sharbrough J, Brud E, Llopart A. 2015. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. *Evolution* 69:1973–1986.
- Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial-DNA. *Proc Natl Acad Sci U S A.* 76:1967–1971.
- Burton RS, Barreto FS. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol Ecol* 21:4942–4957.
- Burton RS, Pereira RJ, Barreto FS. 2013. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Syst* 44:281–302.
- Camacho C, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc Natl Acad Sci U S A.* 101:17741–17746.
- Christensen AC. 2014. Genes and junk in plant mitochondria-repair mechanisms and selection. *Genome Biol Evol.* 6:1448–1453.
- Davis SL, Delph LF. 2005. Prior selfing and gynomonoeicy in *Silene noctiflora* L. (Caryophyllaceae): Opportunities for enhanced outcrossing and reproductive assurance. *Int J Plant Sci.* 166:475–480.
- Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* 289:2342–2344.
- Desfeux C, Maurice S, Henry JP, Lejeune B, Gouyon PH. 1996. Evolution of reproductive systems in the genus *Silene*. *Proc Roy Soc Biol Sci.* 263:409–414.
- Dohm JC, et al. 2013. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505:546–549.
- Dowling DK. 2014. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochim Biophys Acta.* 1840:1393–1403.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Ellison CK, Burton RS. 2010. Cytonuclear conflict in interpopulation hybrids: the role of RNA polymerase in mtDNA transcription and replication. *J Evol Biol.* 23:528–538.
- Ellison CK, Burton RS. 2008. Genotype-dependent variation of mitochondrial transcriptional profiles in interpopulation hybrids. *Proc Natl Acad Sci U S A.* 105:15831–15836.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2:953–971.
- Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Gagnaire PA, Normandeau E, Bernatchez L. 2012. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Mol Biol Evol.* 29:2909–2919.
- Gershoni M, Templeton AR, Mishmar D. 2009. Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays* 31:642–650.
- Gong L, Olson M, Wendel JF. 2014. Cytonuclear evolution of rubisco in four allopolyploid lineages. *Mol Biol Evol.* 31:2624–2636.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol.* 28:583–600.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. 2008. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc Natl Acad Sci U S A.* 105:18424–18429.
- Havird JC, Hall MD, Dowling DK. 2015a. The evolution of sex: A new hypothesis based on mitochondrial mutational. *Bioessays* 37:951–958.

- Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial vs. nuclear genomes. *Mol Biol Evol.* 33:3042–3053.
- Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015b. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution* 69:3069–3081.
- Hill GE. 2015. Mitonuclear Ecology. *Mol Biol Evol.* 32:1917–1927.
- Hjort K, Goldberg AV, Tsaousis AD, Hirt RP, Embley TM. 2010. Diversity and reductive evolution of mitochondria among microbial eukaryotes. *Philos Trans R Soc Lond B Biol Sci.* 365:713–727.
- Huang M, et al. 2013. Construction of plastid reference proteomes for maize and *Arabidopsis* and evaluation of their orthologous relationships; the concept of orthoproteomics. *J Proteome Res.* 12:491–504.
- Karnkowska A, et al. 2016. A eukaryote without a mitochondrial organelle. *Curr Biol.* 26:1274–1284.
- Lavrov DV, Pett W. 2016. Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in Nonbilaterian lineages. *Genome Biol Evol.* 8:2896–2913.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Lynch M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol Biol Evol.* 14:914–925.
- Lynch M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol Biol Evol.* 13:209–220.
- Lynch M, Blanchard JL. 1998. Deleterious mutation accumulation in organelle genomes. *Genetica* 102-103:29–39.
- Lynch M, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A.* 105:9272–9277.
- MacManes MD. 2014. On the optimal trimming of high-throughput mRNA sequence data. *Front Genet.* 5:13.
- Marechal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytol* 186:299–317.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol.* 7:135.
- Nabholz B, Ellegren H, Wolf JB. 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Mol Biol Evol.* 30:272–284.
- Neiman M, Taylor DR. 2009. The causes of mutation accumulation in mitochondrial genomes. *Proc Biol Sci.* 276:1201–1209.
- Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet.* 39:197–218.
- Nielsen R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* 86:641–647.
- Osada N, Akashi H. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol Biol Evol.* 29:337–346.
- Parkinson CL, et al. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evol Biol.* 5:73.
- Parmakelis A, Kotsakiozi P, Rand D. 2013. Animal mitochondria, positive selection and cyto-nuclear coevolution: insights from pulmonates. *Plos One* 8:e61970.
- Pett W, Lavrov D. 2015. Mito-nuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biol Evol.* 7:2089–2101.
- Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol.* 31:1929–1936.
- Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679.
- Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE. 2013. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol Biol Evol.* 30:347–355.
- Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol Biol Evol.* 13:735–748.
- Rautenberg A, Sloan DB, Alden V, Oxelman B. 2012. Phylogenetic Relationships of *Silene multinervis* and *Silene* Section *Conoimorpha* (Caryophyllaceae). *Systematic Botany* 37:226–237.
- Rockenbach K, et al. 2016. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. *Genetics* 204:1507–1522.
- Sloan DB, et al. 2012a. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *Plos Biology* 10:e1001241.
- Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012b. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the Angiosperm genus *Silene*. *Genome Biol Evol.* 4:294–306.
- Sloan DB, Fields PD, Havird JC. 2015. Mitonuclear linkage disequilibrium in human populations. *Proc Biol Sci.* 282:20151704.
- Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. *BMC Evol Biol.* 9:260.
- 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.
- Stajich JE, et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* 12:1611–1618.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 57:758–771.
- Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Mol Biol Evol.* 28:63–70.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tamura K, Stecher G, Peterson D, Filipksi A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Taylor DR, Keller SR. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61:334–345.
- Wallace DC. 2005. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet.* 39:359–407.
- Warren JM, Simmons MP, Wu Z, Sloan DB. 2016. Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. *Genome Biol Evol.* 8:364–374.
- Watterson GA. 1978. The homozygosity test of neutrality. *Genetics* 88:405–417.
- Weng ML, Ruhlman TA, Jansen RK. 2016. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biol Evol.* 8:1824–1838.
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2014. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* 32:820–832.
- Willett CS, Burton RS. 2004. Evolution of interacting proteins in the mitochondrial electron transport system in a marine copepod. *Mol Biol Evol.* 21:443–453.

- 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.
- Wu ZQ, Cuthbert JM, Taylor DR, Sloan DB. 2015. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc Natl Acad Sci U S A*. 112:10185–10191.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Zaegel V, et al. 2006. The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18:3548–3563.
- Zeng K, Shi S, Wu CI. 2007. Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol*. 24:1898–1908.
- Zhang F, Broughton RE. 2013. Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes?. *Genome Biol Evol*. 5:1781–1791.
- Zhang J, Ruhlman TA, Sabir J, Blazier JC, Jansen RK. 2015. Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. *Plant Cell* 27:563–573.

Associate Editor: Dennis Lavrov