Positive Selection in Rapidly Evolving Plastid–Nuclear Enzyme Complexes

Kate Rockenbach,* Justin C. Havird,* J. Grey Monroe,† Deborah A. Triant,‡ Douglas R. Taylor,§ and Daniel B. Sloan*,1

*Department of Biology and †Department of Bioagricultural Sciences and Pest Management, Colorado State University, Fort Collins, Colorado 80523, ‡Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, and §Department of Biology, University of Virginia, Charlottesville, Virginia 22904

ABSTRACT Rates of sequence evolution in plastid genomes are generally low, but numerous angiosperm lineages exhibit accelerated evolutionary rates in similar subsets of plastid genes. These genes include *clpP1* and *accD*, which encode components of the caseinolytic protease (CLP) and acetyl-coA carboxylase (ACCase) complexes, respectively. Whether these extreme and repeated accelerations in rates of plastid genome evolution result from adaptive change in proteins (*i.e.*, positive selection) or simply a loss of functional constraint (*i.e.*, relaxed purifying selection) is a source of ongoing controversy. To address this, we have taken advantage of the multiple independent accelerations that have occurred within the genus *Silene* (Caryophyllaceae) by examining phylogenetic and population genetic variation in the nuclear genes that encode subunits of the CLP and ACCase complexes. We found that, in species with accelerated plastid genome evolution, the nuclear-encoded subunits in the CLP and ACCase complexes are also evolving rapidly, especially those involved in direct physical interactions with plastid-encoded proteins. A massive excess of nonsynonymous substitutions between species relative to levels of intraspecific polymorphism indicated a history of strong positive selection (particularly in CLP genes). Interestingly, however, some species are likely undergoing loss of the native (heteromeric) plastid ACCase and putative functional replacement by a duplicated cytosolic (homomeric) ACCase. Overall, the patterns of molecular evolution in these plastid–nuclear complexes are unusual for anciently conserved enzymes. They instead resemble cases of antagonistic coevolution between pathogens and host immune genes. We discuss a possible role of plastid–nuclear conflict as a novel cause of accelerated evolution.

KEYWORDS chloroplast; cytonuclear interactions; McDonald–Kreitman test; plastome

PLASTIDS carry reduced genomes that reflect an evolutionary history of extensive gene loss and transfer to the nucleus since their ancient endosymbiotic origin roughly 1 billion years ago (Timmis *et al.* 2004; Keeling 2010; Gray and Archibald 2012). Many of the proteins encoded by genes that have been transferred to the nuclear genome are trafficked back into the plastid (Gould *et al.* 2008), where they interact closely with proteins encoded by genes remaining in the plastid genome. These interacting proteins are key not only to photosynthesis, but also to transcription, translation, and critical nonphotosynthetic metabolic functions of the plastid. The interactions between these gene products create the opportunity for coevolution between plastid and nuclear genomes. Thus, studying the nuclear genes that contribute to plastid complexes is a valuable tool for understanding the processes underlying plastid genome evolution and cytonuclear coevolution.

Within angiosperms, most plastid genomes are highly conserved in sequence and structure (Jansen *et al.* 2007; Wicke *et al.* 2011), but multiple independent lineages have experienced accelerated rates of aa substitution in similar subsets of nonphotosynthetic genes (Jansen *et al.* 2007; Erixon and Oxelman 2008; Greiner *et al.* 2008b; Guisinger *et al.* 2008, 2010, 2011; Straub *et al.* 2011; Sloan *et al.* 2012a, 2014a; Barnard-Kubow *et al.* 2014; Weng *et al.* 2014; Dugas *et al.* 2015; Williams *et al.* 2015; Zhang *et al.* 2016). Several mechanisms have been hypothesized to explain these repeated accelerations including positive selection, reduced effective population size (N_e), altered DNA repair, changes in

doi: 10.1534/genetics.116.188268
Manuscript received February 18, 2016; accepted for publication October 4, 2016; published Early Online October 5, 2016.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.188268/-/DC1.
[1]Corresponding author: Colorado State University, 1878 Campus Delivery, Fort Collins, CO 80523. E-mail: dan.sloan@colostate.edu

gene expression, and pseudogenization following gene transfer to the nucleus (see above citations). Distinguishing among these hypotheses has proved challenging, and the ultimate cause or causes of the extreme differences in rates of molecular evolution among genes within plastid genomes remain unclear.

In many cases of extreme plastid genome evolution, accelerations have disproportionately affected nonsynonymous sites, resulting in elevated ratios of nonsynonymous to synonymous substitution rates (d_N/d_S) (*e.g.*, Erixon and Oxelman 2008; Guisinger *et al.* 2008; Barnard-Kubow *et al.* 2014; Sloan *et al.* 2014a), which indicates that changes in selection are likely involved. In addition, recent studies showed correlated increases in d_N/d_S between nuclear- and plastid-encoded subunits in ribosomal (Sloan *et al.* 2014b; Weng *et al.* 2016) and RNA polymerase complexes (Zhang *et al.* 2015), providing further evidence for changes in selection pressures. However, these studies could not confidently distinguish between two alternative explanations for increased d_N/d_S, positive selection and relaxed purifying selection, which can be difficult to disentangle based on sequence divergence data alone. Because these selection pressures can have very different effects on population genetic variation, analyses that combine data on intraspecific polymorphism and interspecific divergence (McDonald and Kreitman 1991) can detect positive selection even in cases where it is not readily identifiable based only on d_N/d_S (Rausher *et al.* 2008). However, most studies of accelerated plastid genome evolution and plastid–nuclear coevolution have not included the necessary intraspecific polymorphism data to perform these analyses.

In contrast to recent analyses of plastid genetic machinery (*i.e.*, ribosomal and RNA polymerase genes; Sloan *et al.* 2014b; Zhang *et al.* 2015; Weng *et al.* 2016), the potential for molecular coevolution involving nuclear-encoded subunits in other plastid complexes remains largely unexplored. Two such complexes are the caseinolytic protease (CLP), which is an ATP-dependent protease required for proper plastid function (Nishimura and van Wijk 2015), and the heteromeric acetyl-coA carboxylase (ACCase), which is involved in fatty acid biosynthesis (Sasaki and Nagano 2004; Salie and Thelen 2016). The CLP complex and ACCase each contain a single plastid-encoded subunit (ClpP1 and AccD, respectively) and multiple subunits of nuclear origin. In most angiosperms, the sequences of the *clpP1* and *accD* genes are generally conserved, but they are among the plastid-encoded genes that exhibit elevated rates of sequence evolution in multiple independent lineages. The *clpP1* gene, in particular, exhibits recent and dramatically increased rates of nonsynonymous substitutions and indels (*e.g.*, Erixon and Oxelman 2008; Sloan *et al.* 2014a).

In addition to phylogenetic and population genetic analyses, examining patterns of aa substitutions relative to protein structure can help distinguish between relaxed and positive selection. In the model angiosperm *Arabidopsis thaliana*, the CLP complex is made up of two stacked heptameric rings, comprising nine different types of paralogous and structurally related subunits that are derived from the single subunit found in the ancestral homotetradecameric form of this enzyme (Peltier *et al.* 2004; Yu and Houry 2007; Olinares *et al.* 2011). The P-ring is formed entirely of the nuclear-encoded subunits CLPP3, 4, 5, and 6 in a 1:2:3:1 stoichiometric ratio, and the R-ring contains the plastid-encoded subunit ClpP1 and the nuclear-encoded subunits CLPR1, 2, 3, and 4 in a 3:1:1:1:1 ratio. The CLPP subunits all contain a conserved catalytic Ser-His-Asp triad, which is lacking from the CLPR subunits (Peltier *et al.* 2004), meaning that ClpP1 is the only catalytic subunit within the R-ring. Other nuclear-encoded subunits such as CLPC, CLPD, CLPF, CLPS, CLPT1, and CLPT2 are physically associated with the core CLP complex and act as adapters, chaperones, and accessory proteins, helping to regulate the proteolytic activity of CLP (Peltier *et al.* 2004; Nishimura and van Wijk 2015; Nishimura *et al.* 2015).

Most flowering plants contain two different types of ACCase enzymes: a eukaryotic-like homomeric multidomain ACCase in the cytosol and a bacterial-like heteromeric ACCase within the plastids. The heteromeric ACCase consists of proteins encoded by four different genes (ACCA, B, C, and D; Sasaki and Nagano 2004; Salie and Thelen 2016). ACCC is a biotin carboxylase. In an ATP-dependent reaction, it carboxylates a biotin molecule attached to ACCB (a biotin carboxyl carrier protein), with bicarbonate serving as the donor of the carboxyl group (White *et al.* 2005). The nuclear-encoded ACCA and plastid-encoded AccD closely interact with each other and represent the α and β carboxyltransferase subunits, respectively. Each of these subunits first homodimerizes, and then they combine as a heterotetramer, forming the functional enzyme that transfers the carboxyl group from biotin to acetyl-CoA (Cronan and Waldrop 2002). Together, these enzymes convert acetyl-CoA to malonyl-CoA within the plastid, which is the first step in the fatty acid biosynthesis pathway (White *et al.* 2005). In some lineages, the homomeric ACCase has undergone a duplication, and one copy is targeted to the plastid while the other remains in the cytosol (Konishi and Sasaki 1994; Schulte *et al.* 1997; Babiychuk *et al.* 2011; Parker *et al.* 2014).

The angiosperm tribe *Sileneae* (Caryophyllaceae) has emerged as model for studying organelle genomes under divergent rates of sequence evolution (Mower *et al.* 2007; Erixon and Oxelman 2008; Sloan *et al.* 2012a,b, 2014a). This group contains multiple lineages with phylogenetically independent accelerations in rates of plastid genome evolution (Sloan *et al.* 2012a, 2014a). In contrast, closely related *Sileneae* lineages have largely maintained low ancestral rates of evolution. This rate variation among closely related species presents a powerful contrast to analyze the evolutionary mechanisms responsible for accelerated plastid genome evolution and test for correlated changes in nuclear-encoded counterparts. Here, we use transcriptome sequencing data coupled with structural information to identify variation in nuclear genes both within and among *Sileneae* species with highly divergent rates of plastid genome evolution.

Specifically, we asked if there is evidence of selection on the sequences of nuclear-encoded subunits of the CLP and ACCase complexes in *Silene* species whose plastid-encoded counterparts have experienced recent accelerations in rates of evolution.

Materials and Methods

Taxon sampling, mRNA-seq, and transcriptome assembly

Silene conica, *S. noctiflora*, and *S. paradoxa* were all previously identified as having highly accelerated rates of nonsynonymous substitutions in a subset of plastid genes, with the most dramatic effects observed in *clpP1* (Sloan *et al.* 2012a, 2014a). The *accD* gene also exhibited increased nonsynonymous substitution rates (albeit much less pronounced), as well as the accumulation of large indels in these species. In contrast, there was little or no evidence of accelerated sequence evolution in photosynthetic genes in these species. *Silene latifolia*, *S. vulgaris*, and *Agrostemma githago* were chosen as representatives of closely related lineages that have maintained low rates of evolution throughout their entire plastid genomes (Sloan *et al.* 2012a, 2014a). Transcriptomes for these six species (*S. conica*, *S. latifolia*, *S. noctiflora*, *S. paradoxa*, *S. vulgaris*, and *Ag. githago*) were taken from previously described data sets (Sloan *et al.* 2014b) that were each generated from a single individual and assembled with Trinity r20120608 (Grabherr *et al.* 2011). These data sets were used for all phylogenetic analyses and correspond to NCBI Sequence Read Archive (SRA) accessions SRX353031, SRX353047, SRX353048, SRX353049, SRX353050, and SRX352988. For two genes (*CLPC* and *CLPP6*), some sequences were extracted from separate SOAPdenovo-Trans v1.02 (Xie *et al.* 2014) assemblies of the same reads because the Trinity assemblies were fragmented or complex.

Seeds from 19 geographically dispersed *S. conica* collections, including ABR which was used for the original *S. conica* transcriptome referenced above and one collection of the close relative *S. macrodonta* (Supplemental Material, Table S1), were germinated on soil in either July or August 2014 (Fafard 2SV mix supplemented with vermiculite and perlite) and grown under a 16/8-hr light/dark cycle with regular watering and fertilizer treatments in greenhouse facilities at Colorado State University. Plants were grown for 7–9 weeks, and total RNA was extracted from 2 to 3 leaves of a single individual from each collection using an RNeasy Plant Mini Kit (QIAGEN, Valencia, CA). Rosette leaves were used for all individuals with the exception of the ARZ and PDA samples, for which cauline leaves were used. The resulting RNA was sent to the Yale Center for Genome Analysis for Illumina mRNA-seq library preparation and sequencing. For all but two samples, polyA selection was used during library construction, while for the ABR sample of *S. conica* and the single *S. macrodonta* sample, mRNA selection was performed using a Ribo-Zero Plant Leaf rRNA Removal Kit (Illumina) in an effort to capture more organellar transcripts (as part of an unrelated project). Resulting strand-specific Illumina libraries were sequenced on two lanes of an Illumina HiSeq 2500 to generate paired-end 151-bp (2 × 151) reads. Raw (*i.e.*, nonnormalized and untrimmed) reads were then assembled using Trinity r20140717 with default parameters (note that the strand-specificity of the reads was not taken into account during assembly). Transcriptome assembly statistics and numbers of reads were similar among the 20 samples, except for an ~50% reduction in the average and total length of assembled transcripts for samples where Ribo-Zero was used (Table S2).

Extraction and alignment of orthologous sequences from Sileneae species

The focus of our study was the nuclear-encoded components of the plastid CLP and ACCase complexes (Table S3). In addition, sets of gene sequences were obtained from photosystem I (PSI) and the mitochondrial-targeted CLP protease (mtCLP) to serve as a basis for comparison. PSI was selected because it contains subunits from both the nuclear and plastid genomes but, unlike CLP and ACCase, the plastid-encoded subunits have been highly conserved even in *Silene* species with accelerated rates of evolution in other plastid genes (Sloan *et al.* 2012a, 2014a). The mtCLP complex was chosen because it is homologous to the plastid CLP, but consists entirely of nuclear-encoded subunits and is targeted to a different cellular compartment (the mitochondria). The mtCLP complex has a homotetradecamer core consisting entirely of CLPP2 subunits that interacts with the chaperones CLPX1, 2, and 3 (van Wijk 2015). Additionally, 50 genes with a minimum coding sequence length of 600 bp were selected at random from a published list of single-copy nuclear genes in angiosperms to test for global increases in evolutionary rates within the nuclear genome (Table S4; Duarte *et al.* 2010). Genes that were annotated as being targeted to the mitochondria or plastids were excluded from this random set. The *A. thaliana* sequences for selected genes were obtained through the TAIR database (https://www.arabidopsis.org/) with accession numbers from the literature (Table S3 and Table S4; Peltier *et al.* 2004; Duarte *et al.* 2010; Olinares *et al.* 2011; van Wijk 2015).

BLAST+ 2.2.31 (Camacho *et al.* 2009) was used to run tblastn searches (default settings) with the selected *A. thaliana* aa sequences as queries against each of the assembled *Sileneae* transcriptomes. The top hit in each transcriptome was retrieved with a custom Perl script using BioPerl modules (Stajich *et al.* 2002). In cases where there were multiple paralogous genes, manual curation aided by exploratory tree-building was performed to identify orthologs. Genes that were absent from the transcriptomes or for which orthologs could not be confidently identified were excluded from further analysis. For the random set, genes were excluded and replaced with another randomly selected gene if one or more species lacked orthologous sequence or had a partially assembled transcript that was less than two-thirds the length of the

coding sequence. Extracted sequences were aligned by nucleotide using the MUSCLE algorithm embedded within MEGA v6.0 (Tamura *et al.* 2013). The longest ORF in the *Ag. githago* sequence was predicted with the NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and other sequences were trimmed accordingly.

Sequences were then realigned by codon using MUSCLE. TargetP v1.1 (Emanuelsson *et al.* 2000) was used with translated ORFs from *Ag. githago* to predict the length of the N-terminal signal peptide (for plastid- and mitochondrial-targeted proteins), which was then removed from all sequences. In a few cases, TargetP was unable to predict a targeting peptide based on *Ag. githago* sequence, so *S. latifolia*, *S. vulgaris*, or *A. thaliana* was used instead to identify and remove signal peptides. Concatenated sequences for sets of genes in each complex were generated from final alignments with a custom Perl script using BioPerl modules.

Phylogenetic analysis of rates of sequence evolution

For each gene individually and for the concatenated sets of nuclear-encoded genes of each complex, we conducted multiple analyses of rates of synonymous and nonsynonymous substitutions by using the codeml program within PAML version 4.8 (Yang 2007). An F1 × 4 codon frequency model was applied in each analysis, and a constrained tree topology was used with the species in *Silene* subgenus *Behenantha* (*S. conica*, *S. latifolia*, *S. noctiflora*, and *S. vulgaris*) collapsed as a polytomy. First, we implemented a free branch model (model = 1 in PAML) to estimate d_N/d_S for each branch independently. Branches that were identified as having d_N/d_S values > 1 were then tested for significance using a LRT that compared the free branch model to a model that constrained d_N/d_S for the individual branch in question to a value of one. Finally, we classified species/branches into two groups, "fast" and "slow," based on known rates of plastid genome evolution (Sloan *et al.* 2012a, 2014a) and estimated separate d_N/d_S values for each group (model = 2 in PAML). The terminal branches for *S. conica*, *S. noctiflora*, and *S. paradoxa* were assigned to the fast group, while those for *S. latifolia*, *S. vulgaris*, and *Ag. githago* were assigned to the slow group. The internal branch connecting the common ancestor of *Silene* to the base of *Silene* subgenus *Behenantha* was also included in the slow group. As above, we used an LRT to test for significance in cases in which individual genes or concatenated sequences for entire complexes had an estimated d_N/d_S value > 1 for the fast group (no such cases where identified for the slow group). For the constrained model in these LRT comparisons, the d_N/d_S value for the fast group was set to one.

McDonald–Kreitman (MK) tests

MK tests (McDonald and Kreitman 1991) were performed using sequences from the *S. conica* population genetic data set. Sequences were extracted, aligned, and trimmed following the same methodology described above for the phylogenetic analysis. The tests were implemented with the web server described by Egea *et al.* (2008). For each gene, the neutrality index (NI) was calculated by dividing the ratio of nonsynonymous to synonymous polymorphisms within *S. conica* (P_n/P_s) by the ratio of nonsynonymous to synonymous divergence from a closely related outgroup (see below) species (D_n/D_s) (Rand and Kann 1996). NI values < 1 are indicative of positive selection, with statistical significance assessed by a standard contingency-table χ^2 analysis. We also calculated the direction of selection (DoS) for each gene (Stoletzki and Eyre-Walker 2011). Positive DoS values are indicative of positive selection and an excess of nonsynonymous substitutions. We looked for evidence of selection in sets of related genes by summing polymorphism and divergence counts for genes belonging to the plastid CLP, ACCase, PSI, or mtCLP complexes, as well as for the set of random genes. Because summing across contingency tables can introduce statistical bias, we also calculated NI_{TG} for each combined set of related genes, which is an unbiased estimator of NI (Stoletzki and Eyre-Walker 2011). Two separate sets of analyses were carried out, using either *S. latifolia* or *S. macrodonta* as the outgroup. Extracted sequences from the *S. conica* and *S. macrodonta* transcriptome assemblies that could not be confidently identified as orthologous were removed from the analysis. Specifically, *CLPR2*, *CLPX1*, and one randomly selected gene (*AXS2*: AT1G08200) showed evidence of recent duplications in the *S. conica* lineage, leading to apparent chimeric assembly artifacts. Therefore, these genes were not used for MK tests. In addition, all three *CLPX* genes and 32 of the randomly selected nuclear genes had low coverage and fragmented assemblies in the *S. macrodonta* data set, so MK tests for these genes were only performed with *S. latifolia* as an outgroup. The low coverage for many nuclear genes in the *S. macrodonta* assembly was likely related to the use of Ribo-Zero in the construction of that library (Table S2).

Analysis of protein structure and position of substitutions

To gain insight into the functional consequences of aa changes observed in fast-evolving *Silene* species, we mapped substitutions onto plastid CLP and ACCase protein structures. Ancestral *Silene* sequences were inferred using codeml in PAML with the guide tree containing the five *Silene* species encoded as a polytomy, with *Ag. githago* and *Escherichia coli* as outgroups. Partial sequences were excluded when inferring ancestral sequences. For each CLPP and CLPR subunit (including the plastid-encoded ClpP1), changes that were inferred to have occurred in *S. conica*, *S. paradoxa*, or *S. noctiflora* from the ancestral *Silene* sequence were mapped onto the structure of an individual *E. coli* ClpP subunit (PDB accession 1YG6; Bewley *et al.* 2006; Yu and Houry 2007). Likewise, changes in ACCase subunits were also mapped onto solved *E. coli* structures (PDB accessions 4HR7 and 2F9Y: Bilder *et al.* 2006; Broussard *et al.* 2013). Template structures from *E. coli* were used because no plant CLP or ACCase structures have been solved. While it is likely that there have been structural changes within these complexes between bacteria

and plants, most of these subunits are anciently conserved and can be reliably aligned (average aa identity = 43.3%).

Data availability

Raw Illumina reads and assembled transcriptome sequences are available via the NCBI SRA and Transcriptome Shotgun Assembly (TSA) database, respectively. Accession numbers are provided in Table S2. Sequence alignments used in PAML, MK, and phylogenetic analyses are provided in File S1.

Results

CLP and ACCase gene content in the tribe Sileneae

We were able to recover most of the expected genes from the *Sileneae* transcriptomes, as the gene content in these species was largely similar to that of *Arabidopsis*. However, we did find that some genes had experienced recent duplications or losses. We identified orthologs of all eight of the nuclear-encoded *CLPP* and *CLPR* genes that are targeted to the plastid in *Arabidopsis* (Figure S1), and we found that *CLPP5* has been duplicated (with the resulting copies designated as *CLPP5a* and *CLPP5b*). The duplication appears to have occurred prior to the divergence between *Agrostemma* and *Silene*, but only one of these copies (*CLPP5b*) was recovered from the *Agrostemma* transcriptome (Figure S1). As described in the *Materials and Methods*, there may also have been more recent duplications of genes such as *CLPR2* in individual species. In addition to the subunits that make up the core proteolytic rings of the plastid *CLP* complex, we also identified orthologs of the associated chaperones, adapters, and accessory proteins that have been described in *Arabidopsis* (Table 1; Nishimura and van Wijk 2015). The newly discovered *CLPF* adapter was also identified in our data set but was not included in the present analysis (Nishimura *et al.* 2015). *Arabidopsis* contains three paralogous chaperone genes that contribute to the plastid CLP complex (*CLPC1*, *CLPC2*, and *CLPD*). We found evidence that multiple copies of *CLPC/D* also exist in the *Sileneae*, but that the assemblies of these long genes were often fragmented, and we were only able to successfully extract one set of orthologs, which we refer to as *CLPC*. In addition to these plastid-targeted subunits, we also found orthologs of the mtCLP genes that have been identified in *Arabidopsis* (*CLPP2*, *CLPX1*, *CLPX2*, and *CLPX3*; van Wijk 2015).

Sileneae genes were successfully identified from each of the three classes of nuclear-encoded subunits of the heteromeric plastid ACCase (*ACCA*, *ACCB*, and *ACCC*), including two divergent copies of *ACCB*. Two copies of this gene also exist in *Arabidopsis* (Fukuda *et al.* 2013), but it was not readily apparent from phylogenetic analysis if there is an orthologous relationship between *Sileneae* and *Arabidopsis* copies or if they are the product of independent duplication events (data not shown). Although all of the heteromeric ACCase genes were identified in this clade, we found evidence of recent gene loss in some of the *Silene* species, which is described in detail below (see "Gene Loss and Accelerated Evolution of Some Subunits in the ACCase Complex").

With respect to the homomeric ACCase that is typically targeted to the cytosol, our transcriptome data indicated that *Sileneae* species express two distinct copies of this gene (Figure S2 and Figure S3), and that one of the resulting proteins has an N-terminal extension that is strongly predicted to act as a plastid-targeting peptide (with a specificity > 0.95 based on TargetP analysis). Duplication of the homomeric ACCase and retargeting to the plastids has occurred repeatedly and independently during angiosperm evolution (Figure S3). The observed duplication in our data set preceded the divergence between *Agrostemma* and *Silene*, but it was independent from similar duplications in grasses (Konishi and Sasaki 1994) and the Brassicaceae (Schulte *et al.* 1997; Babiychuk *et al.* 2011; Parker *et al.* 2014).

Nuclear-encoded components of the plastid CLP complex show elevated rates of aa substitution in species with rapidly evolving plastid genomes

We found that nuclear-encoded CLP genes have dramatically elevated d_N/d_S values in *Silene* species with recent accelerations in the evolutionary rates of plastid-encoded *clpP1* (Table 1 and Table 2). When concatenated, all 13 nuclear-encoded *CLP* genes had a d_N/d_S value significantly > 1 for both *S. conica* and *S. noctiflora*, and nearly $= 1$ for *S. paradoxa* (Table 1). In contrast, concatenated *CLP* genes had d_N/d_S values between 0.05–0.16 for closely related species with typical rates of *clpP1* evolution (Table 1). The extreme variance in d_N/d_S estimates resulted from elevated nonsynonymous substitution rates, whereas synonymous substitution rates were very similar across species (Figure 1). Rate differences were most pronounced for CLPR subunits, which occupy the same structural ring as the plastid-encoded CLPP1 subunit (van Wijk 2015). All 12 d_N/d_S estimates for *CLPR* genes within the "fast" species were > 1, with eight found to be significantly > 1 (Table 1). The d_N/d_S estimates for the *CLPP* genes within these fast species were also highly elevated, but only eight of the 15 were > 1, and only one was significantly so (Table 1). The adaptor gene *CLPS* was a noticeable outlier, being generally conserved in species regardless of their rates of plastid genome evolution (Figure 2).

Gene loss and accelerated evolution of some subunits in the ACCase complex

The three species with high rates of plastid-encoded *accD* evolution exhibited varied patterns with respect to nuclear-encoded ACCase genes, including some cases of accelerated evolution and other examples of outright gene loss. Notably, none of the nuclear-encoded ACCase genes were identified in the assembled *S. noctiflora* transcriptome. We confirmed the loss of these genes by searching a draft assembly of the *S. noctiflora* nuclear genome (D. B. Sloan, unpublished data). The nuclear genome assembly contained only pseudogenized fragments of *ACCA*, and none of the other ACCase genes were detected. The *S. paradoxa* transcriptome also appeared to

Table 1 Summary of d_N/d_S estimates

	Gene	*Ag. githago*	*S. paradoxa*	*S. conica*	*S. noctiflora*	*S. latifolia*	*S. vulgaris*
Plastid CLP	*CLPP3*	0.09	0.88	**2.89**	**2.31**	0.52	0.00
	CLPP4	0.02	0.36	0.62	**1.24**	0.21	0.00
	CLPP5a	0.03	0.52	**1.55**	0.53	0.00	0.03
	CLPP5b	0.02	0.94	**2.03**	**1.77**	0.13	0.17
	CLPP6	0.10	0.57	**2.84**	**1.23**	0.35	0.09
	CLPR1	0.04	**1.79**	**8.86**	**2.40**	0.46	0.10
	CLPR2	0.02	**3.65**	**1.16**	**2.41**	0.06	0.00
	CLPR3	0.03	**8.64**	**1.24**	**2.29**	0.22	0.07
	CLPR4	0.01	**1.83**	**2.37**	**2.02**	0.08	0.06
	CLPC	0.02	0.22	0.47	0.75	0.02	0.03
	CLPT1	0.10	**1.66**	**1.31**	**7.31**	0.27	0.10
	CLPT2	0.17	0.67	0.80	**1.24**	0.31	0.36
	CLPS	0.02	0.03	0.05	0.39	0.00	0.04
	Concatenated	0.05	0.87	**1.30**	**1.55**	0.16	0.06
ACC	*ACCA*	0.01	0.99	0.49	–	0.06	0.05
	ACCB1	0.17	–	0.45	–	0.07	0.14
	ACCB2	0.67	0.18	0.11	–	0.42	0.59
	ACCC	0.01	0.08	0.08	–	0.08	0.06
	Concatenated	0.13	0.33	0.21	–	0.16	0.13
PSI	*LHCA2*	0.03	0.00	0.04	0.04	0.09	0.02
	LHCA3	0.02	0.05	0.13	0.06	0.02	0.07
	PSAK	0.00	0.07	0.28	0.45	0.00	0.00
	PSAL	0.02	0.08	0.18	0.11	0.00	0.00
	PSAN	0.07	0.14	0.09	0.05	0.09	0.00
	PSAO	0.09	0.00	0.20	0.13	0.21	0.05
	PSAP	0.18	0.12	0.03	0.14	0.18	0.16
	Concatenated	0.05	0.05	0.14	0.11	0.08	0.04
mtCLP	*CLPP2*	0.05	0.48	0.09	0.00	0.12	0.21
	CLPX1	0.06	0.21	0.08	0.12	0.07	0.02
	CLPX2	0.16	0.13	0.26	0.20	0.22	0.18
	CLPX3	0.12	0.52	0.32	0.08	0.31	0.30
	Concatenated	0.10	0.28	0.18	0.11	0.16	0.15
Random	Concatenated	0.12	0.15	0.12	0.15	0.13	0.15

Values > 1 are highlighted in bold, with underlined text indicating statistical significance based on likelihood ratio tests. Cells containing "–" indicate that the particular gene was not found in the corresponding transcriptome. Only concatenated results are reported for the set of 50 random genes. Individual gene results for this random set are available in Table S5. Note that CLPP5a and CLPP5b are recently duplicated paralogs and that only one of these copies was recovered from the *Agrostemma* transcriptome. Therefore, this *Agrostemma* sequence was used in both CLPP5 analyses. CLP, caseinolytic protease; ACC, acetyl-coA carboxylase; PSI, photosystem I; mtCLP, mitochondrial-targeted CLP protease.

lack a full complement of functional nuclear-encoded ACCase genes. Most of the species contained two *ACCB* paralogs, but we did not detect a copy of *ACCB1* in the *S. paradoxa* transcriptome. In addition, the assembly of the *S. paradoxa ACCA* transcript was incomplete, covering only 543 nt of a 2133-nt alignment and exhibiting a d_N/d_S of 0.99 (Table 1). This partial *ACCA* transcript was aberrantly spliced, resulting in a 9-nt insertion that introduced a premature in-frame stop codon (Figure S4). Therefore, despite being transcribed, *ACCA* is likely a pseudogene in *S. paradoxa*. In contrast, *ACCB2* and *ACCC* were both intact with very low d_N/d_S values in *S. paradoxa*. Finally, unlike in *S. noctiflora* and *S. paradoxa*, all four nuclear-encoded ACCase genes were present and intact in the transcriptomes of *S. conica* and all three "slow" species.

The ACCA subunit, which interacts directly with the plastid-encoded ACCD subunit, was highly divergent in aa sequence in *S. conica*. We initially estimated an elevated d_N/d_S value of 0.49 for *ACCA* in *S. conica*, and that value increased to 0.94 when we analyzed the full-length of the gene by excluding the partial *S. paradoxa* sequence from the alignment (PAML was run with the "cleandata" option, which ignores alignment positions for which gaps are present in any of the sequences). There was a striking difference between this elevated d_N/d_S in *S. conica* and the very low values (≤ 0.06) for *ACCA* in the slow group species (Table 1). The *S. conica ACCB1* gene also exhibited substantially higher d_N/d_S values than in any of the other species, whereas d_N/d_S was very low in *ACCB2* and *ACCC* (Table 1). In general, rates of nonsynonymous substitution in the slow species were very low ($d_N/d_S < 0.2$) for ACCase genes, with the exception of *ACCB2*. Interestingly, the d_N/d_S values for this gene showed a converse pattern, in which d_N/d_S was elevated in the slow species relative to *S. conica* and *S. paradoxa* (Table 1).

Low rates of nonsynonymous substitutions in mtCLP, PSI, and randomly selected genes

For both the PSI and mtCLP concatenated gene sets, there was significantly higher d_N/d_S in the fast group (Table 2). However, d_N/d_S values for both PSI and mtCLP were generally low in all species, and the differences between the two species groups were very small, especially in comparison to the

Table 2 Summary of d_N/d_S for constrained sets of fast and slow lineages based on concatenations of genes in each functional complex

Genes	d_N/d_S		*P*-Value
	Slow Lineages	Fast Lineages	
Plastid CLP	0.07	1.25	0.000
ACC	0.14	0.26	0.014
Photosystem I	0.05	0.10	0.002
mtCLP	0.11	0.20	0.002
Random	0.13	0.13	0.894

The *P*-values indicate significant differences in rate between the two sets of lineages based on LRTs. d_N/d_S, nonsynonymous to synonymous substitution rates; CLP, caseinolytic protease; ACC, acetyl-coA carboxylase; mtCLP, mitochondrial-targeted CLP protease.

differences observed for the plastid CLP complex and some ACCase genes (Figure S5 and Table 1). The lower overall d_N/d_S estimate for the slow species group appeared to be largely driven by the low values for the *Ag. githago* branch, which has a disproportionate effect on the estimation of d_N/d_S in this group because it represents more divergence time and a large fraction of total observed substitutions. The randomly selected nuclear genes showed no significant difference in d_N/d_S between the fast and slow species groups (Table 1, Table 2, and Table S5). Thus, it does not appear that there is a global elevation of d_N/d_S in the nuclear genomes of species with rapidly evolving plastid genomes.

MK tests reveal excess of nonsynonymous divergence between species

Despite high levels of observed nonsynonymous divergence in *ACCA* and the majority of the nuclear genes that encode components of the plastid CLP complex, most of the segregating variants within *S. conica* are synonymous. Thus, in *ACCA* and the concatenated set of nuclear-encoded *CLP* genes, there was a large and highly significant excess of nonsynonymous divergence from the outgroup *S. latifolia* relative to levels of nonsynonymous and synonymous polymorphism within *S. conica* (Table 3). The PSI genes had extremely low D_n/D_s values, but the P_n/P_s values were even lower, again resulting in a significant excess of nonsynonymous divergence for the concatenated gene set (Table 3). In contrast, there were no indications of a similar excess in the mtCLP genes or the set of randomly selected nuclear genes, as their concatenated sequences had NI values very close to one (Table 3 and Table S6). Repeating the MK analysis with a more closely related outgroup (*S. macrodonta*) produced similar results (Table S6 and Table S7).

Substitutions in nuclear-encoded subunits preferentially occur at interfaces with plastid-encoded subunits within the CLP complex, but not within the ACCase complex

To investigate the effect of physical interactions between CLP complex subunits on substitution patterns, we mapped observed changes in CLPP and CLPR subunits onto the solved structure of the representative ClpP subunit from *E. coli* (Figure 3). This analysis showed that numerous substitutions have occurred throughout the entirety of the plastid-encoded ClpP1 subunit in fast-evolving *Silene* species. These include substitutions in: (1) the handle domain, which physically interconnects the two heptameric rings and likely stabilizes ring–ring interactions; (2) the head domain, which likely stabilizes interactions between subunits within a single ring; and (3) the axial loop regions, which form the axial pores and mediate interactions with associated chaperones (Figure 3C) (Yu and Houry 2007). Several individual residues that have been implicated in ring stability or substrate interactions (Wang *et al.* 1997) were also observed to have undergone changes in ClpP1 (Figure 3).

Given the extreme levels of divergence in the plastid-encoded ClpP1 subunit, we reasoned that substitutions might be nonrandomly distributed within interacting nuclear-encoded subunits. Specifically, we predicted that the nuclear-encoded CLPR subunits would have an abundance of changes in the head domain because the ClpP1 subunit assembles with nuclear-encoded CLPR subunits to form the R-ring (Nishimura and van Wijk 2015), and the head domain contains residues that are likely to maintain intraring interactions (Yu and Houry 2007). Similarly, we predicted that nuclear-encoded CLPP subunits would have a disproportionate number of substitutions in their handle domains because these subunits form the P-ring, and their handle domains are likely involved in interactions with the highly divergent copies of ClpP1 in the R-ring. As predicted, aa substitutions were significantly overrepresented in the head domains of CLPR subunits and the handle domains of CLPP subunits in all three fast-evolving species (Table 4).

In most plants, both the nuclear- and plastid-encoded CLPP subunits (but not the CLPR subunits) retain the Ser-His-Asp triad that confers protease activity (Peltier *et al.* 2004). However, we found that in fast-evolving *Silene* species, many of these subunits have experienced substitutions at the His or Asp positions within this highly conserved triad (whereas the catalytic Ser is universally conserved in our data set; Table S8).

Relative to the CLP complex, there were fewer changes in the ACCase subunits, and only a fraction of these could be analyzed in a structural context because several regions of ACCase proteins lack structural information (Broussard *et al.* 2013). In particular, there is a large N-terminal extension of AccD that is highly variable among angiosperms (Greiner *et al.* 2008a) and absent entirely from *E. coli*. Likewise, the C-terminal half of ACCA is also unique to plants. In contrast to the pattern observed in the CLP complex, the ACCase changes that were able to be mapped to the *E. coli* structure occurred away from protein–protein interfaces and generally did not involve functionally important residues (Figure 4). This was also true for a site at which large insertions are present in the AccD subunit in both *S. conica* and *S. paradoxa* (Figure 4B).

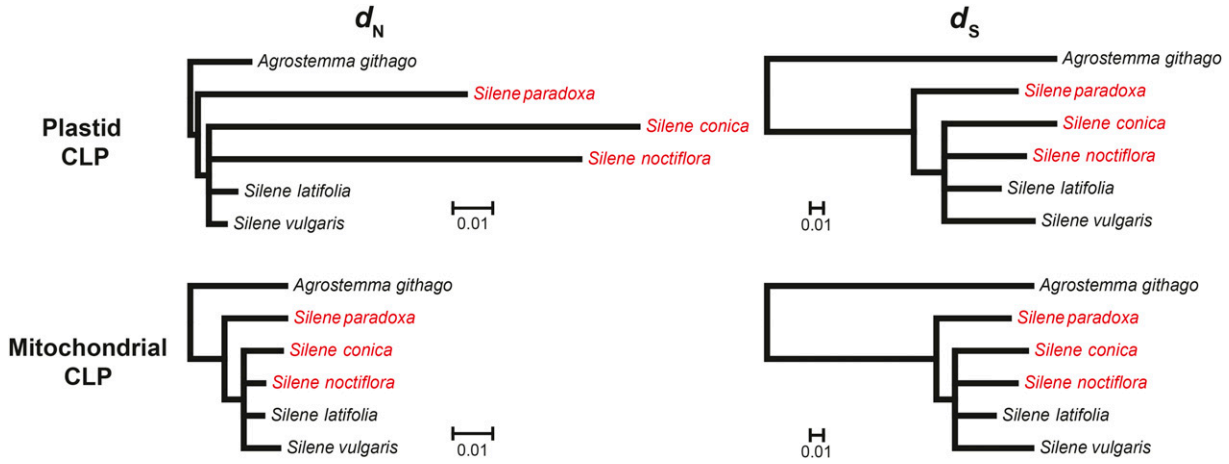

Figure 1 Rates of sequence evolution in nuclear genes coding for subunits of the plastid and mitochondrial CLP complexes. Branch lengths are scaled to the amount of nonsynonymous (d_N) and synonymous (d_S) divergence per site. Species with rapid rates of plastid genome evolution (Sloan *et al.* 2014a) are highlighted in red. CLP, caseinolytic protease.

Discussion

In numerous angiosperm lineages, a subset of plastid genes, including *clpP1* and *accD*, display accelerated evolutionary rates, but the causes of this recurring phenomenon have remained unclear (Jansen *et al.* 2007; Erixon and Oxelman 2008; Greiner *et al.* 2008b; Guisinger *et al.* 2008, 2010, 2011; Straub *et al.* 2011; Sloan *et al.* 2012a, 2014a; Barnard-Kubow *et al.* 2014; Weng *et al.* 2014; Dugas *et al.* 2015; Williams *et al.* 2015; Blazier *et al.* 2016; Zhang *et al.* 2016). We investigated the nuclear genes that contribute to the multisubunit complexes that include ClpP1 and AccD, and incorporated population genetic and structural data to distinguish between relaxed purifying selection and positive selection as drivers of elevated d_N/d_S values.

Our analysis revealed different patterns of selection on the nuclear-encoded CLP and ACCase genes, which may reflect the contrasting evolutionary histories of the plastid-encoded subunits in these two complexes. The patterns of *clpP1* sequence divergence in some lineages are truly remarkable and include both structural changes (*i.e.*, indels and loss of introns) and extreme increases in nonsynonymous substitution rates (*e.g.*, Erixon and Oxelman 2008; Sloan *et al.* 2012a). For example, the aa identity between the *clpP1* copies in the fast-evolving species *S. conica* and *S. noctiflora* is only 34%, and many *Silene* species have elevated d_N/d_S values (up to 5.9 in *S. fruticosa*; Erixon and Oxelman 2008). In contrast, slow-evolving species from the same genus, such as *S. latifolia*, retain up to 58% identity with free-living cyanobacteria, so these recent accelerations have led to far more divergence in the last few million years than has typically accumulated since the endosymbiotic origins of photosynthetic eukaryotes roughly one billion years ago. The contrasts between fast and slow lineages for *accD* are far less stark. The increased rates of aa substitution in fast lineages are only modest, and most of the sequence change is caused by indels (Sloan *et al.* 2012a, 2014a). Furthermore, even in species with typical, slow-evolving plastomes, it is primarily the catalytic C-terminal domain of AccD that is highly conserved, whereas the N-terminal domain, which is plant-specific and has an unknown function, accumulates substantial structural divergence (Greiner *et al.* 2008a). The evolution of *accD* is further complicated in some angiosperm lineages by functional transfer to the nucleus (Magee *et al.* 2010; Rousseau-Gueutin *et al.* 2013) or by functional replacement with a duplicated and retargeted copy of the homomeric ACCase (Konishi and Sasaki 1994). In contrast, there is no evidence (to our knowledge) of functional transfer of *clpP1* to the nucleus in green plants.

The complex history of the plastid *accD* gene in angiosperms is mirrored by the varied evolutionary histories that we observed within *Silene* for the nuclear-encoded ACCase subunits. One lineage (*S. noctiflora*) has experienced the outright loss of the heteromeric ACCase complex, and another lineage (*S. paradoxa*) appears to be undergoing gene loss/pseudogenization with signatures of relaxed selection. However, in a third fast species (*S. conica*), all the ACCase subunits are retained and one gene shows clear evidence of positive selection. In contrast to this heterogeneity in the evolution of ACCase genes, we found a consistent signal of positive selection throughout nearly all the subunits in the plastid CLP complex in all three fast species. It was especially striking to find d_N/d_S values significantly > 1 when averaged over more than a dozen nuclear-encoded CLP genes.

Loss of plastid heteromeric ACCase

The finding that *S. noctiflora* has completely lost the nuclear-encoded heteromeric ACCase genes is consistent with previous observations that the copy of the plastid-encoded *accD* may be a pseudogene in this species (Sloan *et al.* 2012a). Although the *accD* reading frame is intact in *S. paradoxa*, the gene is highly divergent and contains multiple large insertions, raising questions as to its functionality (Sloan *et al.*

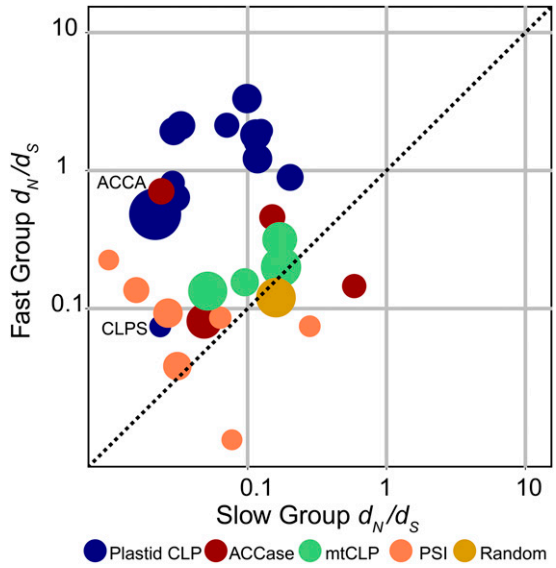

Figure 2 Gene-by-gene comparison of d_N/d_S in fast *vs.* slow sets of lineages. Points are color-coded by functional complex, and the diameter of each point is proportional to gene length. All points represent individual genes except the "random" point, which is based on the concatenation of all 50 genes in that set. The size of that point is scaled to the average length of each gene rather than the total concatenation length. ACCA and CLPS, which exhibit distinct patterns from the other ACCase and CLP subunits, are labeled individually. ACCase, acetyl-coA carboxylase; CLP, caseinolytic protease; PSI, photosystem I.

2014a). These questions extend to the entire ACCase complex, as we found evidence of gene loss/decay in nuclear-encoded *S. paradoxa* ACCase genes (*i.e.*, the apparent loss of *ACCB1* and pseudogenization of *ACCA*). In at least two angiosperm lineages, the plastid *accD* gene has been transferred to the nucleus (Magee *et al.* 2010; Rousseau-Gueutin *et al.* 2013). However, we found no evidence of a functional nuclear copy of *accD* in any *Silene* species examined. Instead, the presence of a duplicated homomeric ACCase that is predicted to be targeted to the plastids may be compensating for the lost or altered function of the heteromeric ACCase, as shown in grasses (Konishi and Sasaki 1994). Because the duplication of the homomeric ACCase appears to have happened long before the divergence of *Agrostemma* and *Silene* (Figure S3), the heteromeric and duplicated homomeric ACCases have remained conserved and expressed for more than 20 MY (Sloan *et al.* 2009) in many lineages within this clade. This raises intriguing questions for future investigation about the respective roles of these enzymes and why functional loss of the heteromeric ACCase has occurred in some lineages, while others have retained all of the subunits and even exhibit evidence of strong positive selection in one case (see below).

Positive selection acting on nuclear-plastid enzyme complexes

We found signatures of intense positive selection acting on the plastid CLP complex. In many genes within the fast species, d_N/d_S is > 1 (often significantly so; Table 1), which is an especially powerful signature of positive selection because any effects are averaged across the entire length of each gene and likely dampened by purifying selection acting on many residues. Although an increase in the rate of nonsynonymous substitutions can also be indicative of reduced functional importance or even a pseudogene, we can reject that hypothesis based on the population genetic data. The vast majority of the *CLP* sequence polymorphism that is segregating within *S. conica* is synonymous (Table 3), meaning that these genes are still functionally constrained and that purifying selection is purging most new nonsynonymous mutations from the population. Instead, the large observed excess of nonsynonymous divergence between species (relative to intraspecific polymorphism) is an indication that a specific subset of aa substitutions have been preferentially driven to fixation due to positive selection (McDonald and Kreitman 1991).

Interestingly, some of the observed sequence changes affected residues in the catalytic triad in *Silene* CLPP subunits (Table S8). Substitutions in the catalytic triad of the plastid-encoded ClpP1 subunit in *Acacia* have previously been interpreted as evidence of pseudogenization (Williams *et al.* 2015). However, the strong selection acting on the plastid CLP complex in *Silene* suggests that both plastid- and nuclear-encoded CLPP subunits may retain an important functional role despite changes in the catalytic triad (Table S8). Although the Ser-His-Asp catalytic site is a widely conserved feature across the diversity of life, some of the same substitutions in this triad have been observed in other atypical serine proteases and functionally related enzymes (Schrag *et al.* 1991; Ekici *et al.* 2008; Zeiler *et al.* 2013). Notably, substitutions in the catalytic triad were only observed in one nuclear-encoded *CLPP* gene in each of the fast species. It is possible that such changes are tolerable as long as some of the subunits in the P-ring retain the canonical catalytic residues.

Given the evidence of gene loss and relaxed selection on the heteromeric ACCase in *S. noctiflora* and *S. paradoxa*, we initially suspected that the elevated d_N/d_S for *ACCA* in *S. conica* (0.94 for the full-length gene) was also due to relaxed selection. However, the population genetic data demonstrated that *ACCA* and the heteromeric ACCase complex are still functionally constrained in *S. conica*, as most of the intraspecific polymorphisms were synonymous (Table 3). Therefore, the unusually high rate of fixed *ACCA* aa substitutions in this lineage are most likely the result of positive selection and adaptive evolution.

We randomly selected a set of nuclear genes that are not targeted to the plastids or mitochondria, as well as genes from the mtCLP and PSI complexes. These were chosen with the *a priori* expectation that they would be similar between the fast and slow species, because they either do not interact with organelle-encoded subunits (random genes and mtCLP), or they interact with plastid-encoded subunits that show typical, slow rates of evolution in all of the species (PSI). The mtCLP complex, which is comprised solely of nuclear subunits, and the set of random nuclear genes largely supported

Table 3 Summary of MK tests using intraspecific polymorphism within *S. conica* relative to interspecific divergence between *S. conica* and *S. latifolia*

	Gene	P_s	P_n	D_s	D_n	NI	DoS	*P*-Value
Plastid CLP	*CLPP3*	8	3	16	83	0.07	0.57	0.000
	CLPP4	18	0	31	51	0.00	0.62	0.000
	CLPP5a	5	2	4	19	0.08	0.54	0.006
	CLPP5b	0	0	11	25	–	–	–
	CLPP6	0	2	11	59	–	−0.16	0.543
	CLPR1	1	1	22	144	0.15	0.37	0.256
	CLPR3	3	2	48	120	0.27	0.31	0.128
	CLPR4	1	1	22	80	0.28	0.28	0.337
	CLPC	4	1	55	50	0.28	0.28	0.226
	CLPT1	4	0	10	34	0.00	0.77	0.001
	CLPT2	7	1	19	39	0.07	0.55	0.002
	CLPS	3	0	11	2	0.00	0.15	0.467
	Concatenated	54	13	260	706	0.09 (0.11)	0.54	0.000
ACC	*ACCA*	13	7	42	111	0.20	0.38	0.001
	ACCB1	0	1	14	11	–	−0.56	0.270
	ACCB2	4	3	10	7	1.07	−0.02	0.939
	ACCC	10	3	22	18	0.37	0.22	0.160
	Concatenated	27	14	88	147	0.31 (0.31)	0.28	0.001
Photosystem I	*LHCA2*	4	0	19	6	0.00	0.24	0.271
	LHCA3	13	1	28	9	0.24	0.17	0.167
	PSAK	1	0	13	7	0.00	0.35	0.468
	PSAL	5	0	15	5	0.00	0.25	0.211
	PSAN	11	2	18	10	0.33	0.20	0.183
	Concatenated	34	3	93	37	0.22 (0.20)	0.20	0.011
mtCLP	*CLPP2*	1	1	21	4	5.25	−0.34	0.233
	CLPX2	4	5	26	23	1.41	−0.09	0.634
	CLPX3	9	3	12	8	0.50	0.15	0.387
	Concatenated	14	9	59	35	1.08 (1.05)	−0.02	0.866
Random	Concatenated	368	178	1297	570	1.10 (1.07)	−0.02	0.371

For each gene, the NI and the DoS were calculated, with NI values < 1 and DoS values > 0 indicative of positive selection (*i.e.*, an excess of nonsynonymous divergence). For concatenations of genes within each complex, NI_{TG} (an unbiased estimator of NI) is reported in parentheses. Only concatenated results are reported for the set of random genes. Individual gene results for this random set are available in Table S6. MK, McDonald–Kreitman test; P_s, synonymous polymorphisms; P_n, nonsynonymous polymorphisms; D_s, synonymous divergence; D_n, nonsynonymous divergence; NI, neutrality index; DoS, direction of selection; CLP, caseinolytic protease; ACC, acetyl-coA carboxylase; mtCLP, mitochondrial-targeted CLP protease.

this expectation. The concatenated mtCLP genes in fast species exhibited very similar (albeit slightly higher) d_N/d_S levels compared to slow species, and there was no significant difference for the random genes (Table 1 and Table 2). Furthermore, the MK tests found no evidence of positive selection for either of these data sets (Table 3, Table S6, and Table S7). In contrast, the nuclear PSI genes did exhibit a significant excess of nonsynonymous divergence (Table 3 and Table S7), even though the absolute rate of aa substitutions was extremely low. This evidence suggests that, although rare, some of the nonsynonymous substitutions in PSI are adaptive changes that spread under positive selection rather than fixing by drift.

An alternate interpretation for the observed excess of nonsynonymous divergence between species is that there was an ancestral bottleneck, which could have led to an increased frequency of weakly deleterious alleles spreading to fixation because of the reduced efficiency of selection in small populations (Hughes 2007). However, we identified three lines of evidence that lead us to reject this possibility and conclude that our results are, in fact, indicative of strong positive selection. First, if a demographic bottleneck had occurred, we would expect an excess of nonsynonymous divergence across all genes, but we did not observe this in the mtCLP genes or the randomly chosen nuclear genes. Previous studies in *Silene* also support the conclusion that the observed changes are not the result of genome-wide demographic effects; analysis of 140 cytosolic ribosomal proteins and seven mitochondrial-targeted complex II genes in *S. conica* and *S. noctiflora* did not show elevated d_N/d_S relative to other *Silene* species (Sloan *et al.* 2014b; Havird *et al.* 2015). Second, we observed elevated levels of nonsynonymous divergence across different timescales in our MK tests (Table 3 and Table S7) by using two different outgroups: *S. latifolia* (~5.7 MY divergence time) and *S. macrodonta* (~1.8 MY divergence time; Rautenberg *et al.* 2012). Therefore, separate bottlenecks at a minimum of two different historical points would had to have taken place. Third, the magnitude of the observed effects is inconsistent with a bottleneck. The relaxed selection associated with a reduced N_e should not increase d_N/d_S to values significantly > 1, which were frequently observed in our data set (Table 1). Thus, we conclude that, although it is possible that an ancestral bottleneck in *S. conica* might have contributed to some minor increases in aa

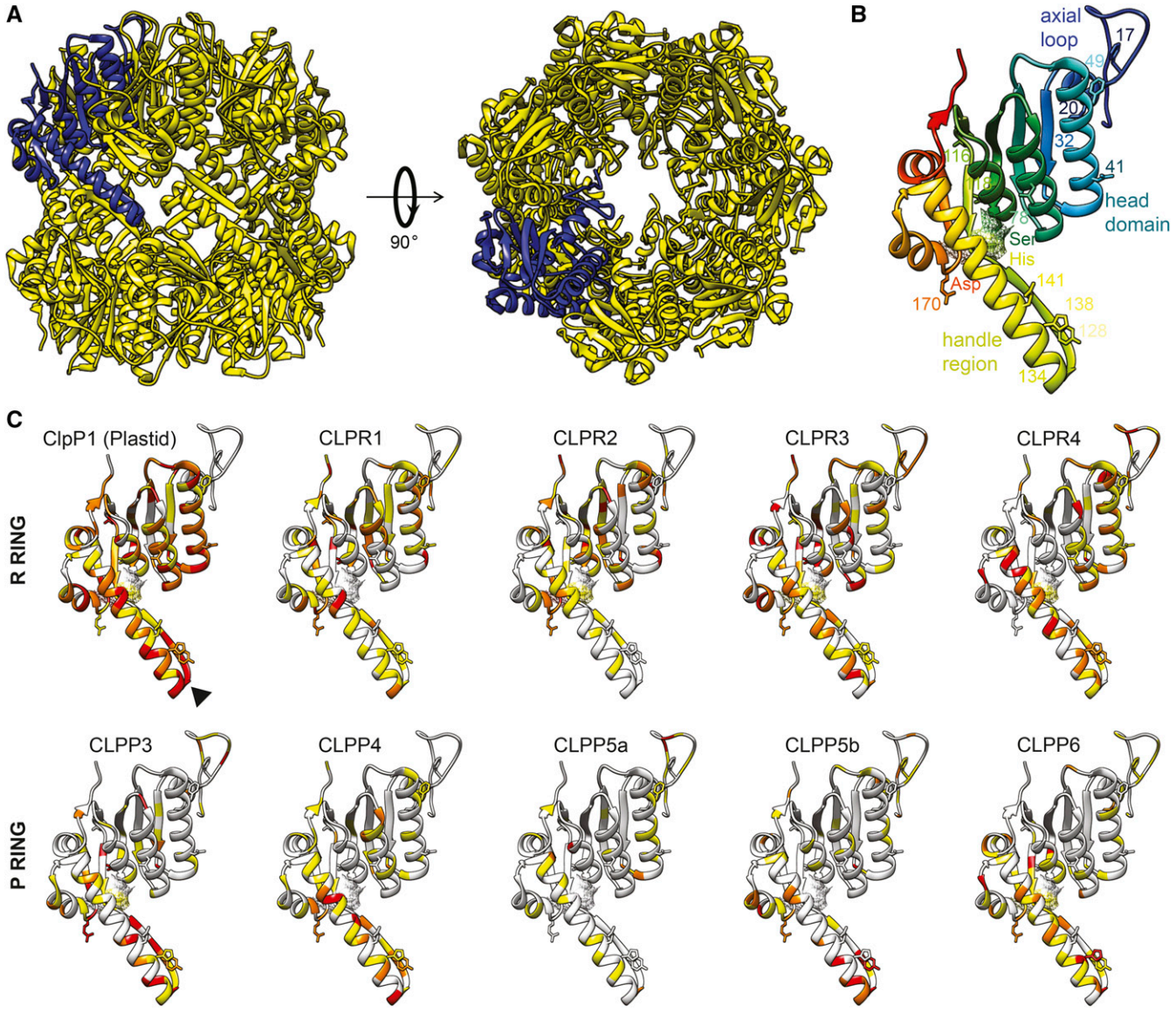

Figure 3 (A) ClpP protease structure from *E. coli* (PDB 1YG6; Yu and Houry 2007) oriented with the two heptameric rings stacked on top of each other (left) or in front and behind (right). One of the 14 (identical) *E. coli* subunits is highlighted in blue. (B) A single *E. coli* ClpP subunit (as in part (A)), with the head, handle, and axial-loop domains highlighted. The surface of the catalytic triad is indicated in mesh and the individual residues (Ser, His, and Asp) are labeled. Other important residues are numbered and indicated with stick models, including those that interact with substrates (Phe-17 and Phe-49) or contribute to heptamer stability by forming hydrophobic bonds (Asn-41/Tyr-20, Asn-41/Thr-32, Asp-78/Asn-116, and Asp-171/Try-128) or ion pairs (Arg-118/Glu-141, Arg-170/Glu-134, and Asp-171/His-138). (C) Locations of substitutions in *Silene* species with fast-evolving CLP sequences (*S. conica*, *S. noctiflora*, and *S. paradoxa*). Residues with changes (relative to the inferred ancestral *Silene* sequence) in one, two, or three species are indicated in yellow, orange, and red, respectively. The black triangle highlights the ClpP1 site at which large insertions have occurred in *S. conica* (39 aa) and *S. noctiflora* (seven aa). CLP, caseinolytic protease.

substitution rates, the massive rate increases (*e.g.*, in CLP genes) are more likely to have been driven by positive selection than by a temporary reduction in N_e.

Antagonistic coevolution and plastid–nuclear conflict

The accelerated aa substitution rates in both the nuclear- and plastid-encoded components of CLP and ACCase are very unusual for anciently conserved enzyme complexes, but are similar in many ways to the patterns that result from antagonistic coevolution between pathogens and host immune genes (Hughes and Nei 1988; Borghans *et al.* 2004). Selfish interactions and "arms races" can occur within a cell (*i.e.*, intragenomic conflict) when there is opportunity for genetic elements to enhance their own transmission at the expense of organismal-level fitness (Burt and Trivers 2006). Such conflicts are common between the nucleus and cytoplasmic genomes. For example, copies of mitochondrial genomes with large deletions can confer a replication advantage within the cell, even if they harm overall fitness by reducing or eliminating the cell's ability to respire (Taylor *et al.* 2002;

Table 4 The ratio of aa substitutions in the head *vs.* handle regions is higher in CLPR subunits than CLPP subunits

Species	Head:Handle Ratio		Fisher's Exact Test *P*-Value
	CLPP	CLPR	
S. conica	85:63	136:61	0.031
S. noctiflora	50:37	109:36	0.006
S. paradoxa	49:40	127:22	0.000

Values represent counts of substitutions summed across all nuclear-encoded genes in each category. CLP, caseinolytic protease.

Clark *et al.* 2012; Phillips *et al.* 2015). In addition, because most cytoplasmic genomes are inherited maternally, they can benefit from manipulating sexual reproduction to increase female reproduction and fitness (Perlman *et al.* 2015). Examples of this phenomenon include chimeric ORFs in plant mitochondrial genomes that induce cytoplasmic male sterility (CMS; Ingvarsson and Taylor 2002; Touzet and Budar 2004; Fujii *et al.* 2011) and numerous bacterial endosymbionts that manipulate sexual reproduction in animal hosts (Werren *et al.* 2008).

Interestingly, some of the earliest hypotheses about cytonuclear conflict were developed based on observations of differential rates of replication of plastid genomes in heteroplasmic plants (Grun 1976; reviewed in Greiner *et al.* 2015). Since that point, however, research on cytonuclear conflict in plants has overwhelmingly focused on mitochondria, particularly their role in CMS. Although plastids are often viewed as being relatively benign, in principle, the same evolutionary pressures could apply to these maternally inherited organelles. One possibility is that there are limited pathways available for plastids to exploit in a selfish fashion. For example, the major role of plastids (specifically chloroplasts) is in photosynthesis, and male reproductive tissues are generally nonphotosynthetic. However, plastids also perform other important processes (including CLP and ACCase activity). Recent studies have provided support for the possibility of selfish plastid–nuclear interactions within complexes such as the heteromeric ACCase and the CLP complex. In particular, reproductive incompatibilities (including male sterility) between wild and domesticated lines of peas were recently attributed to variation in nuclear- and plastid-encoded components of the heteromeric ACCase (Bogdanova *et al.* 2015). Plastid–nuclear incompatibilities have also been implicated in male sterility in *Oenothera* (Stubbe and Steiner 1999). Disruption of the synthesis of fatty acids and their derivatives, such as jasmonic acid, has also been associated with sterility phenotypes (Park *et al.* 2002). In addition, a recent proteomic analysis in wheat found that plastid-encoded *clpP1* was one of the most upregulated genes in the anthers of male-sterile individuals, suggesting that it may play important functional roles in male reproductive tissues (Li *et al.* 2015).

Although the correlated increases in evolutionary rates and signatures of positive selection in *Silene* plastid–nuclear complexes could indicate a history of genomic conflict, they are not conclusive evidence that plastid and nuclear genes are locked in an arms race, or even that they are coevolving in any fashion (Lovell and Robertson 2010). General changes in selection for CLP or ACCase function could simultaneously affect all subunits, without interactions among the subunits being a major source of selection. It is also possible that the overall structure or subunit composition of the CLP complex has been radically disrupted or reorganized. Furthermore, even if coevolutionary dynamics are at play, they could involve mutually beneficial changes rather than antagonistic interactions (Rand *et al.* 2004). Adaptive changes in one genome may alter fitness landscapes and facilitate subsequent adaptive changes in the other genome, though it is unclear what forces might trigger such runaway adaptive evolution in these systems. A more conventional model of compensatory change in the nucleus in response to accumulation of deleterious changes in asexual organelle genomes (Rand *et al.* 2004; Osada and Akashi 2012) seems less likely, particularly for the CLP complex, given the evidence that accelerated rates of sequence evolution in the plastid-encoded *clpP1* evolution are driven largely by positive selection (Erixon and Oxelman 2008; Sloan *et al.* 2012a; Barnard-Kubow *et al.* 2014).

It is worth noting that the largest increases in substitution rates were found in the nuclear-encoded subunits that interact most directly with plastid-encoded subunits. Specifically, the greatest elevation of d_N/d_S in the CLP complex occurred in the CLPR subunits (Table 1), which assemble with plastid-encoded ClpP1 subunits to form the proteolytic R-ring (Nishimura and van Wijk 2015). More detailed structural analysis of the CLP complex (Figure 3 and Table 4) showed that, even within subunits, there was an enrichment for substitutions in domains that have intimate interactions with ClpP1 (*i.e.*, the head domains of CLPR proteins and handle domains of CLPP proteins; Yu and Houry 2007; Nishimura and van Wijk 2015). For the heteromeric ACCase, we found evidence of positive selection (Table 3) only on the nuclear-encoded ACCA subunit in *S. conica*, which interacts directly with the plastid-encoded ACCD subunit to make up the carboxyltransferase (Sasaki and Nagano 2004). However, our structural analysis did not detect an enrichment of substitutions at the interface between ACCA and AccD (Figure 4), and it is difficult to draw firm conclusions about ACCase structure given the lack of knowledge about the functions and interactions of the plant-specific portions of ACCA and AccD.

Therefore, it appears more likely that plastid–nuclear interactions and coevolution have played a role in generating positive selection and the observed accelerations in rates of sequence evolution in the CLP complex than in the heteromeric ACCase. We speculate that the mysterious rate accelerations that have occurred repeatedly in *clpP1* in *Silene*, and throughout the diversification of flowering plants (and possibly those that have occurred in other plastid genes as well), are the result of antagonistic coevolution between the plastid and nucleus. An important test of this hypothesis will be to functionally characterize the sequence changes from rapidly

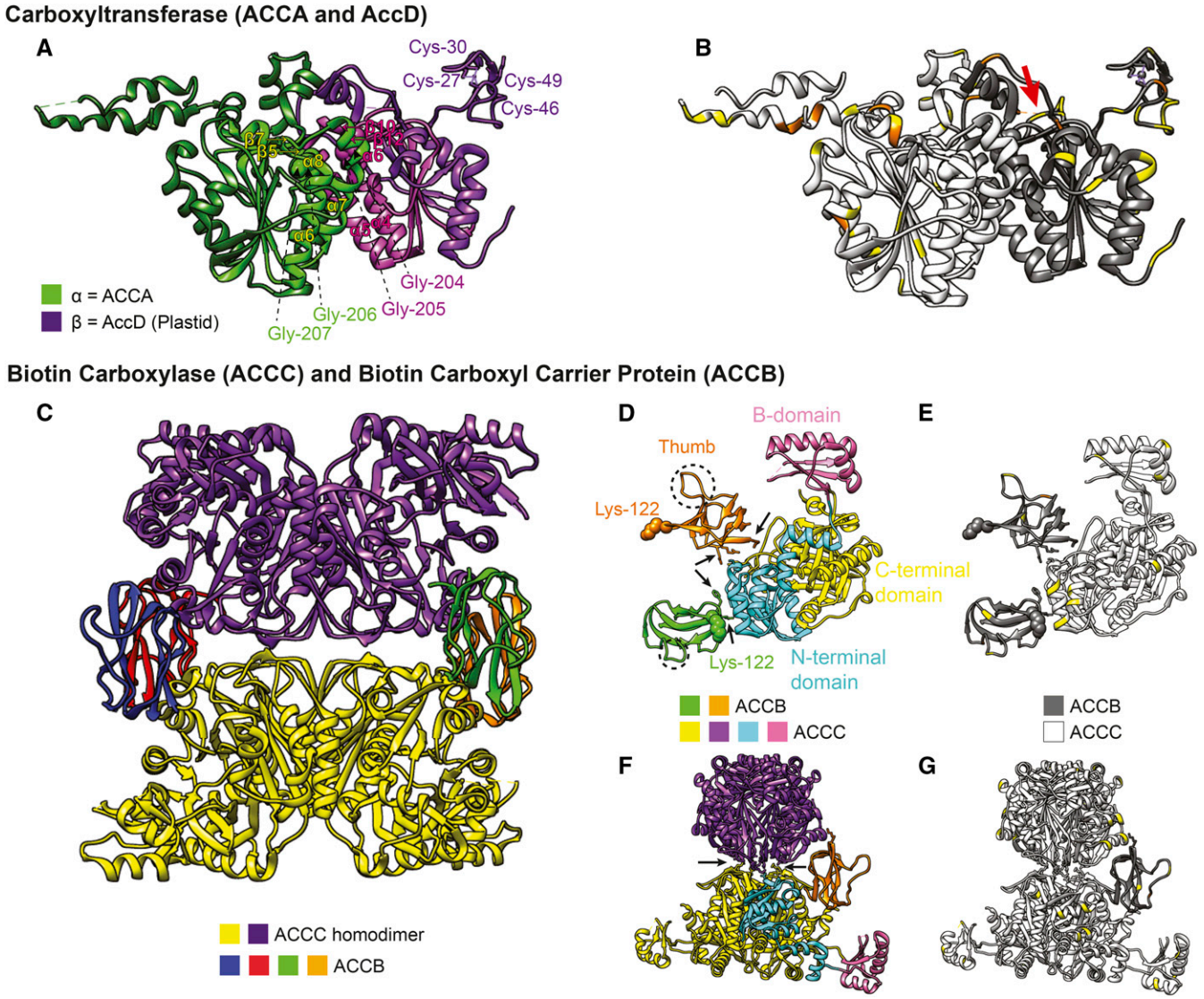

Figure 4 (A) Structure of the carboxyltransferase component from the *E. coli* ACCase (PDB 2F9Y), oriented as in Figure 2C from Bilder *et al.* (2006). Residues of active sites (Gly-204, Gly-205, Gly-206, and Gly-207) are shown as spheres, the four cysteines that form cysteinyl zinc ligands (Cys-27, Cys-30, Cys-46, and Cys-49) are shown as stick models, and the helices and sheets that compose the catalytic platform are indicated. (B) The same structure as in (A) with inferred *Silene* changes mapped as described in Figure 3. The red arrow indicates the location of an AccD site with large insertions in both *S. conica* (26 aa) and *S. paradoxa* (58 aa). Importantly, this model contains only the portions of ACCA and AccD that can be aligned and mapped to *E. coli*, as the copies of these genes in plants are roughly twice the length of their counterparts in *E. coli* owing to large N- or C-terminal extensions of unknown function. (C) Structure of the biotin carboxylase (ACCC)-biotin carboxyl carrier protein (ACCB) complex from the *E. coli* ACCase (PDB 4HR7), as in Figure 2A from Broussard *et al.* (2013). (D) An ACCC monomer complexed with two ACCB monomers, as in Figure 2D in Broussard *et al.* (2013), indicating functional domains and ACCB–ACCC interfaces. The Lys-122 residues that bind biotin and the active site of ACCC (Arg-338) are shown as spheres, while residues that play important roles in ACCB-ACCC interactions are shown as sticks and highlighted with arrows. (E) The same structure as in (D) with inferred *Silene* changes mapped as described in Figure 3. (F) Two ACCC dimers and a single ACCB monomer, showing ACCC–ACCC interfaces as in (A) from Broussard *et al.* (2013). Residues that play important roles in ACCC–ACCC interactions are shown as sticks and highlighted with arrows. (G) The same structure as part (F), with inferred *Silene* changes mapped as described in Figure 3. ACCase, acetyl-coA carboxylase.

evolving lineages, particularly with respect to their phenotypic effects on plastid replication within cells and plant allocation to male *vs.* female reproductive output.

Acknowledgments

We thank Andrea Berardi, Rolland Douzet, Peter Fields, Michael Hood, Andreas König, Arne Saatkamp, the Kew Millenium Seed Bank, the Ornamental Plant Gerplasm Center, and the Ville de Nantes Jardin Botanique for collecting/providing seeds. We also thank Cody Kalous and Jessica Hurley for performing RNA extractions and quality control. We are grateful for valuable comments from Stephen Wright and two anonymous reviewers on an earlier version of this manuscript. This research was supported by grants from the National Science Foundation (NSF) (MCB-1412260 and

MCB-1022128). K.R. is supported by a Graduate Assistance in Areas of National Need graduate fellowship from the U.S. Department of Education (P200A140008) and is a participant in the NSF-funded Generating, Analyzing, and Understanding Sensory and Sequencing Information graduate training program (DGE-1450032). J.C.H. is supported by a National Institutes of Health Postdoctoral Fellowship (F32-GM-116361).

Literature Cited

Babiychuk, E., K. Vandepoele, J. Wissing, M. Garcia-Diaz, R. De Rycke *et al.*, 2011 Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. Proc. Natl. Acad. Sci. USA 108: 6674–6679.

Barnard-Kubow, K. B., D. B. Sloan, and L. F. Galloway, 2014 Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. BMC Evol. Biol. 14: 268.

Bewley, M. C., V. Graziano, K. Griffin, and J. M. Flanagan, 2006 The asymmetry in the mature amino-terminus of ClpP facilitates a local symmetry match in ClpAP and ClpXP complexes. J. Struct. Biol. 153: 113–128.

Bilder, P., S. Lightle, G. Bainbridge, J. Ohren, B. Finzel *et al.*, 2006 The structure of the carboxyltransferase component of acetyl-CoA carboxylase reveals a zinc-binding motif unique to the bacterial enzyme. Biochemistry 45: 1712–1722.

Blazier, J. C., T. A. Ruhlman, M.-L. Weng, S. K. Rehman, J. S. M. Sabir *et al.*, 2016 Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci. Rep. 6: 24595.

Bogdanova, V. S., O. O. Zaytseva, A. V. Mglinets, N. V. Shatskaya, O. E. Kosterin *et al.*, 2015 Nuclear-cytoplasmic conflict in pea (Pisum sativum L.) is associated with nuclear and plastidic candidate genes encoding acetyl-CoA carboxylase subunits. PLoS One 10: e0119835.

Borghans, J. A. M., J. B. Beltman, and R. J. De Boer, 2004 MHC polymorphism under host-pathogen coevolution. Immunogenetics 55: 732–739.

Broussard, T. C., A. E. Price, S. M. Laborde, and G. L. Waldrop, 2013 Complex formation and regulation of Escherichia coli acetyl-CoA carboxylase. Biochemistry 52: 3346–3357.

Burt, A., and R. Trivers, 2006 *Genes in Conflict: The Biology of Selfish Genetic Elements*. Harvard University Press, Cambridge, MA.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Clark, K. A., D. K. Howe, K. Gafner, D. Kusuma, S. Ping *et al.*, 2012 Selfish little circles: transmission bias and evolution of large deletion-bearing mitochondrial DNA in Caenorhabditis briggsae nematodes. PLoS One 7: e41433.

Cronan, J. E., and G. L. Waldrop, 2002 Multi-subunit acetyl-CoA carboxylases. Prog. Lipid Res. 41: 407–435.

Duarte, J. M., P. K. Wall, P. P. Edger, L. L. Landherr, H. Ma *et al.*, 2010 Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol. Biol. 10: 61.

Dugas, D. V., D. Hernandez, E. J. M. Koenen, E. Schwarz, S. Straub *et al.*, 2015 Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Sci. Rep. 5: 16958.

Egea, R., S. Casillas, and A. Barbadilla, 2008 Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res. 36: W157–W162.

Ekici, O. D., M. Paetzel, and R. E. Dalbey, 2008 Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. Protein Sci. 17: 2023–2037.

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

Erixon, P., and B. Oxelman, 2008 Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast clpP1 gene. PLoS One 3: e1386.

Fujii, S., C. S. Bond, and I. D. Small, 2011 Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc. Natl. Acad. Sci. USA 108: 1723–1728.

Fukuda, N., Y. Ikawa, T. Aoyagi, and A. Kozaki, 2013 Expression of the genes coding for plastidic acetyl-CoA carboxylase subunits is regulated by a location-sensitive transcription factor binding site. Plant Mol. Biol. 82: 473–483.

Gould, S. B., R. F. Waller, and G. I. McFadden, 2008 Plastid evolution. Annu. Rev. Plant Biol. 59: 491–517.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Gray, M. W., and J. M. Archibald, 2012 Genomics of chloroplasts and mitochondria, pp. 1–30 in *Genomics of Chloroplasts and Mitochondria*, *Advances in Photosynthesis and Respiration*, edited by R. Bock, and V. Knoop. Springer Netherlands, Dordrecht.

Greiner, S., X. Wang, U. Rauwolf, M. V. Silber, K. Mayer *et al.*, 2008a The complete nucleotide sequences of the five genetically distinct plastid genomes of Oenothera, subsection Oenothera: I. Sequence evaluation and plastome evolution. Nucleic Acids Res. 36: 2366–2378.

Greiner, S., X. Wang, R. G. Herrmann, U. Rauwolf, K. Mayer *et al.*, 2008b The complete nucleotide sequences of the 5 genetically distinct plastid genomes of Oenothera, subsection Oenothera: II. A microevolutionary view using bioinformatics and formal genetic data. Mol. Biol. Evol. 25: 2019–2030.

Greiner, S., J. Sobanski, and R. Bock, 2015 Why are most organelle genomes transmitted maternally? BioEssays 37: 80–94.

Grun, P., 1976 *Cytoplasmic Genetics and Evolution*. Columbia University Press, New York.

Guisinger, M. M., J. V. Kuehl, J. L. Boore, and R. K. Jansen, 2008 Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. Proc. Natl. Acad. Sci. USA 105: 18424–18429.

Guisinger, M. M., T. W. Chumley, J. V. Kuehl, J. L. Boore, and R. K. Jansen, 2010 Implications of the plastid genome sequence of typha (typhaceae, poales) for understanding genome evolution in poaceae. J. Mol. Evol. 70: 149–166.

Guisinger, M. M., J. V. Kuehl, J. L. Boore, and R. K. Jansen, 2011 Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol. Biol. Evol. 28: 583–600.

Havird, J. C., N. S. Whitehill, C. D. Snow, and D. B. Sloan, 2015 Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. Evolution 69: 3069–3081.

Hughes, A. L., 2007 Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity (Edinb) 99: 364–373.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170.

Ingvarsson, P. K., and D. R. Taylor, 2002 Genealogical evidence for epidemics of selfish genes. Proc. Natl. Acad. Sci. USA 99: 11265–11269.

Jansen, R. K., Z. Cai, L. A. Raubeson, H. Daniell, C. W. Depamphilis *et al.*, 2007 Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc. Natl. Acad. Sci. USA 104: 19369–19374.

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

Konishi, T., and Y. Sasaki, 1994 Compartmentalization of two forms of acetyl-CoA carboxylase in plants and the origin of their tolerance toward herbicides. Proc. Natl. Acad. Sci. USA 91: 3598–3601.

Li, Y.-Y., Y.-Y. Li, Q.-G. Fu, H.-H. Sun, and Z.-G. Ru, 2015 Anther proteomic characterization in temperature sensitive Bainong male sterile wheat. Biol. Plant. 59: 273–282.

Lovell, S. C., and D. L. Robertson, 2010 An integrated view of molecular coevolution in protein-protein interactions. Mol. Biol. Evol. 27: 2567–2575.

Magee, A. M., S. Aspinall, D. W. Rice, B. P. Cusack, M. Sémon *et al.*, 2010 Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 20: 1700–1710.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652–654.

Mower, J. P., P. Touzet, J. S. Gummow, L. F. Delph, and J. D. Palmer, 2007 Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7: 135.

Nishimura, K., and K. J. van Wijk, 2015 Organization, function and substrates of the essential Clp protease system in plastids. Biochim. Biophys. Acta 1847: 915–930.

Nishimura, K., J. Apitz, G. Friso, J. Kim, L. Ponnala *et al.*, 2015 Discovery of a unique Clp component, ClpF, in chloroplasts: a proposed binary ClpF-ClpS1 adaptor complex functions in substrate recognition and delivery. Plant Cell 27: 2677–2691.

Olinares, P. D. B., J. Kim, and K. J. van Wijk, 2011 The Clp protease system; a central component of the chloroplast protease network. Biochim. Biophys. Acta 1807: 999–1011.

Osada, N., and H. Akashi, 2012 Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. Mol. Biol. Evol. 29: 337–346.

Park, J.-H., R. Halitschke, H. B. Kim, I. T. Baldwin, K. A. Feldmann *et al.*, 2002 A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in Arabidopsis due to a block in jasmonic acid biosynthesis. Plant J. 31: 1–12.

Parker, N., Y. Wang, and D. Meinke, 2014 Natural variation in sensitivity to a loss of chloroplast translation in Arabidopsis. Plant Physiol. 166: 2013–2027.

Peltier, J.-B., D. R. Ripoll, G. Friso, A. Rudella, Y. Cai *et al.*, 2004 Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. J. Biol. Chem. 279: 4768–4781.

Perlman, S. J., C. N. Hodson, P. T. Hamilton, G. P. Opit, and B. E. Gowen, 2015 Maternal transmission, sex ratio distortion, and mitochondria. Proc. Natl. Acad. Sci. USA 112: 10162–10168.

Phillips, W. S., A. L. Coleman-Hulbert, E. S. Weiss, D. K. Howe, S. Ping *et al.*, 2015 Selfish mitochondrial DNA proliferates and diversifies in small, but not large, experimental populations of Caenorhabditis briggsae. Genome Biol. Evol. 7: 2023–2037.

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. 13: 735–748.

Rand, D. M., R. A. Haney, and A. J. Fry, 2004 Cytonuclear coevolution: the genomics of cooperation. Trends Ecol. Evol. 19: 645–653.

Rausher, M. D., Y. Lu, and K. Meyer, 2008 Variation in constraint *vs.* positive selection as an explanation for evolutionary rate variation among anthocyanin genes. J. Mol. Evol. 67: 137–144.

Rautenberg, A., D. B. Sloan, V. Aldén, and B. Oxelman, 2012 Phylogenetic relationships of Silene multinervia and Silene section Conoimorpha (Caryophyllaceae). Syst. Bot. 37: 226–237.

Rousseau-Gueutin, M., X. Huang, E. Higginson, M. Ayliffe, A. Day *et al.*, 2013 Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. Plant Physiol. 161: 1918–1929.

Salie, M. J., and J. J. Thelen, 2016 Regulation and structure of the heteromeric acetyl-CoA carboxylase. Biochim. Biophys. Acta 1861: 1207–1213.

Sasaki, Y., and Y. Nagano, 2004 Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. Biosci. Biotechnol. Biochem. 68: 1175–1184.

Schrag, J. D., Y. G. Li, S. Wu, and M. Cygler, 1991 Ser-His-Glu triad forms the catalytic site of the lipase from Geotrichum candidum. Nature 351: 761–764.

Schulte, W., R. Töpfer, R. Stracke, J. Schell, and N. Martini, 1997 Multi-functional acetyl-CoA carboxylase from Brassica napus is encoded by a multi-gene family: indication for plastidic localization of at least one isoform. Proc. Natl. Acad. Sci. USA 94: 3465–3470.

Sloan, D. B., B. Oxelman, A. Rautenberg, and D. R. Taylor, 2009 Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae. BMC Evol. Biol. 9: 260.

Sloan, D. B., A. J. Alverson, M. Wu, J. D. Palmer, and D. R. Taylor, 2012a 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, D. B., A. J. Alverson, J. P. Chuckalovcak, M. Wu, D. E. McCauley *et al.*, 2012b Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 10: e1001241.

Sloan, D. B., D. A. Triant, N. J. Forrester, L. M. Bergner, M. Wu *et al.*, 2014a A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol. Phylogenet. Evol. 72: 82–89.

Sloan, D. B., D. A. Triant, M. Wu, and D. R. Taylor, 2014b Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol. Biol. Evol. 31: 673–682.

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz *et al.*, 2002 The Bioperl toolkit: perl modules for the life sciences. Genome Res. 12: 1611–1618.

Stoletzki, N., and A. Eyre-Walker, 2011 Estimation of the neutrality index. Mol. Biol. Evol. 28: 63–70.

Straub, S. C. K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks *et al.*, 2011 Building a model: developing genomic resources for common milkweed (Asclepias syriaca) with low coverage genome sequencing. BMC Genomics 12: 211.

Stubbe, W., and E. Steiner, 1999 Inactivation of pollen and other effects of genome-plastome incompatibility in Oenothera. Plant Syst. Evol. 217: 259–277.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar, 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Taylor, D. R., C. Zeyl, and E. Cooke, 2002 Conflicting levels of selection in the accumulation of mitochondrial defects in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA 99: 3690–3694.

Timmis, J. N., M. A. Ayliffe, C. Y. Huang, and W. Martin, 2004 Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. 5: 123–135.

Touzet, P., and F. Budar, 2004 Unveiling the molecular arms race between two conflicting genomes in cytoplasmic male sterility? Trends Plant Sci. 9: 568–570.

van Wijk, K. J., 2015 Protein maturation and proteolysis in plant plastids, mitochondria, and peroxisomes. Annu. Rev. Plant Biol. 66: 75–111.

Wang, J., J. A. Hartling, and J. M. Flanagan, 1997 The structure of ClpP at 2.3 a resolution suggests a model for ATP-dependent proteolysis. Cell 91: 447–456.

Weng, M.-L., J. C. Blazier, M. Govindu, and R. K. Jansen, 2014 Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol. Biol. Evol. 31: 645–659.

Weng, M.-L., T. A. Ruhlman, and R. K. Jansen, 2016 Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol. Evol. 8: evw115.

Werren, J. H., L. Baldo, and M. E. Clark, 2008 Wolbachia: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6: 741–751.

White, S. W., J. Zheng, Y.-M. Zhang, and Rock, 2005 The structural biology of type II fatty acid biosynthesis. Annu. Rev. Biochem. 74: 791–831.

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

Williams, A. V., L. M. Boykin, K. A. Howell, P. G. Nevill, and I. Small, 2015 The complete sequence of the Acacia ligulata chloroplast genome reveals a highly divergent clpP1 gene. PLoS One 10: e0125768.

Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson *et al.*, 2014 SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30: 1660–1666.

Yang, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586–1591.

Yu, A. Y. H., and W. A. Houry, 2007 ClpP: a distinctive family of cylindrical energy-dependent serine proteases. FEBS Lett. 581: 3749–3757.

Zeiler, E., A. List, F. Alte, M. Gersch, R. Wachtel *et al.*, 2013 Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. Proc. Natl. Acad. Sci. USA 110: 11302–11307.

Zhang, J., T. A. Ruhlman, J. Sabir, J. C. Blazier, and R. K. Jansen, 2015 Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. Plant Cell 27: 563–573.

Zhang, J., T. A. Ruhlman, J. Sabir, J. C. Blazier, M.-L. Weng *et al.*, 2016 Coevolution between nuclear encoded DNA replication, recombination and repair genes and plastid genome complexity. Genome Biol. Evol. 8: 622–634.

Communicating editor: S. I. Wright

File S1 Sequence alignments used in PAML, MK, and phylogenetic analyses. (.zip, 304 KB)

Available for download as a .zip file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.188268/-/DC1/FileS1.zip