Contrasting Patterns of Nucleotide Substitution Rates Provide Insight into Dynamic Evolution of Plastid and Mitochondrial Genomes of *Geranium*

Seongjun Park[1], Tracey A. Ruhlman[1], Mao-Lun Weng[1,2], Nahid H. Hajrah[3], Jamal S.M. Sabir[3], and Robert K. Jansen[1,3,*]

[1]Department of Integrative Biology, University of Texas at Austin

[2]Department of Biology and Microbiology, South Dakota State University

[3]Genomic and Biotechnology Research Group, Department of Biological Science, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

*Corresponding author: E-mail: jansen@austin.utexas.edu.

Accepted: July 3, 2017

Data deposition: All sequences used in this study have been submitted to NCBI Genbank and accession numbers are reported in supplementary tables S10–12, Supplementary Material online.

Abstract

Geraniaceae have emerged as a model system for investigating the causes and consequences of variation in plastid and mitochondrial genomes. Incredible structural variation in plastid genomes (plastomes) and highly accelerated evolutionary rates have been reported in selected lineages and functional groups of genes in both plastomes and mitochondrial genomes (mitogenomes), and these phenomena have been implicated in cytonuclear incompatibility. Previous organelle genome studies have included limited sampling of *Geranium*, the largest genus in the family with over 400 species. This study reports on rates and patterns of nucleotide substitutions in plastomes and mitogenomes of 17 species of *Geranium* and representatives of other Geraniaceae. As detected across other angiosperms, substitution rates in the plastome are 3.5 times higher than the mitogenome in most *Geranium*. However, in the branch leading to *Geranium brycei*/*Geranium incanum* mitochondrial genes experienced significantly higher d_N and d_S than plastid genes, a pattern that has only been detected in one other angiosperm. Furthermore, rate accelerations differ in the two organelle genomes with plastomes having increased d_N and mitogenomes with increased d_S. In the *Geranium phaeum*/*Geranium reflexum* clade, duplicate copies of *clpP* and *rpoA* genes that experienced asymmetric rate divergence were detected in the single copy region of the plastome. In the case of *rpoA*, the branch leading to *G. phaeum*/*G. reflexum* experienced positive selection or relaxation of purifying selection. Finally, the evolution of acetyl-CoA carboxylase is unusual in Geraniaceae because it is only the second angiosperm family where both prokaryotic and eukaryotic ACCases functionally coexist in the plastid.

Key words: *accD*, *clpP*, *rpoA*, gene duplication, positive selection, accelerated substitution rate, cytonuclear incompatibility.

Introduction

Plant cells harbor two endosymbiotic organelles (plastids and mitochondria) that have retained their own genomes (Lang et al. 1999; Keeling 2010). In the early stages of endosymbiosis the two showed parallel evolutionary trajectories that have shaped their genomes, including genome reduction, massive gene loss, DNA transfer to the host nuclear genome and functional replacement of organelle genes (Timmis et al. 2004). The majority of organelle proteins are nuclear-encoded, synthesized in the cytosol and then transported into the organelles (Timmis et al. 2004; Kleine et al. 2009). Many of these nuclear genes are involved in the maintenance of organelle genome stability (Day and Madesis 2007; Sloan and Taylor 2012), and a number are targeted to both plastids and mitochondria (Maréchal and Brisson 2010). Although many proteins involved in DNA replication, repair, and recombination (DNA-RRR) are common to both plastids and mitochondria (Carrie and Small 2013), their genomes exhibit variability in architecture and rates of nucleotide substitution.

Among angiosperms, the contrast between the two organelle genomes is particularly sharp when considering structural organization, which appears to be much more complicated in mitochondria compared with plastids (Mower et al. 2012; Ruhlman and Jansen 2014). For example, plant mitochondrial genomes (mitogenomes) contain an abundance of noncoding sequence, including repetitive and foreign DNA, and have a dynamic, multipartite organization due to active recombination between repeat regions. In contrast, most plastid genomes (plastomes) typically have a quadripartite structure with an inverted repeat (IR) separated by large and small single copy (LSC and SSC) regions. However, several lineages, such as Campanulaceae, Ericaceae, Fabaceae, Geraniaceae, and Oleaceae, have experienced extreme and rapid divergence in plastome architecture (Chumley et al. 2006; Lee et al. 2007; Haberle et al. 2008; Cai et al. 2008; Blazier et al. 2011; Guisinger et al. 2011; Fajardo et al. 2013; Martínez-Alberola et al. 2013; Weng et al. 2014; Ruhlman et al. 2017).

Plastomes and mitogenomes also exhibit nucleotide substitution rate variation. Plastid substitution rates are generally higher than those in the mitochondria of seed plants (Wolfe et al. 1987; Drouin et al. 2008; Richardson et al. 2013). Examples of mitochondrial-specific rate accelerations have been documented, including gene-specific heterogeneity in synonymous substitution rates within the genome in a number of angiosperm lineages: Geraniaceae, *Ajuga*, *Plantago*, *Silene*, and *Viscum* (Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2012a; Zhu et al. 2014; Skippington et al. 2015). However, only two studies in *Silene* and *Ajuga* (Sloan et al. 2012b; Zhu et al. 2014) have compared substitution rates of plastid and mitochondrial gene sequences across the entire genomes to assess the evolutionary parallels between the organelle genomes.

The functional transfer of plastid and mitochondrial genes to the nucleus is an ongoing evolutionary process (Timmis et al. 2004; Kleine et al. 2009). In angiosperms, the majority of recent, functional relocations has involved the transfer of mitochondrial genes to the nucleus (Adams and Palmer 2003; Liu et al. 2009; Park et al. 2014, 2015a), with multiple independent transfers of only four plastid genes documented (*infA*, *rpl22*, *rpl32*, and *accD*) (Millen et al. 2001; Ueda et al. 2007; Magee et al. 2010; Sabir et al. 2014; Park et al. 2015b). Although the role of selection pressure in intracellular gene transfer (IGT) remains uncertain, rates of organelle mutation accumulation could influence selection for IGT to the nucleus (Brandvain and Wade 2009).

Substitution of a nuclear-encoded protein to complement pseudogenization or loss of an organelle-encoded protein has also been documented. Most gene substitutions involve nuclear gene duplication, with one copy supplying the cytosol and the other supplying the organelle(s). For example, the product of the nuclear *rpl23* gene in *Spinacia oleracea* and *Geranium maderense* substitutes for the missing plastid *rpl23* gene product (Bubunenko et al. 1994; Weng et al. 2016). A different kind of substitution has been documented for *accD* in Brassicaceae (Schulte et al. 1997) and Poaceae (Konishi et al. 1996; Gornicki et al. 1997). With the loss of plastid-encoded *accD*, the eukaryotic monomeric ACC gene product substituted the function of the multisubunit plastid ACCase.

Gene substitutions have occurred through different mechanisms. A protein already targeted to an organelle could be duplicated and the copy could undergo transit peptide evolution. For example, a duplicate copy of nuclear-encoded Rps16 that substitutes for the mitochondrial-encoded protein in *Medicago* and *Populus* was recruited through adaptation of its transit peptide to allow import into the plastid as well (Ueda et al. 2008). Likewise, transit peptide recruitment has facilitated replacement of proteins encoded in the mitochondria with nuclear-encoded proteins of plastid origin, which now serve both compartments. Examples include Rpl10 in Brassicaceae and monocots (Kubo and Arimura 2010) and Rps13 in *Arabidopsis* (Mollier et al. 2002).

Geraniaceae organelle genomes offer an excellent system to study the evolutionary dynamics of structural organization and rates of nucleotide substitution. Sequencing of organelle genomes from this family has provided unprecedented examples of genomic change and highly accelerated nucleotide substitution rates, and improper DNA-RRR machinery has been suggested to cause this syndrome (Parkinson et al. 2005; Guisinger et al. 2008, 2011; Blazier et al. 2011, 2016a; Weng et al. 2014). *Geranium* plastomes exhibit extreme reconfiguration including reduction of the IR, operon disruption, extensive gene duplication and a high frequency of repetitive DNA (Guisinger et al. 2011; Zhang et al. 2016). Exploration of the evolutionary phenomena in *Geranium* mitogenomes revealed extensive horizontal gene transfer (HGT), IGT, and loss of introns and RNA editing sites (Park et al. 2015a). In particular, two *Geranium* species, *Geranium brycei* and *Geranium incanum*, contain considerable foreign organellar DNA from diverse eudicots, including many parasitic plants.

Little is known, however, about nucleotide substitution rates within *Geranium* organelle genomes. To assess rates of nucleotide substitution of protein coding genes and the relationship between *Geranium* plastome and mitogenome evolution, we analyzed 98 plastid and mitochondrial genes of 17 *Geranium* species and representatives of other genera of Geraniaceae. We examined whether substitution rates were elevated and variable within the two organelle genomes and assessed the rate correlation between them. In addition, we characterized *accD* and *clpP* in the plastids of Geraniaceae and correlate structural variation with substitution rates in these genes. We document the evolutionary fate of the plastid-encoded *accD* and discuss the complex evolutionary history of Acetyl-Co-A carboxlase in Geraniaceae.

Materials and Methods

Organelle Gene Sequence Analysis

To compare rates of sequence evolution in plastid and mitochondrial genes of *Geranium*, protein-coding genes from 17 species spanning all three currently recognized subgenera (*Erodioidea*, *Geranium*, and *Robertium*; Aedo et al. 1998) were selected along with the three outgroup taxa *California macrophylla*, *Erodium texanum* and *Monsonia emarginata* from Geraniaceae (supplementary table S1, Supplementary Material online). DNA isolation and Illumina sequencing methods were described in Park et al. (2015a). Plastid protein-coding genes for the 17 *Geranium* species were obtained by de novo assembly of paired-end Illumina reads with Velvet v. 1.2.08 (Zerbino and Birney 2008) using multiple *k*-mer sizes from 69 to 91 at the Texas Advanced Computing Center (TACC). Plastid genes were identified in each draft genome by performing BlastN searches using BLAST+ v.2.2.31 (Camacho et al. 2009) against published protein-coding genes from *G. palmatum* (NC_014573) and related Geraniaceae as query sequences. For mitochondrial protein-coding genes, previously published gene sequences of 17 species of *Geranium* and the outgroups *C. macrophylla*, *E. texanum* and *M. emarginata* were used (Park et al. 2015a). Genbank accession numbers for both mitochondrial and plastid sequences of all species are in supplementary table S1, Supplementary Material online.

Estimation of Substitution Rate Variation

Ninety-eight organelle protein-coding genes were shared by all 20 taxa sampled (supplementary table S2, Supplementary Material online). Individual gene alignments were generated based on "back-translate" with MAFFT v. 7.017 (Katoh et al. 2002) in Geneious R6 v. 6.1.8 (Biomatters Ltd., Auckland, New Zealand, http://www.geneious.com) and all shared organelle protein-coding genes were concatenated into a single alignment. The N-terminal portion of the *clpP* was extremely divergent so only the conserved domain of the gene was used. The 5′ end of *matK* was also removed due to unalignable sequences. In *Geranium phaeum* and *Geranium reflexum*, which contain two distinct copies of *clpP* and *rpoA*, the two copies were arbitrarily named copies 1 and 2 and only the first copy (*G. phaeum1* and *G. reflexum1*) was used for rate analyses. Phylogenetic relationships from the concatenated sequence were estimated using maximum likelihood methods in RAxML v. 7.2.8 (Stamatakis 2006) with the "GTRGAMMA" model under the rapid bootstrap algorithm (1,000 replicates) on TACC and the resulting ML tree used as the constraint tree.

To estimate rates of nucleotide substitution, nonsynonymous (d_N) and synonymous (d_S) rates were calculated in PAML v. 4.8 (Yang 2007) using the codeml option with codon frequencies estimated with the F3 × 4 model. Gapped regions were removed by using "cleandata = 1" option. Rate estimations were performed on the following data sets: 1) each concatenated plastid and mitochondrial data set, 2) concatenated gene sets for the functional groups ATP synthase (*atp*), cytochrome b_6f (*pet*), cytochrome c oxidase (*cox*), cytochrome c maturation (*ccm*), NADH-plastoquinone oxidoreductase (*ndh*), NADH-ubiquinone oxidoreductase (*nad*), photosystem I (*psa*), photosystem II (*psb*), RNA polymerase (*rpo*), and small (*rps*) and large (*rpl*) ribosomal subunits, and 3) all individual genes from both organelles. For *accD* and *clpP*, extensive comparisons of d_N and d_S were performed to investigate the relationship between the rates of nucleotide substitution and structural divergence of across selected angiosperms (supplementary table S3, Supplementary Material online). Five plastid genes (*atpB*, *ndhF*, *psaA*, *psbA*, and *rbcL*) were chosen to generate phylogenetic trees, which were used as a constraint tree for rate comparisons of *clpP* and *accD* genes (supplementary figs. S1 and S2, Supplementary Material online). Only the conserved domains of *clpP* and *accD* were used because other regions of the genes were too divergent to generate reliable nucleotide alignments.

To test for evidence of selection acting on duplicated genes, d_N/d_S was calculated. Likelihood ratio tests (LRTs) were performed to compare the goodness fit of two models, H_0 (d_N/d_S is fixed across entire tree) and H_1 (d_N/d_S allowed to be different in a given clade). Statistical analyses were conducted by using R v. 3.2.0 (R Development Core Team 2015) and the Bonferroni correction for multiple comparisons was applied.

Prediction of Biological Features and Structures in Genes and Proteins

The effective number of codons (EN_C) and GC content at synonymous third codon positions (GC_{3S}) for each organelle gene were calculated using CodonW (http://codonw.sourceforge.net). The NCBI Conserved Domain Database (CDD v. 3.14) was used for conserved domain annotation (Marchler-Bauer et al. 2015). Gene conversion was examined using GENECONV in the Recombination Detection Program (RDP) v. 4.56b (Martin et al. 2015). Previously assembled transcriptomes (Ruhlman et al. 2015; Zhang et al. 2015) for three *Geranium* species and the other genera from Geraniales were used to identify IGT to the nucleus or substitution of a nuclear gene for a plastid gene. Transferred or substituted genes were identified using BLAST+ v.2.2.31 (Camacho et al. 2009) by performing "TBlastN" of the *ACC* sequences from *Arabidopsis thaliana* (*ACC*1, *ACC*2, *accA*, *accB*, and *accC*, NC_003070; *accD*, NC_000932) against the transcriptomes. Organelle transit peptides were predicted using TargetP v.1.1 (Emanuelsson et al. 2007). Phylogenetic analysis was performed on data set of ACC homologs (supplementary table S4, Supplementary Material online). Divergence time

was estimated by Bayesian MCMC analysis using BEAST v.2.1.3 (Bouckaert et al. 2014) following the method described in Park et al. (2015a).

Results

Nucleotide Substitution Rates in *Geranium* Organelle Genes

All 17 *Geranium* species and representative species of the related Geraniaceae genera *California*, *Erodium*, and *Monsonia* shared 72 plastid-encoded and 26 mitochondrial-encoded genes (supplementary table S2, Supplementary Material online). To examine rate variation in organelle genes among *Geranium* species, nonsynonymous (d_N) and synonymous (d_S) substitution rates were estimated using a constraint tree generated by phylogenetic analyses of a concatenated data set of all 98 genes (supplementary fig. S3*A*, Supplementary Material online). Estimates of nucleotide substitution rates in *Geranium* showed that plastid genes generally evolved faster than mitochondrial genes in both d_N and d_S (figs. 1 and 2). Linear regression of plastid and mitochondrial substitution rates for all branches, except for the branch leading to *G. brycei* and *G. incanum*, showed that plastid rates are 3.5 times faster for d_N and 3.4 times faster for d_S than mitochondrial rates (fig. 2*A*). A significant correlation between the organelles was found for both d_N and d_S ($P < 0.0001$, fig. 2*A*). Examination of rate variation in individual organelle genes or gene groups revealed both gene-specific and lineage-specific accelerations in substitution rates within *Geranium* (fig. 1*B* [see supplementary fig. S4, Supplementary Material online, for outlier information], supplementary figs. S5–S7, Supplementary Material online). For example, the average value of d_N for the plastid-encoded *clpP* gene was ~3 times higher than other plastid-encoded genes (fig. 1*B*), and this difference was significant (pairwise Wilcoxon rank-sum test, $P < 0.001$ after Bonferroni correction; supplementary table S5, Supplementary Material online). A significant difference of d_S for mitochondrial-encoded *atp* gene group was detected compared with most other mitochondrial-encoded genes or gene groups (supplementary table S6, Supplementary Material online). The rate of synonymous substitutions in *atp6* and *atp9* were highest among mitochondrial genes across *Geranium* species (supplementary fig. S5, Supplementary Material online). Accelerated d_S was detected for several other mitochondrial genes in three lineages, A2, A4, and B (shaded in gray in supplementary fig. S7, Supplementary Material online). Lineage-specific accelerations of d_N of the plastid genes *cemA*, *clpP*, *matK* and genes encoding ribosomal proteins and RNA polymerase subunits were observed within clades A, A1, A2, and B (supplementary fig. S6, Supplementary Material online).

Comparison of the effective number of codons (EN_C) and GC content at synonymous third codon positions (GC_{3S}) showed that most *Geranium* plastid genes fall below Wright's curve (Wright 1990), reflecting low GC content at synonymous third codon positions (supplementary fig. S8*A*, Supplementary Material online). Mitochondrial genes were consistent with the plastid results except for *matR*, which has GC_{3S} values between 53.2% and 55.4% (supplementary fig. S8*B*, Supplementary Material online). The accelerated mitochondrial *atp9* gene was below Wright's curve (supplementary fig. S8*B*, Supplementary Material online), and showed a significant negative correlation between GC_{3S} and d_S (Pearson's correlation, $R = -0.7993$, $P < 0.001$; supplementary fig. S9*A*, Supplementary Material online) and a minor but not significant increase in EN_C relative to other genes (supplementary fig. S9*B*, Supplementary Material online).

Elevated and Variable Mitochondrial Substitution Rates in a Specific *Geranium* Lineage

Substitution rates of organelle genes in the *G. brycei*/*G. incanum* clade showed the opposite pattern from other *Geranium* species in which the plastid genes had much higher rates than mitochondrial genes (figs. 1*A* and 2). The d_N and d_S of mitochondrial genes from *G. brycei* and *G. incanum* were 2.9 times and 2.4 times higher than plastid genes, respectively, and this difference was significant (Wilcoxon rank-sum test, $P < 0.00001$; fig. 2*B*). LRTs indicated that d_N/d_S values for mitochondrial-encoded *ccmFn*, *matR*, and *nad2* genes were significantly different in the *G. brycei*/*G. incanum* clade compared with other *Geranium* mitochondrial genes (supplementary table S7, Supplementary Material online).

To address whether accelerated nucleotide substitution rates in *G. brycei* and *G. incanum* were affected by HGT (i.e., these two species contain intact native and foreign copies of *atp8*, *cox2*, *nad2*, and *rps1*; Park et al. 2015a), native copies were compared with their orthologs from other *Geranium* species. Significant differences in rate variation were identified in six of the eight comparisons; the two exceptions were d_S of *atp8* and d_N of *cox2* (Wilcoxon rank-sum test, $P < 0.05$; supplementary fig. S10, Supplementary Material online).

Elevated Substitution Rates Following Duplication of Plastid-Encoded *clpP* and *rpoA* Genes

Both *G. phaeum* and *G. reflexum* contained two divergent copies of *clpP* (75.2% and 75.8% nucleotide identity between paralogs) and *rpoA* (88.5% and 87.8% nucleotide identity between paralogs) genes in the plastome single copy (SC) region, and all copies include complete conserved domains (fig. 3). No gene conversion between the paralogs of *clpP* and *rpoA* genes was detected by GENECONV (Martin et al. 2015). Phylogenetic analyses of these copies with other *Geranium* orthologs indicated that the extra copies derived from duplication events that occurred in the common ancestor of *G. phaeum* and *G. reflexum* (fig. 3). To examine

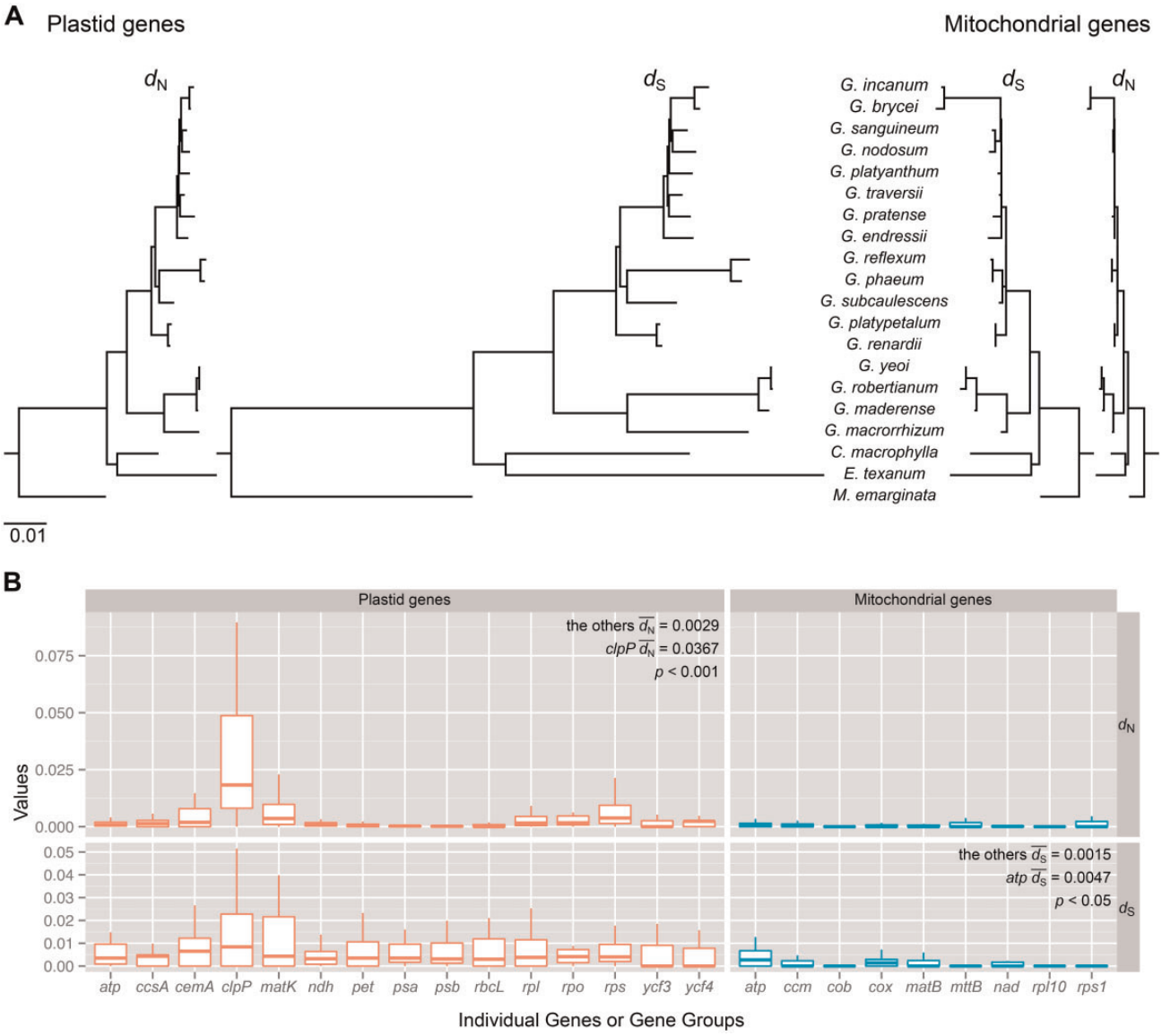

Fig. 1.—Organelle gene divergence among 17 species of *Geranium*. (*A*) Plastid and mitochondrial phylograms of nonsynonymous (d_N) or synonymous (d_S) substitution rates based on 72 plastid and 26 mitochondrial genes, respectively. All trees are drawn to the same scale and scale bar indicates the number of substitutions per site. C., *California*; *E.*, *Erodium*; *G.*, *Geranium*; *M.*, *Monsonia*. (*B*) Box plots of the values of d_N and d_S for *Geranium* plastid (red) and mitochondrial (blue) individual genes or functional groups of genes (see supplementary table S2, Supplementary Material online). The box represents values between quartiles, solid lines extend to minimum and maximum values, horizontal lines in boxes show median values but outliers were eliminated (see supplementary fig. S4 for outlier information, Supplementary Material online). Significance of fit was evaluated by pairwise Wilcoxon rank sum tests in the R package.

sequence evolution of the duplicated genes, pairwise comparisons of d_N and d_S among their homologs were computed using a constraint tree (supplementary fig. S3*B*, Supplementary Material online). Branch lengths of d_N for *clpP* for *G. phaeum1* and *G. reflexum1* were 2.0 times and 2.4 times longer than their paralogs *G. phaeum2* and *G. reflexum2*, whereas d_S branch lengths for *G. phaeum2* and *G. reflexum2* were 7.5 times and 7.4 times longer than their paralogs, respectively (fig. 3*A*). Fourteen branches with $d_N/d_S > 1$ were detected (shown in red font in supplementary table S8, Supplementary Material online) but LRTs showed that none of them were significantly different ($P = 1.00$ after Bonferroni correction, supplementary table S8, Supplementary Material online). d_N branches for *G. phaeum2* and *G. reflexum2 rpoA* were 15.4 times and 9.6 times longer than their paralogs *G. phaeum1* and *G. reflexum1* and d_S branches were 21.5 times and 7.0 times longer, respectively (fig. 3*B*). Six branches with $d_N/d_S > 1$ were detected (shown in

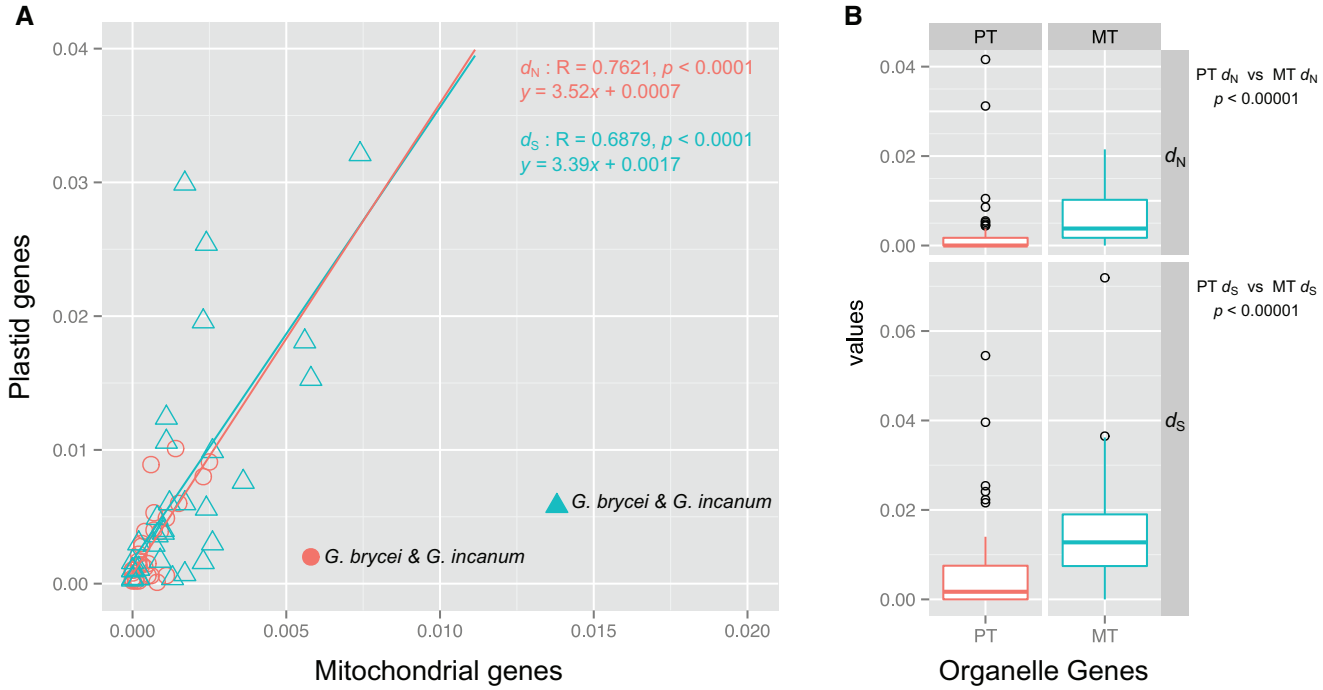

Fig. 2.—Plastid and mitochondrial sequence divergence among *Geranium* species. (*A*) Correlation of nonsynonymous (d_N, red circles) and synonymous (d_S, blue triangles) substitution rates between each shared branch in the phylogram shown in supplementary fig. S3*A*, Supplementary Material online. Linear regression analyses included all branches except for the branch leading to *G. brycei* and *G. incanum* (closed circle and triangle). (*B*) Boxplot distribution of the d_N and d_S values for all individual organelle genes from *G. brycei* and *G. incanum*. The box represents values between quartiles, solid lines extend to minimum and maximum values, outliers are shown as circles and horizontal lines in boxes show median values. Significance of fit was evaluated by Pearson's correlation and pairwise Wilcoxon rank sum tests in the R package. PT, plastid; MT, mitochondrion.

red font in supplementary table S9, Supplementary Material online). LRTs showed that only one branch, leading to *G. phaeum* and *G. reflexum*, had a significantly different d_N/d_S for *rpoA* ($P < 0.001$ after Bonferroni correction, supplementary table S9, supplementary fig. S3*b*, Supplementary Material online), indicating positive selection ($d_N/d_S = 4.7$) in the ancestor of these two species. Lineage-specific rate accelerations of the other three plastid-encoded polymerase (PEP) genes *rpoB*, *rpoC1*, and *rpoC2* were observed on the branch leading to *G. phaeum* and *G. reflexum* compared with their sister species *Geranium subcaulescens* (supplementary fig. S11, Supplementary Material online), although among the four PEP genes only *rpoB* had a significantly different d_N/d_S in LRTs ($P < 0.05$ after Bonferroni correction; supplementary table S10, Supplementary Material online).

Structural Evolution of Plastid-Encoded *clpP* Gene

Geranium and *Monsonia* plastomes contain *clpP*-like open reading frames (ORFs) that are intronless but retain conserved domains (supplementary fig. S12*A*, Supplementary Material online). The first 71 nucleotides of the ORFs is a highly divergent region that represents the former first exon (43–55% nucleotide sequence identity compared with *Nicotiana tabacum*, supplementary fig. S13, Supplementary Material online). The C-terminal portions of the ORFs included the conserved domain and the remaining sequence of the *Geranium* and *Monsonia clpP*-like ORFs was highly variable, ranging from 252 to 488 bp in length (supplementary fig. S12*A*, Supplementary Material online). The conserved C to U editing site of the *clpP* gene, which was characterized by Chateigner-Boutin et al. (2008), was also detected in the 3′ end of the sequence (supplementary fig. S14, Supplementary Material online). To investigate a correlation between structural evolution of the plastid-encoded *clpP* gene and substitution rates, d_N and d_S of *clpP* for representative angiosperms and Geraniales were calculated using sequences from the conserved domain. Multiple lineage-specific accelerations of d_N and d_S were detected within the groups that experienced the intron loss in *clpP* (supplementary fig. S12*B*, Supplementary Material online). The d_N and d_S values were more strongly correlated in the groups that contain structural changes in *clpP* (Pearson's correlation, $R = 0.8985$, $P < 0.0001$) than in other angiosperms ($R = 0.7704$, $P < 0.0001$), and Wilcoxon rank-sum test showed that d_N is significantly different between these groups (d_N, $P < 0.001$; d_S, $P = 0.0708$; supplementary fig. S15*A*, Supplementary Material online).

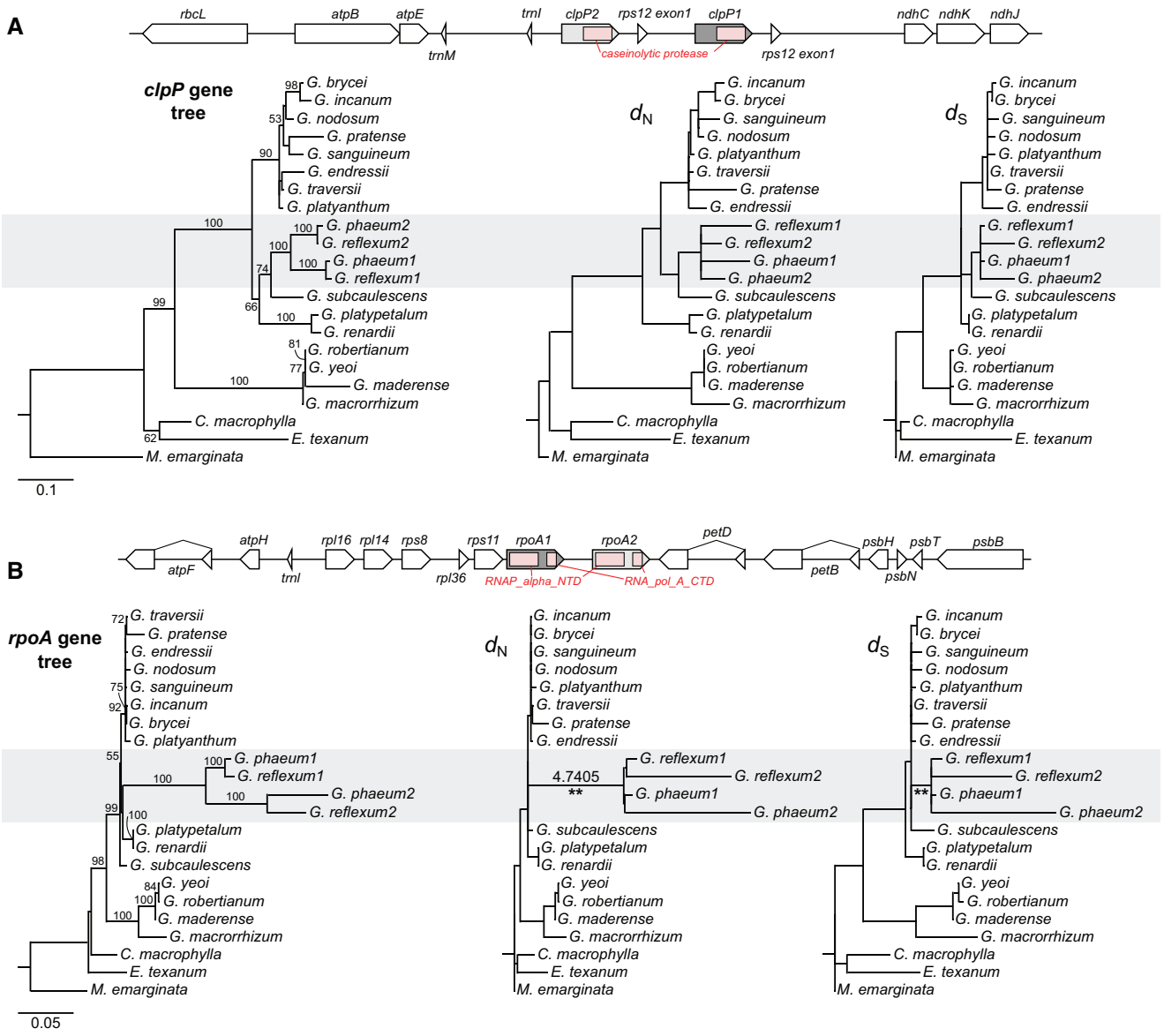

Fig. 3.—Duplication of plastid-encoded *clpP* and *rpoA* genes in *G. phaeum* and *G. reflexum*. (*A*) Schematic diagram of the genomic regions surrounding the plastid *clpP* homologs from *G. phaeum* and *G. reflexum*. Pink box indicates conserved domain (caseinolytic protease). Maximum likelihood (ML) gene tree (left) based on *clpP* nucleotide sequences. ML phylograms used the constraint tree in supplementary fig. S3*B*, Supplementary Material online to show nonsynonymous (d_N) or synonymous (d_S) substitution rates of *clpP* (middle, right, respectively). (*B*) Schematic diagram of the genomic regions surrounding the *rpoA* homologs from *G. phaeum* and *G. reflexum* plastomes. Pink box indicates two conserved domains (RNA_alpha_NTD and RNA_pol_A_CTD). ML gene tree (left) based on *rpoA* nucleotide sequences. ML phylograms used the constraint tree in supplementary fig. S3*B*, Supplementary Material online to show d_N and d_S of *rpoA* (middle, right, respectively). Branches with significantly higher d_N/d_S (4.7405) detected by likelihood ratio test (LRT) are marked with two asterisks ($P < 0.001$ after Bonferroni correction; supplementary table S9, Supplementary Material online). Clade A2 within *Geranium* (supplementary fig. S3*A*, Supplementary Material online) that has duplication events is highlighted in gray. The numbers "1" and "2" after each species in the phylograms represent paralogs for *clpP* and *rpoA*. Bootstrap support values >50% are shown on the branches.

LRTs showed that the branches leading to clades A2 (branch 18), A3 (branch 24), B (branch 32), and *California* (branch 34) had significantly different d_N/d_S, and that the branches within Geraniaceae including *Geranium* also had significantly different d_N/d_S (indicated in red font in supplementary table S11, Supplementary Material online).

Interruption of Plastid-Encoded *accD*

Plastomes of all *Geranium* species examined and *C. macrophylla* have an expanded acetyl-CoA carboxylase subunit β (*accD*) ORF compared with *N. tabacum* (1,539 bp), ranging in size from 2,748 bp in *C. macrophylla* to 3,576 bp in *G. brycei*. The conserved domain of *accD* is split into two portions

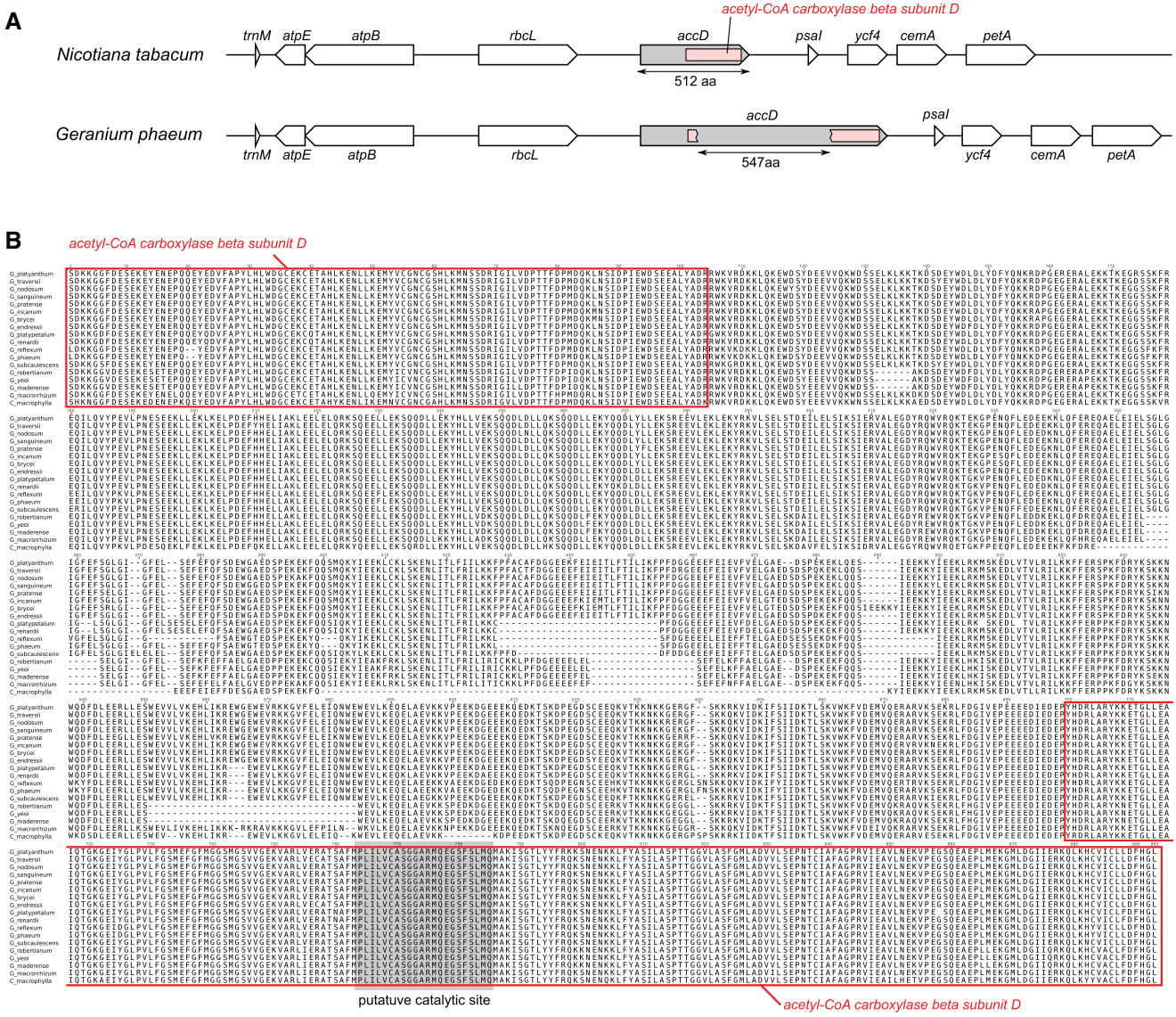

Fig. 4.—Characterization of the plastid-encoded acetyl-CoA carboxylase beta subunit (*accD*) genes of *Geranium*. (*A*) Schematic diagram of the genomic region surrounding *accD* in the *G. phaeum* plastome compared with *Nicotiana tabacum*. Boxes inside *accD* gene (gray) indicate the conserved domain (acetyl-CoA carboxylase beta subunit; pink). (*B*) Amino acid sequence alignment of plastid-encoded acetyl-CoA carboxylase beta subunit conserved domain including the insertion regions from 17 *Geranium* species and *California macrophylla*. Red boxes indicate the conserved domain of acetyl-CoA carboxylase beta subunit and shaded gray box indicates well-known putative catalytic site (Lee et al. 2004).

by a large insertion of 547 amino acids (fig. 4). Interruption of the conserved domain was also detected in *accD*-like ORFs from three other genera of Geraniales (i.e., *Pelargonium alternans*, NC_023261; *Pelargonium cotyledonis*, KM459516; and *Viviania marifolia*, NC_023259).

To investigate the effect of the insertion on nucleotide substitution rates of *accD*, d_N, and d_S were calculated for Geraniaceae and representative angiosperms using only the conserved domain sequences. Higher values of d_N and d_S were identified in *accD* genes on branches leading to Geraniaceae and *Viviania* compared with other angiosperms (supplementary fig. S16, Supplementary Material online). Both d_N and d_S values were more strongly correlated in the divergent groups with the interrupted *accD* conserved domain (Pearson's correlation, $R = 0.9699$, $P = 0.0301$) than in angiosperms with intact domains ($R = 0.6598$, $P < 0.0001$), and Wilcoxon rank-sum test showed that branch values of d_N and d_S were significantly different between the two groups (d_N, $P < 0.001$; d_S, $P < 0.001$; supplementary fig. S15*B*, Supplementary Material online).

Coordination of Acetyl-CoA Carboxylase Gene Function between Plastid and Nuclear Genomes

Examination of transcriptomes (Ruhlman et al. 2015; Zhang et al. 2015) provided evidence that all Geraniaceae except *Hypseocharis* contained two transcripts for the nuclear, eukaryotic acetyl-CoA carboxylase (*ACC*) gene (*ACC*1 and *ACC*2; fig. 5). One of these copies (arbitrarily referred to as *ACC*2) has an extended sequence at the 5′ end predicted by TargetP to be a plastid transit peptide in the product (supplementary table S12, Supplementary Material online). Three nuclear-encoded transcripts (*accA*, *accB*, and *accC*) for plastid-targeted proteins similar to *Arabidopsis* homologs were also detected across Geraniaceae (fig. 5*B*, supplementary table S12, Supplementary Material online). In contrast, an *accD* gene transfer from the plastid to the nucleus was identified in the *Hypseocharis* and *Monsonia* transcriptomes and the predicted ORF contained a transit peptide sequence (supplementary fig. S17, supplementary table S12, Supplementary Material online). The target peptides were 75 and 76 amino acids long in *Hypseocharis* and *Monsonia*, respectively, and their alignment length was 78 amino acids long with a sequence identity of 19.2% (supplementary fig. S17, Supplementary Material online). The nuclear-encoded *accD* of these two species contains a large insertion that interrupted the conserved domains (supplementary fig. S17*A*, Supplementary Material online), similar to the plastid-encoded *accD* in Geraniaceae (fig. 4).

Phylogenetic analyses of nuclear-encoded *ACC* homologs from Geraniaceae showed that each of the orthologs grouped together (fig. 5*C*), suggesting that the duplication event occurred in the common ancestor of Geraniaceae excluding *Hypseocharis* (fig. 5*D*). Pairwise analysis of d_N and d_S for *ACC* homologs showed that the *ACC*1 ortholog had elevated d_S and lower d_N values (supplementary fig. S18, Supplementary Material online) and Wilcoxon rank-sum test showed that d_N is significantly different between *ACC* orthologs ($P = 0.01017$, supplementary fig. S18*B*, Supplementary Material online). On the basis of the organismal divergence time in Geraniaceae (Park et al. 2015a), the interruption of the conserved domains of plastid *accD* occurred between 61 Ma and 95 Ma (fig. 5*D*). Divergence time estimates suggest that the nuclear *ACC* gene duplication occurred 51 Ma in Geraniaceae excluding *Hypseocharis* (fig. 5*D*; supplementary fig. S19, Supplementary Material online for more detailed divergence time estimates).

Discussion

Organelle sequencing across the angiosperm family Geraniaceae revealed accelerated substitution rates and rapid structural evolution in plastomes (Guisinger et al. 2008, 2011; Weng et al. 2014, 2017; Blazier et al. 2016b; Ruhlman et al. 2017) as well as highly accelerated substitution rates in mitogenomes (Parkinson et al. 2005, Bakker et al. 2006, Weng et al. 2012). However, because only one to three species of the largest genus *Geranium* with >400 species (Aedo et al. 1998) were included in previous studies, little is known about the evolution of their organelle genomes. In this study, we analyzed all shared protein coding genes from both plastomes and mitogenomes of 17 *Geranium* species and closely related species in the family to compare patterns of evolutionary rates in both organelles. We identified several unusual phenomena, including gene-specific and lineage-specific rate accelerations with one lineage showing significantly higher rates in mitochondrial genes than plastid genes, different rates of nucleotide substitutions in duplicate copies of plastid genes, a correlation between accelerated rates and structural variation in the plastid-encoded *clpP* gene and a complex evolutionary history of gene interruption, transfer and duplication of both the eukaryotic and prokaryotic acetyl-CoA carboxylase genes. Our discussion will focus on these unusual phenomena, which highlight the dynamic evolutionary history of *Geranium* organellar genomes.

Substitution Rate Variation in *Geranium* Organelle Genomes

Extensive variation in organelle substitution rates was detected in *Geranium*. Our study identified gene-specific patterns of rate variation in organelle protein coding genes (figs. 1*B*, supplementary figs. S6 and S7, Supplementary Material online). Among plastid genes, several had accelerated d_N, including *cemA*, *clpP*, *matK*, and genes encoding large and small subunit ribosomal proteins and PEP subunits. These are the same plastid genes/functional groups that had significantly accelerated d_N in a family-wide comparison (Guisinger et al. 2008). The most extreme case was *clpP*, which had significantly higher d_N with a nearly 13-fold increase. In the case of mitochondrial-encoded genes, rate accelerations were restricted to synonymous sites with members of the *atp* functional group having a significantly higher d_S with a 3-fold increase. This pattern of accelerated d_N in plastid genes and d_S in mitochondrial genes in *Geranium* is similar to that previously reported in *Pelargonium*, another large genus of Geraniaceae (Weng et al. 2012).

Lineage-specific accelerated rates of nucleotide substitutions occur in both genomes, which are especially evident from the longer branches in clades A2 and A4 that include *G. incanum*/G. *brycei* and *G. phaeum*/*G. reflexum* (fig. 1*A*, supplementary figs. S6 and S7, Supplementary Material online). The rate accelerations in mitochondrial and plastid genes occur in different clades indicating that different lineages are affected by rate variation in the two genomes. Previous studies have shown that substitution rate is generally three times slower in mitogenomes than in plastomes (Wolfe et al. 1987; Drouin et al. 2008). Some lineages (Geraniaceae, *Ajuga*, *Plantago*, and *Silene*) have experienced a substantial rate increase in the mitogenome (Cho et al. 2004;

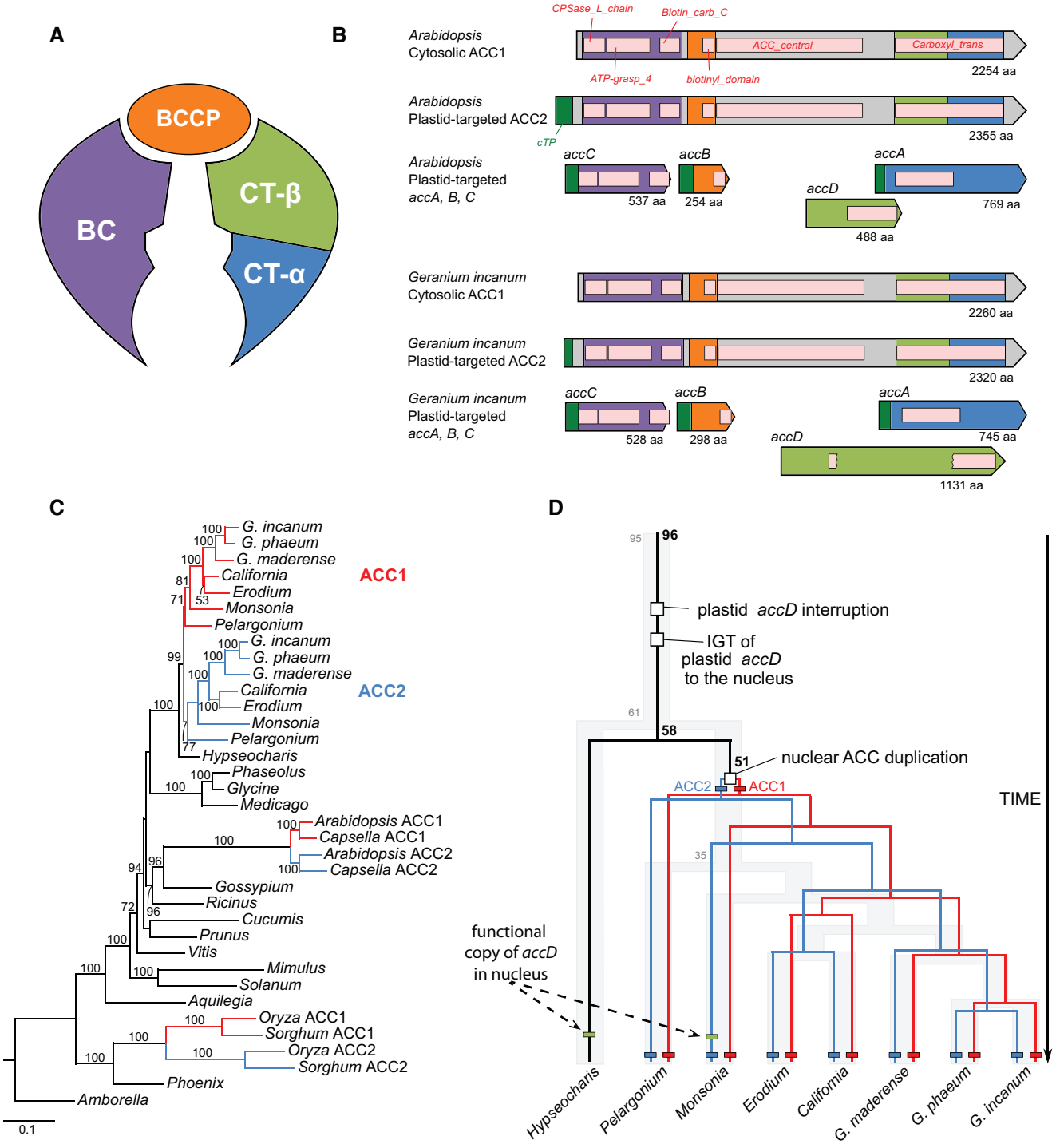

Fig. 5.—Evolutionary implications of acetyl-CoA carboxylase variation in Geraniaceae. (*A*) Schematic diagram of Acetyl-coenzyme A carboxylase (ACCase) composed of four subunits. BC = Biotin carboxylase (purple), BCCP = Biotin carboxyl carrier protein (orange), CT = Carboxyl transferase (alpha: blue, beta: green). (*B*) Architecture of ACCase from *Geranium incanum* compared with *Arabidopsis thaliana*. Colors indicate subunits of ACCase corresponding to the schematic diagram in (*A*). Pink boxes indicate the conserved domains of ACCase. Dark green boxes in N-terminus indicate a transit peptide. (*C*) Maximum likelihood (ML) gene tree based on nuclear-encoded *ACC* homologs among angiosperms. The numbers "1" and "2" after each species in the phylogram represents paralogs for *ACC* (red and blue). Bootstrap support values >50% are shown on the branches. (*D*) Summary of gene duplication and intracellular gene transfer (IGT) events with divergence times. See supplementary fig. S19, Supplementary Material online for more detailed divergence time estimates for Geraniaceae. The evolutionary events are marked with squares and are shown in colored boxes. Numbers at nodes indicate divergence time estimates in Ma. Background tree with numbers (gray) adapted from Park et al. (2015a). Color branches correspond to each ortholog (red, *ACC1*; blue, *ACC2*).

Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2008; Zhu et al. 2014). *Geranium brycei* and *G. incanum* exhibit accelerated rates in mitochondrial genes compared with their plastid counterparts (fig. 2*B*). This pattern contrasts with a previous study that showed that in most angiosperms nonsynonymous and synonymous substitution rates are 1.8 and 4.9 times faster in the plastome than the mitogenome, respectively (Zhu et al. 2014). Rate comparisons in *Ajuga* by Zhu et al. (2014) showed a similar pattern of higher rates of nucleotide substitutions in the mitogenome than the plastome. They suggested that localized hypermutation or mRNA retroprocessing was likely responsible for this phenomenon. A previous study in *Geranium* (Park et al. 2015a) showed that all examined *Geranium* mitochondrial genes had similar editing sites, suggesting that the accelerated rates in *G. brycei* and *G. incanum* genes are not caused by RNA editing frequency. The mitogenomes of these two species have acquired a number of foreign organelle genes (*atp8*, *cox2*, *nad2*, and *rps1*) via HGT (Park et al. 2015a). It is possible that the substitution rates correlate with gene conversion between the native and foreign copies, however, gene conversion was only found in the *cox2* gene of *G. brycei* (Park et al. 2015a). The mitochondrial rate accelerations in *G. brycei* and *G. incanum* do not appear to be a gene-specific phenomenon because many genes show higher rates in these species (supplementary figs. S5 and S7, Supplementary Material online). Furthermore, accelerated rates of these same genes have occurred in several other *Geranium* lineages. One possible cause for this pattern in *Geranium* is dysfunction of the DNA-RRR machinery, which has been suggested previously in groups with accelerated mitochondrial rates (Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2012a). The correlation between nucleotide substitution rates and structural evolution in *Geranium*, particularly where mitochondrial-targeted DNA-RRR genes are involved, should be explored to test this hypothesis.

Rate Variation among Duplicated Plastid Genes

Most plastid gene duplications are caused by IR expansion (Jansen and Ruhlman 2012), and duplicate copies are homogenized by gene conversion during homologous recombination between the repeats. However, partial or complete gene duplications have been documented in SC regions in a few lineages including Geraniaceae (Guisinger et al. 2011; Ruhlman et al. 2017), Fabaceae (Cai et al. 2008; Sabir et al. 2014; Sveinsson and Cronk 2014), Campanulaceae (Haberle et al. 2008), and Caryophyllaceae (Erixon and Oxelman 2008). *Geranium phaeum* and *G. reflexum* plastomes each contain two divergent copies of *clpP* and *rpoA* genes in the LSC region (fig. 3). The phylogenetic placement of the duplicate copies indicates that *clpP* and *rpoA* were duplicated in the common ancestor of two species (fig. 3). On the basis of previously published divergence times (Park et al. 2015a), the duplication events occurred between 2.0 and 8.7 Ma.

Elevated substitution rates and positive selection or relaxed functional constraints are often associated with gene duplication (Ohta 1994). Duplicate copies of *clpP* and *rpoA* in *G. phaeum* and *G. reflexum* have experienced asymmetric rate divergence (fig. 3). Analysis of d_N/d_S for *clpP* did not detect positive selection on the branch leading to *G. phaeum* and *G. reflexum* (supplementary table S8, Supplementary Material online). In the case of *rpoA*, however, the branch leading to *G. phaeum*/*G. reflexum* had a d_N/d_S significantly >1 after Bonferroni-correction, indicating positive selection or relaxation of purifying selection acting on the gene in the ancestor of these two species (fig. 3*B*, supplementary table S9, Supplementary Material online). In *Campanulastrum americanum* (Campanulaceae) positive selection was also detected in *clpP* and some populations had a partial duplication of the first exon (Barnard-Kubow et al. 2014). The authors suggested that accelerated evolution in the plastome may contribute to cytonuclear incompatibility. This hypothesis was supported in a recent study of rapidly evolving plastid-nuclear protein complexes in *Silene* (Rockenbach et al. 2016).

Duplications of genes in the SC regions have been reported for several genes in unrelated groups of angiosperms and in some cases this involves the same two genes, *clpP* and *rpoA*, that are duplicated in *Geranium*. In two genera of the tribe Sileneae (*Silene* and *Lychnis*, Caryophyllaceae), high substitution rates were accompanied by gene duplications, loss of introns and positive selection in *clpP* (Erixon and Oxelman 2008). A similar pattern was observed in the *G. phaeum*/*G. reflexum* clade in which the tandemly duplicated *clpP* genes lack both introns. Another example of gene duplication in SC regions was recently reported for *rpl20* in *M. emarginata* in Geraniaceae (Ruhlman et al. 2017). In this case, there are five copies included in a larger repeat in this IR-lacking species that have very high nucleotide sequence identity (96%) and there is GC biased gene conversion between the copies. Furthermore, replication dependent recombination between the *rpl20* repeats produces alternative plastome arrangements. This species also has duplicated, intronless copies of *clpP*.

Duplication of *rpoA* has also been documented in *Pelargonium* but in this case duplicate copies reside in the IR (Blazier et al. 2016a; Weng et al. 2017). Up to six copies of *rpoA* are present in each IR in some species and although rates of nucleotide substitutions are very high and levels of amino acid sequence identity is as low as 19% compared with the outgroup *Eucalyptus*, this gene is experiencing purifying selection and functional domains remain intact, suggesting that the *rpoA* is functional in *Pelargonium*. In this case, the multiple copies are homogenized via gene conversion.

Evolution of the ATP-Dependent Clp Protease Proteolytic Subunit

In plant cells, the ATP-dependent Clp protease proteolytic subunit gene (*clpP*) that contains two introns is encoded in

plastids and is involved in protein metabolism (Peltier et al. 2004; Stanne et al. 2009). Several angiosperm lineages have experienced loss or structural evolution (i.e., lack of first intron or both introns) of the plastid-encoded *clpP* (Erixon and Oxelman 2008; Martinez-Alberola et al. 2013; Sloan et al. 2014; Barnard-Kubow et al. 2014; Wang et al. 2016; Ruhlman et al. 2017). We identified an extremely divergent, intronless *clpP* gene that was previously annotated as a pseudogene or loss in *Geranium* and *Monsonia* plastomes (Guisinger et al. 2011; Weng et al. 2014). Intron loss may be explained by direct genomic deletion, exonization of introns, retroprocessing and HGT of an intronless version of a gene followed by gene conversion with native copies (reviewed in Hepburn et al. 2012). The most parsimonious interpretation for *clpP* intron loss in Geraniaceae involves the loss of the first intron in the common ancestor of *Geranium*, *California*, *Erodium*, and *Monsonia* followed by loss of the second intron independently in *Geranium* and *Monsonia*. The mechanism of the first intron loss is indiscernible from the present data, although two possible mechanisms are exonization of the intron or intron loss followed by an insertion. However, the second intron is likely to have been lost in *Geranium* and *Monsonia* through retroprocessing, which involves genetic integration of an intronless cDNA intermediate produced by reverse transcription of a spliced mRNA. This explanation is supported by the fact that *Geranium* and *Monsonia clpP* genes show the parallel loss of introns and a RNA editing site (fig. 3*A*, supplementary fig. S14, Supplementary Material online).

Extensive analyses of the plastid-encoded *clpP* gene across angiosperms revealed that structural changes in *clpP* genes are correlated with elevated substitution rates in several lineages (supplementary fig. S12, Supplementary Material online). In Geraniaceae lineages, high substitution rates were accompanied by positive or relaxed purifying selection. Our results also showed that increased nonsynonymous substitution rates in *clpP* were recurrent in the family. A similar situation was observed in some *Silene* lineages (Erixon and Oxelman 2008; Sloan et al. 2014) where *clpP* genes had elevated d_N and d_S associated with the loss of the introns. Gene-specific rate accelerations could be explained by localized hypermutation (Sloan et al. 2009; Magee et al. 2010) or mutagenic retroprocessing (Parkinson et al. 2005; Zhu et al. 2014). Consistent with the mechanism of the intron losses, mutagenic retroprocessing is favored as a potential explanation for *clpP* in *Geranium* and *Monsonia*.

Evolution of Acetyl-CoA Carboxylase

Acetyl-CoA carboxylase (ACCase) is crucial to fatty acid metabolism in most living organisms (Wakil et al. 1983; Kode et al. 2005) and two types of ACCase (i.e., eukaryotic and prokaryotic) are known (Wakil et al. 1983). The eukaryotic ACCase is a single multifunctional polypeptide chain encoded in the nucleus that operates in the cytoplasm. The prokaryotic ACCase comprises four subunits: The acetyl-CoA carboxylase subunit α (*accA*), the biotin carboxyl carrier protein subunit (*accB*), the biotin carboxylase subunit (*accC*), all nuclear-encoded, and the acetyl-CoA carboxylase subunit β (*accD*), which is encoded in the plastid (fig. 5*A* and *B*). Plant cells contain both types of ACCase, the eukaryotic type and the prokaryotic type, and the latter assembles in the plastid subsequent to the import of the three nuclear-encoded subunits. The plastid-encoded *accD* gene has been lost multiple times across angiosperms (Ruhlman and Jansen 2014), however, because the plastid ACCase is essential a functional replacement must occur. Several studies have shown that the missing plastid-encoded *accD* gene has been transferred to the nucleus (Rousseau-Gueutin et al. 2013; Sabir et al. 2014; Li et al. 2016), or alternatively that the prokaryotic ACCase has been replaced by the nuclear-encoded eukaryotic type (Konishi et al. 1996; Gornicki et al. 1997). Substitution of the duplicated eukaryotic ACCase for the plastid ACCase has been documented in *Brassica* (Schulte et al. 1997) and Poaceae (Konishi et al. 1996; Gornicki et al. 1997). Co-existence of the plastid-targeted eukaryotic ACCase and the plastid-encoded, prokaryotic ACCase has been documented in *Arabidopsis* (Babiychuk et al. 2011) and likely occurs more widely in Brassicaceae (Rousseau-Gueutin et al. 2013). According to Babiychuk et al. (2011) the prokaryotic ACCase, including the subunit encoded by *accD*, provides most of the ACCase activity.

The evolutionary history of ACCase in Geraniaceae is complex (fig. 5*D*). All species examined have copies of the three nuclear encoded prokaryotic subunits (A, B, C; supplementary table S12, Supplementary Material online) and the plastid-encoded *accD* with a large insertion that interrupts the conserved domain. The acetyl-CoA carboxylase subunit β, encoded by plastid *accD*, is likely functional in Geraniaceae as it retains the putative catalytic site at the C-terminus (Rousseau-Gueutin et al. 2013). Large insertions in *accD* have been documented in other lineages including cupressophyte conifer clade (Hirao et al. 2008; Yi et al. 2013) and the legume *Medicago truncatula* (Gurdon and Maliga 2014). Like Geraniaceae, in both of these cases *accD* preserves an intact reading frame suggesting functionality. The conserved domain of *accD* likely was interrupted in the common ancestor of Geraniaceae. Two examined Geraniaceae, *Hypeocharis* and *Monsonia*, have a plastid-to-nucleus functional transfer of the formerly plastid-encoded *accD* gene, and this likely also occurred in the ancestor of the family because the nuclear copies have the interrupted conserved domain (supplementary fig. S17, Supplementary Material online). Although transfer of DNA from the plastid to the nuclear genome is frequent (Noutsos et al. 2005), functionalization of transferred sequences is far less common. The low level of amino acid identity (19.2%) indicates that the transit peptides have diverse origins within the nuclear genomes of each species. The Geraniaceae

is only the second angiosperm family in which a plastid-targeted eukaryotic ACCase and the plastid-encoded, prokaryotic ACCase are known to co-exist.

Organelle genomes of *Geranium* exhibit both lineage-specific and gene-specific rate accelerations where mitogenomes have increased rates in synonymous substitutions whereas nonsynonymous substitutions are accelerated in plastomes. The two most unusual phenomena in *Geranium* organelle genomes concern substitution rates in the mitogenome and the evolutionary history of acetyl-CoA carboxylase. *Geranium* provides only the second example among angiosperms in which rates are significantly higher in the mitogenome than the plastome. Evolution of ACCase is complex in the Geraniaceae because of the interruption and transfer of the prokaryotic, plastid encoded *accD* subunit and the duplication of the eukaryotic subunit with one copy now targeted to the plastid, making it only the second example among angiosperms of the co-existence a plastid-targeted eukaryotic ACCase and the plastid-encoded, prokaryotic ACCase. Finally, our results support the suggestion that accelerated evolution in the plastome in angiosperm lineages like Geraniaceae may contribute to cytonuclear incompatibility (Barnard-Kubow et al. 2014; Zhang et al. 2015; Rockenbach et al. 2016; Ruhlman et al. 2017).

Supplementary Material

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

Acknowledgments

Support was provided by the National Science Foundation (IOS-1027259 to R.J.K. and T.A.R.) and from the President of King Abdulaziz University, Prof Dr Abdulrahman O. Alyoubi (to J.S.M.S., H.H.H., and R.K.J.). The authors thank Robin Parer at www.Geraniaceae.com for providing living material of Geraniaceae species. We also thank the Genome Sequencing and Analysis Facility at the University of Texas at Austin for performing the Illumina sequencing, the Texas Advanced Computing Center (TACC) at UT Austin for access to supercomputers, the Plant Resources Center (TEX-LL) at UT Austin for housing voucher specimens and two anonymous reviewers for valuable suggestions on an earlier version of the manuscript.

Literature Cited

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogen Evol. 29:380–395.

Aedo C, Garmenia FM, Pando F. 1998. World checklist of *Geranium* L. (Geraniaceae). Anal Jard Bot Madrid. 56:211–252.

Babiychuk E, et al. 2011. Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. Proc Natl Acad Sci U S A. 108:6674–6679.

Bakker FT, Breman F, Merckx V. 2006. DNA sequence evolution in fast evolving mitochondrial DNA *nad1* exons in Geraniaceae and Plantaginaceae. Taxon 55:887–896.

Barnard-Kubow KB, Sloan DB, Galloway LF. 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.

Blazier JC, Guisinger MM, Jansen RK. 2011. Recent loss of plastid-encoded *ndh* genes within *Erodium* (Geraniaceae). Plant Mol Biol. 76:263–272.

Blazier JC, et al. 2016a. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci Rep. 6:24595.

Blazier JC, et al. 2016b. Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann Bot. 117:1209–1220.

Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10:e1003537.

Brandvain Y, Wade MJ. 2009. The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. Genetics 182:1129–1139.

Bubunenko MG, Schmidt J, Subramanian AR. 1994. Protein substitution in chloroplast ribosome evolution: a eukaryotic cytosolic protein has replaced its organelle homologue (L23) in spinach. J Mol Biol. 240:28–41.

Cai Z, et al. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. J Mol Evol. 67:696–704.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinform. 10:421.

Carrie C, Small I. 2013. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochim Biophys Acta. 1833:253–259.

Chateigner-Boutin AL, et al. 2008. CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. Plant J. 56:590–602.

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.

Chumley TW, et al. 2006. The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 23:2175–2190.

Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: Bock R, editor. Cell and molecular biology of plastids. Berlin (Germany): Springer. p. 65–119.

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 49:827–831.

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.

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

Fajardo D, et al. 2013. Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing. Tree Genet Genomes. 9:489–498.

Gornicki P, et al. 1997. Plastid-localized acetyl-CoA carboxylase of bread wheat is encoded by a single gene on each of the three ancestral chromosome sets. Proc Natl Acad Sci U S A. 94:14179–14184.

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.

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.

Gurdon C, Maliga P. 2014. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *accD* coding region in *Medicago truncatula*. DNA Res. 21:417–427.
Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. J Mol Evol. 66:350–361.
Hepburn NJ, Schmidt DW, Mower JP. 2012. Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. Mol Biol Evol. 29:3111–3120.
Hirao T, Watanabe A, Kurita M, Kondo T, Takata K. 2008. Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. BMC Plant Biol. 8:70.
Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. Genomics of chloroplasts and mitochondria. Netherlands: Springer. pp. 103–126
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Keeling PJ. 2010. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 365:729–748.
Kleine T, Maier UG, Leister D. 2009. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 60:115–138.
Kode V, Mudd EA, Iamtham S, Day A. 2005. The tobacco plastid *accD* gene is essential and is required for leaf development. Plant J. 44:237–244.
Konishi T, Shinohara K, Yamada K, Sasaki Y. 1996. Acetyl-CoA carboxylase in higher plants: most plants other than Gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol. 37:117–122.
Kubo N, Arimura S. 2010. Discovery of the *rpl10* gene in diverse plant mitochondrial genomes and its probable replacement by the nuclear gene for chloroplast RPL10 in two lineages of angiosperms. DNA Res. 17:1–9.
Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annu Rev Genet. 33:351–397.
Lee HL, Jansen RK, Chumley TW, Kim KJ. 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. Mol Biol Evol. 24:1161–1180.
Lee SS, et al. 2004. Characterization of the plastid-encoded carboxyltransferase subunit (*accD*) gene of potato. Mol Cells. 17:422–429.
Li J, et al. 2016. Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of Sciadopityaceae. Sci Rep. 6:20934.
Liu SL, Zhuang Y, Zhang P, Adams KL. 2009. Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. Mol Biol Evol. 26:875–891.
Magee AM, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 20:1700–1710.
Marchler-Bauer A, et al. 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43:D222–D226.
Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1:vev003.
Martínez-Alberola F, et al. 2013. Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. PLoS ONE. 8:e79685.
Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. New Phytol. 186:299–317.
Millen RS, et al. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell. 13:645–658.
Mollier P, Hoffmann B, Debast C, Small I. 2002. The gene encoding *Arabidopsis thaliana* mitochondrial ribosomal protein S13 is a recent duplication of the gene encoding plastid S13. Curr Genet. 40:405–409.
Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JH, editor. Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics. New York: Springer. p. 123–144.
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.
Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res. 15:616–628.
Ohta T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics 138:1331–1337.
Park S, et al. 2015a. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 208:570–583.
Park S, Jansen RK, Park SJ. 2015b. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. BMC Plant Biol. 15:40.
Park S, et al. 2014. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. BMC Genomics. 15:405.
Parkinson CL, et al. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol. 5:73.
Peltier JB, 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.
Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipera*: Ancestral gene content and order, ancestral editing sites and extraordinarily low mutation rate. BMC Biol. 11:29.
R Development Core Team. 2015. R: A language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/.
Rockenbach K, et al. 2016. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics 204:1507–1522.
Rousseau-Gueutin M, 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.
Ruhlman TA, et al. 2015. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. BMC Plant Biol. 15:100.
Ruhlman TA, Jansen RK. 2014. The plastid genomes of flowering plants. In: Maliga P, editor. Chloroplast biotechnology: methods and protocols. Vol. 1132. New York: Springer Science and Business Media, LLC. p. 3–38.
Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK. 2017. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot. 104:1–14.
Sabir J, et al. 2014. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. Plant Biotechnol J. 12:743–754.
Schulte W, Töpfer R, Stracke R, Schell J, Martini N. 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 U S A. 94:3465–3470.
Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. Proc Natl Acad Sci U S A. 112:E3515–E3524.
Sloan DB, et al. 2012a. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 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, Barr CM, Olson MS, Keller SR, Taylor DR. 2008. Evolutionary rate variation at multiple levels of biological organization in plant mitochondrial DNA. Mol Biol Evol. 25:243–246.
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, Taylor DR. 2012. Evolutionary rate variation in organelle genomes: the role of mutational processes. In: Bullerwell CE, editor. Organelle genetics. Heidelberg (Germany): Springer-Verlag. p. 123–146.
Sloan DB, et al. 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol Phylogenet Evol. 72:82–89.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.
Stanne TM, Sjögren LL, Koussevitzky S, Clarke AK. 2009. Identification of new protein substrates for the chloroplast ATP-dependent Clp protease supports its constitutive role in *Arabidopsis*. Biochem J. 417:257–268.
Sveinsson S, Cronk Q. 2014. Evolutionary origin of highly repetitive plastid genomes within the clover genus (*Trifolium*). BMC Evol Biol. 14:228.
Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet. 5:123–135.
Ueda M, et al. 2007. Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. Gene 402:51–56.
Ueda M, et al. 2008. Substitution of the gene for chloroplast RPS16 was assisted by generation of a dual targeting signal. Mol Biol Evol. 25:1566–1575.
Wakil SJ, Stoops JK, Joshi VC. 1983. Fatty acid synthesis and its regulation. Annu Rev Biochem. 52:537–579.
Wang C-W, Chen S-Y, Zhang X-Z. 2016. Chloroplast genome evolution in Actinidiaceae: *clpP* loss, heterogeneous divergence and phylogenomic practice. PLoS ONE. 11:e0162324.
Weng M-L, Ruhlman TA, Gibby M, Jansen RK. 2012. Phylogeny, rate variation, and genome size evolution of *Pelargonium* (Geraniaceae). Mol Phylog Evol. 64:654–670.
Weng ML, Blazier JC, Govindu M, Jansen RK. 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 ML, Ruhlman TA, Jansen RK. 2016. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol. 8:1824–1838.
Weng ML, Ruhlman TA, Jansen RK. 2017. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 214:842–851.
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.
Wright F. 1990. The effective number of codons used in a gene. Gene 87:23–29.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
Yi X, Gao L, Wamg B, Su Y-J, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. Genome Biol Evol. 5:688–698.
Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821–829.
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.
Zhang J, et al. 2016. Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. Genome Biol Evol. 8:622–634.
Zhu A, Guo W, Jain K, Mower JP. 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. Mol Biol Evol. 31:1228–1236.

Associate editor: Shu-Miaw Chaw