Detecting de novo mitochondrial mutations in angiosperms with highly divergent evolutionary rates

Amanda K. Broz,1,† Gus Waneka,1,† Zhiqiang Wu,1,2,† Matheus Fernandes Gyorfy,1 and Daniel B. Sloan1,1,*

1Department of Biology, Colorado State University, Fort Collins, CO 80523, USA
2Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, 518120 Shenzhen, China
†These authors contributed equally to this work.
*Corresponding author: dan.sloan@colostate.edu

Abstract

Although plant mitochondrial genomes typically show low rates of sequence evolution, levels of divergence in certain angiosperm lineages suggest anomalously high mitochondrial mutation rates. However, de novo mutations have never been directly analyzed in such lineages. Recent advances in high-fidelity DNA sequencing technologies have enabled detection of mitochondrial mutations when still present at low heteroplasmic frequencies. To date, these approaches have only been performed on a single plant species (Arabidopsis thaliana). Here, we apply a high-fidelity technique (Duplex Sequencing) to multiple angiosperms from the genus Silene, which exhibits extreme heterogeneity in rates of mitochondrial sequence evolution among close relatives. Consistent with phylogenetic evidence, we found that Silene latifolia maintains low mitochondrial variant frequencies that are comparable with previous measurements in Arabidopsis. Silene noctiflora also exhibited low variant frequencies despite high levels of historical sequence divergence, which supports other lines of evidence that this species has reverted to lower mitochondrial mutation rates after a past episode of acceleration. In contrast, S. conica showed much higher variant frequencies in mitochondrial (but not in plastid) DNA, consistent with an ongoing bout of elevated mitochondrial mutation rates. Moreover, we found an altered mutational spectrum in S. conica heavily biased towards AT→GC transitions. We also observed an unusually low number of mitochondrial genome copies per cell in S. conica, potentially pointing to reduced opportunities for homologous recombination to accurately repair mismatches in this species. Overall, these results suggest that historical fluctuations in mutation rates are driving extreme variation in rates of plant mitochondrial sequence evolution.

Keywords: Silene; mitogenome; mutation rate; Duplex Sequencing; genome copy number

Introduction

Plant mitochondrial genomes exhibit dramatic variation in rates of nucleotide substitution. Early molecular evolution studies (Wolfe et al. 1987; Palmer and Herbon 1988) established that mitochondrial rates in most angiosperms are about an order of magnitude lower than in the nucleus (Dröin et al. 2008), which contrasts with the rapid evolution of mitochondrial DNA (mtDNA) in many other eukaryotic lineages (Brown et al. 1979; Smith and Keeling 2015; Lavrov and Pett 2016). However, subsequent phylogenetic surveys have identified a number of angiosperms with exceptionally high levels of mtDNA sequence divergence (Palmer et al. 2000; Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2009; Skippington et al. 2015; Zervas et al. 2019). As such, despite the relatively recent origin and diversification of angiosperms (Barba-Montoya et al. 2018), mitochondrial substitution rates are estimated to span a remarkable 5000-fold range across this group (Richardson et al. 2013). At one extreme, Magnolia stellata has a rate of only ~0.01 synonymous substitutions per site per billion years (SSB), while certain Plantago and Silene species have estimated rates of >50 SSB (Mower et al. 2007; Sloan et al. 2012a; Richardson et al. 2013). In some cases, rate accelerations appear to be short-lived with bursts of sequence divergence inferred for internal branches on phylogenetic trees followed by reversions to slower rates of sequence evolution (Cho et al. 2004; Parkinson et al. 2005; Sloan et al. 2009; Skippington et al. 2017).

The angiosperm genus Silene (Caryophyllaceae) is a particularly interesting model for the study of mitochondrial genome evolution and substitution rate variation. This large genus comprises ~850 species (Jafari et al. 2020) and exhibits rate accelerations that rival the magnitude of other extreme cases in genera such as Plantago, Pelargonium, and Viscum (Mower et al. 2007; Skippington et al. 2015). Moreover, Silene is distinct because the observed rate accelerations appear to have occurred very recently (~10 Mya), such that close relatives within the genus exhibit radically different substitution rates (Figure 1, Mower et al. 2007, Sloan et al. 2009; Rautenberg et al. 2012). For example, S. latifolia (section Melandrium) has retained a substitution rate that is roughly in line with the low levels of sequence divergence found in most angiosperms. In contrast, species such as Silene noctiflora and S. conica represent lineages (sections Elisanthe and Conaminorpha, respectively) that are within the same subgenus as section Melandrium but have highly divergent mtDNA sequence. These large differences
Materials and methods

Plant lines and growth conditions

A single family of full siblings was grown for each of three Silene species. Families were taken from lines with previously sequenced mitochondrial and plastid genomes: *S. latifolia* UK2600, *S. noctiflora* OSR, and *S. conica* ABR (Sloan et al. 2012a,b). The latter two species are hermaphroditic and readily self-fertilize, so families were derived from selfed parents. In contrast, *S. latifolia* is dioecious and exhibits substantial inbreeding depression from full-sib crosses (Teixeira et al. 2009). Therefore, we crossed a female from the UK2600 line with a male from a different line (Kew 32982, obtained from the Kew Gardens Millenium Seed Bank) to produce the full-sib family used in this study. Seeds were germinated on ProMix BX soil mix and grown in a growth room under short-day conditions (10-/14-h light/dark cycle). All three species were grown for ~7 weeks to produce sufficiently large rosettes for organelle DNA extractions.

Organelle DNA extractions and Duplex Sequencing

Each full-sib family was divided into three biological replicates, consisting of ~30–40 individual plants per replicate. A total of 35 g of rosette tissue was harvested from each replicate and used for simultaneous extraction of enriched mtDNA and chloroplast DNA (cpDNA) as described previously (Wu et al. 2020a,b). Mitochondria were isolated by differential centrifugation followed by DNase I treatment to remove contaminating DNA not protected by intact mitochondrial membranes. Chloroplasts were isolated on discontinuous sucrose gradients. Following DNA extraction, Duplex Sequencing libraries were constructed as described previously (Wu et al. 2020a). These libraries were multiplexed and sequenced with 2 × 150 bp reads on an Illumina NovaSeq 6000 S4 Lane (NovoGene).

Shotgun Illumina sequencing of total-cellular DNA samples for k-mer database construction

Detection of low-frequency mitochondrial heteroplasmies presents a number of technical challenges that can lead to false
In contrast to the inbred history of the full-sib family, other lines, species (McCauley et al. 2005), are referred to as k-mers (Wu et al. 2020a). The premise of this approach is that variants associated with numts or nupts should be abundant in total-cellular sequencing datasets (as quantified by counting corresponding k-mers in the raw reads) and match the level expected for other nuclear sequences. However, for reasons discussed below (see Results), we expect this filtering approach to be more reliable for numts than for nupts.

To generate a k-mer database for each species we extracted total-cellular DNA from mature leaf tissue using the Qiagen DNeasy kit. For S. noctiflora and S. conica, the sampled individuals were from the same inbred line as the full-sib family used for Duplex Sequencing but separated by at least three generations. In contrast to the inbred history of the S. noctiflora and S. conica lines, S. latifolia is expected to be highly heterozygous, including for numt and nupt alleles. Therefore, in order to capture all numt and nupt alleles that might be segregating in the S. latifolia full-sib family, we generated total-cellular samples for both parents that were crossed to generate the family. Sampling the male parent in this S. latifolia cross also provides reads from its actual mitochondrial genome sequence, which can identify potential false positives resulting from low-frequency paternal transmission of mtDNA. Such “paternal leakage” has been identified in outcrossing Silene species (McCauley et al. 2005; Bentley et al. 2010).

Illumina libraries were generated from each total-cellular DNA sample using the New England Biolabs NEB FS II Ultra kit with ~100 ng of input DNA, a 20 min fragmentation time, and eight cycles of PCR. The S. latifolia and S. noctiflora libraries were multiplexed and sequenced on the same NovaSeq 6000 lane as the above Duplex Sequencing libraries. The S. conica library was sequenced separately with 2 x 150 bp reads on an Illumina HiSeq 4000 lane. The resulting raw reads were used to generate databases of k-mer counts (k = 39 bp) with KMC v3.0.0 (Kokot et al. 2017).

**Duplex Sequencing data analysis and variant calling**

Duplex Sequencing datasets were processed with a previously published pipeline (https://github.com/dbsloan/duplexseq) (Wu et al. 2020a). This pipeline uses Duplex Sequencing barcodes (i.e., unique molecular identifiers) to group raw reads into families corresponding to the two complementary strands of an original DNA fragment, requiring a minimum of three reads for each strand. After trimming barcodes and linker sequences, the consensus sequences for each double-stranded family were mapped to the reference mitochondrial and plastid genomes for the corresponding species. Because of known sequencing artifacts associated with end repair and adapter ligation (Kennedy et al. 2014), the 10 bp at the end of each read were excluded when identifying variants and calculating sequencing coverage.

Reads that contained a single mismatch relative to the reference sequences were used to identify single nucleotide variants (SNVs). Variants with a k-mer count of 10 or greater in the corresponding k-mer database were excluded as likely numts or nupts. This k-mer filtering also ensured elimination of false positives due to paternal leakage from the S. latifolia male parent or errors in the published reference sequence. The latter is an important concern because the published S. conica mitochondrial genome sequence is a draft assembly due to its extreme repetitiveness (Sloan et al. 2012a). Accordingly, we also excluded variants if the corresponding reference sequence was not detected in the k-mer database to account for sites with ambiguities (Ns) or possible errors in the reference. Variants were also filtered using the pipeline’s –recomb_check option, which identifies SNVs that can be explained by recombination between nonidentical repeat sequences within the mitochondrial genome rather than de novo point mutations (Davila et al. 2011). Finally, the pipeline’s –contam_check option was used to provide the reference genome sequences from the other Silene species for filtering of variants arising from contamination between multiplexed libraries.

To calculate SNV frequencies for each library, the number of reads containing an SNV (after filtering) was divided by the total bp of duplex consensus sequence mapped to the genome Differences in SNV frequency among species were tested with a one-way ANOVA implemented, using the aov function in R v3.6.3 followed by post hoc pairwise comparisons with the Tukey’s honestly significant difference (HSD) function.

For calling mitochondrial SNVs, we were able to use the cpDNA Duplex Sequencing libraries to increase our mitochondrial genome coverage because they contained a substantial amount of contaminating mtDNA reads resulting from incomplete enrichment of cpDNA. Importantly, being able to use mitochondrial reads from the cpDNA library for S. conica biological replicate 2 was key because sequencing of the mtDNA library for that replicate failed (see “Results”). However, we did not do the reverse (use mtDNA libraries to supplement plastid datasets) because the mtDNA libraries were expected to have a far higher ratio of nuclear to plastid DNA than the cpDNA libraries, exacerbating the challenges associated with filtering numts.

**Analysis of organelle genome copy number**

To estimate the copy number of mitochondrial and plastid genomes based on the total-cellular sequencing datasets, raw reads were trimmed with cutadapt v1.16 (Martin 2011) to remove adapter sequences with an error tolerance of 0.15 and to trim low-quality ends with a q20 threshold. Read pairs with a minimum length of 50 bp each after trimming were retained. Trimmed reads were then mapped to reference mitochondrial and plastid genomes for the corresponding species using bowtie v2.2.23 (Langmead and Salzberg 2012), and position-specific coverage data were extracted from the resulting alignment (BAM/SAM) files. Average coverage was summarized for 2-kb windows tiled across each organelle genome. To estimate the number of mitochondrial and plastid genomes per nuclear genome copy, we assumed that the remaining unmapped reads were all nuclear. We used the total length of these unmapped reads divided by the nuclear genome size of the corresponding species to estimate the average nuclear genome coverage and obtain ratios of organelle to nuclear coverage. Nuclear genome size estimates of 2.67, 2.78,
and 0.93 Gb were used for S. latifolia, S. noctiflora, and S. conica, respectively (Williams et al. 2020).

The above analysis of genome copy number identified a surprisingly low number of mitochondrial genome copies in S. conica. To validate this unexpected result and compare stoichiometry of mitochondrial, plastid, and nuclear genomes across multiple tissue samples in S. conica, we performed a droplet digital PCR (ddPCR) analysis. We used the same S. conica total-cellular DNA extraction that was used for the Illumina shotgun sequencing. In addition, we harvested leaf tissue from three individuals from a different batch of S. conica ABR plants. For both the original sample and the newer samples, tissue was harvested from the largest rosette leaves. The original sample was grown in a growth chamber under long-day conditions (16-/8-h light/dark cycle) and harvested after 40 days. The tissue sample also differed in that the entire rosette leaf was sampled for light racks in a growth room under short-day conditions (10-/14-h light/dark cycle) and harvested after 40 days. The tissue sampling also differed in that the entire rosette leaf was sampled for the newer replicates, whereas only the distal half of the leaf was taken for the original DNA extraction. In both cases, DNA extractions were performed with a QiaGen DNeasy Kit. A total of six ddPCR markers were developed, with two each targeting the mitochondrial, plastid and, nuclear genomes (Supplementary Table S1). Each ddPCR reaction was set up in a 20 μL volume, containing Bio-Rad QX200 ddPCR EvaGreen Supermix and 0.2 μM of each primer. For mitochondrial and nuclear markers, 5 ng of total-cellular DNA was used as template. To avoid saturation with the higher copy number plastid makers, a 200-fold dilution (25 pg) of the template DNA was used. The reaction volumes were then emulsified in Bio-Rad QX200 Droplet Generation Oil for EvaGreen, using the Bio-Rad QX200 Droplet Generator. PCR amplification was performed on a Bio-Rad C1000 Touch thermal cycler with an initial incubation at 95°C for 5 min, 40 cycles of 90°C for 30 s and 60°C for 1 min, followed by a 4°C incubation for 5 min and a final 90°C incubation for 5 min before holding at 4°C. After amplification, droplets were analyzed on a Bio-Rad QX200 Droplet Reader, and the absolute copy number of each PCR target was estimated based on a Poisson distribution in the Bio-Rad QuantaSoft package. Mitochondrial: nuclear and plastid: nuclear ratios were calculated by dividing organellar marker copy numbers by the mean of the estimates for the two nuclear markers for a given sample.

Results

Duplex Sequencing yield and read mapping

Each of the Duplex Sequencing libraries produced between 88 and 142 M read pairs (Supplementary Table S2), with the exception of the S. conica mtDNA library for biological replicate 2, which was not sequenced due to an apparent loading error. Nevertheless, we were still able to calculate mitochondrial SNV frequencies for S. conica replicate 2 by taking advantage of contaminating mitochondrial reads in the enriched cpDNA library for that replicate. The large number of reads per library translated into sizeable single-stranded read families for construction double-stranded consensus sequences, with modal values of >10 reads per family for most libraries (Supplementary Figure S1). The three species differed in their degree of enrichment in the mtDNA libraries. For both S. noctiflora and S. conica, an average of 79% of the sequences in the mtDNA libraries could be mapped to the reference mitochondrial genome, whereas only 30% of the sequences in the S. latifolia mitochondrial libraries mapped, indicating a much lower level of mtDNA enrichment (Supplementary Table S2). For all three species, between 64% and 67% of sequences from the cpDNA libraries mapped to the plastid genome, with a substantial number of contaminating sequences mapping to the mitochondrial genome (8%, 27%, and 31% for S. latifolia, S. noctiflora, and S. conica, respectively). To make use of as much mitochondrial data as possible, we combined all mitochondrial-derived sequences from both sample types. After collapsing raw reads into duplex consensus sequences and mapping them to the reference genome, we obtained between 85 and 382 Mb of mapped mitochondrial-sequence data for each replicate (based on combing coverage from both mtDNA and cpDNA libraries; Table S2). For plastid genome coverage, we relied solely on the cpDNA libraries, which yielded between 205 and 319 Mb of coverage per replicate (Supplementary Table S2).

Three Silene species differ in their frequencies of mitochondrial SNVs but show little variation in plastid SNV frequencies

Using the variant calls from the duplex consensus sequences (Supplementary File S1), we calculated the frequency of mitochondrial SNVs per mapped bp for each Silene replicate and compared those values to previously published estimates from A. thaliana (Wu et al. 2020a). After applying filtering criteria to exclude false positives associated with numts, low-frequency paternal transmission, chimeric recombination products, or errors in the reference sequence (see ‘Materials and methods’), we found significant variation in (log-transformed) mitochondrial SNV frequency among species (one-way ANOVA, F = 5.3 x 10^{-6}, Figure 2 and Supplementary Table S3).

The three biological replicates of S. latifolia had a mean mitochondrial SNV frequency of 1.7 x 10^{-7} per bp. Silene latifolia was selected for this study because it exhibited very little mitochondrial rate acceleration in previous phylogenetic analyses, suggesting that it has retained the slow rate of sequence evolution that is characteristic of most plant mitochondrial genomes (Mower et al. 2007; Sloan et al. 2009). Accordingly, the S. latifolia estimate was very similar to our previous estimate of 1.8 x 10^{-7} per bp for the mitochondrial SNV frequency in wild type A. thaliana Col-0 (Wu et al. 2020a), another plant species with a typically low rate of mitochondrial sequence evolution (Mower et al. 2007). In contrast to the low historical substitution rates in S. latifolia and Arabidopsis, S. noctiflora exhibits highly accelerated mitochondrial sequence evolution since diverging from other major lineages within the genus Silene (Figure 1). However, we did not find an elevated frequency of mitochondrial SNVs in our Duplex Sequencing analysis of S. noctiflora, suggesting that this species may have experienced a reversion to lower mutation rates. In fact, the mean SNV frequency in S. noctiflora was 6.1 x 10^{-8} per bp, which was approximately 3-fold lower than in S. latifolia or A. thaliana (Tukey’s HSD post hoc test, F = 0.01 for both comparisons). The most noteworthy variation among species was based on observed SNV frequencies in S. conica, representing another Silene lineage with a history of rapid mitochondrial sequence divergence (Figure 1). The mean mitochondrial SNV frequency in S. conica was 1.7 x 10^{-6} per bp, which was 9- to 27-fold higher than in A. thaliana and the other two Silene species (Tukey’s HSD post hoc test, P < 0.0001 for all three comparisons). All of these SNV frequencies substantially exceed the noise threshold of ~2 x 10^{-8} that we previously estimated for this protocol using Escherichia coli samples derived from single colonies as “negative controls” (Wu et al. 2020a).

Silene conica was also distinct in that a large proportion of the identified mitochondrial SNVs (31.7%) were found in two or more biological replicates. Because our biological replicates represented sets of individuals from the same full-sib family, variants that are shared among replicates likely indicate SNVs that were
already heteroplasmic in the parent and then inherited by the offspring. In contrast, none of the identified mitochondrial SNVs in either S. latifolia or S. noctiflora were present in multiple biological replicates. There is reason to expect that our k-mer filtering may have introduced bias against detection of inherited heteroplasmies in S. latifolia (see ‘Discussion’). Nevertheless, even without this filtering, only 3.7% of S. latifolia SNVs (and only 1.8% of S. noctiflora SNVs) would be present in two or more libraries. Therefore, the filtering does not appear to explain this observed difference between S. conica and the other Silene species.

Unlike in the mitochondrial genome, we found no evidence that S. conica has an elevated frequency of SNVs in its plastid genome, as the three Silene species all exhibited similarly low plastid SNV frequencies (Figure 2). We recommend that the estimates of the Silene plastid SNV frequencies and spectra be interpreted cautiously because of the nuclear contamination in these libraries and the fact that nupts are more difficult to reliably filter with our k-mer database than nunts. The challenge that nupts pose relates to the high coverage of true plastid DNA in our total-cellular libraries (>10,000×). As such, even rare sequencing errors in total-cellular libraries have the potential to occur repeatedly and exhibit sizeable counts in our k-mer database, which could lead to exclusion of variant calls that are actually true de novo mutations. Nevertheless, we can confidently conclude that S. conica does not exhibit a major increase in plastid SNV frequency. Even if we performed no k-mer filtering whatsoever on the S. conica plastid samples, SNV frequencies would only increase by 55% on average (Supplementary Table S4), leaving them at a level that is still similar to that of the other Silene species and more than an order of magnitude lower than the (filtered) mitochondrial SNV frequencies in S. conica (Figure 2).

**Variation in mitochondrial mutation spectra among Silene species and extreme GC bias in S. conica**

Previous analysis of low-frequency mitochondrial SNVs in rosette tissue from wild type A. thaliana Col-0 (Wu et al. 2020a) identified a mutation spectrum dominated by GC→AT transitions (Figure 2). The slowly evolving S. latifolia mitochondrial genome exhibited a bias in this same direction with 57% of all observed SNVs being GC→AT transitions (Supplementary Table S3). In contrast, S. noctiflora did not show a similar bias. The low overall SNV frequency in S. noctiflora makes it difficult to precisely estimate the mutation spectrum, but no single substitution type dominated, as GC→AT transitions, AT→GC transitions, and GC→CG transversions all had similar frequencies in the observed set of SNVs (Supplementary Figure S2 and Table S3). Once again, S. conica exhibited the most extreme departure from the other species. Notably, the high rate in S. conica was not driven by an increased frequency of the GC→AT transitions that dominate the spectrum of A. thaliana and S. latifolia. In fact, the frequency of GC→AT transitions in S. conica was lower than in either of those species. Instead, the high overall SNV frequency was largely the result of a massive increase in the frequency AT→GC transitions, which account for 77% of the observed SNVs in S. conica (Figure 2 and Supplementary Table S3). This species also exhibited a substantial increase in the frequency of AT→CG transversions (11% of observed SNVs). As such, both of the dominant types of substitutions in the S. conica mitochondrial mutation spectrum increase guanine→cytosine (GC) content, which is unusual because mutation spectra are generally AT-biased (Hershberg and Petrov 2010; Hildebrand et al. 2010; Sloan and Wu 2014).

**Unusually low mitochondrial genome copy number in S. conica rosette tissue**

By performing deep sequencing of total-cellular DNA to generate a k-mer database for variant filtering, we were also able to estimate the relative copy number of mitochondrial, plastid, and nuclear genomes in each of the three Silene species (Figures 3, Supplementary Figures S3–S5, and Table S5). We found similar plastid genome copy numbers across species, with mean estimates of 378, 263, and 275 plastid genome copies per nuclear genome copy for S. latifolia, S. noctiflora, and S. conica, respectively. If
we assume that each cell is diploid and has not yet undergone DNA replication, the plastid genome copy numbers per cell would be double those values, but that may be an underestimate because many plants undergo extensive endoreduplication, in which the nuclear genome replicates without subsequent cell divisions, resulting in cells with higher nuclear ploidies (Joubes and Chevalier 2000). The number of mitochondrial genome copies was surprisingly low in *S. conica*, with an average of only 0.38 copies per nuclear genome copy. In contrast, *S. latifolia* and *S. noctiflora* had an average of 47.7 and 9.7 mitochondrial genome copies per nuclear genome copy, respectively, which is more consistent with estimates from other plants (Preuten et al. 2010; Oldenburg et al. 2013; Shen et al. 2019).

To validate the finding of extremely low mitochondrial genome copy number in *S. conica*, we performed a ddPCR analysis with two markers each for the mitochondrial, plastid, and nuclear genomes. We first analyzed DNA from the same extraction that was originally used for the total-cellular Illumina shotgun sequencing, obtaining an estimate of 0.42 mitochondrial genomes per nuclear genome copy. This is in close agreement with our estimate of 0.38 from the sequencing data. Based on this ddPCR analysis, we also estimated that there were 737 copies of the plastid genome per nuclear genome copy, which was 2.7-fold higher than our original estimate (Figure 3), possibly indicating that our sequencing estimate was downwardly biased for plastid genome copy number. We then analyzed leaf DNA collected from three new *S. conica* plants that were grown separately from the originally sampled plant to assess whether the original DNA extraction was anomalous in some way. These new samples also produced extremely low estimates for the number of mitochondrial genomes copies with a mean of 1.02 per nuclear genome copy (Figure 3 and Supplementary Table S6). Therefore, these new *S. conica* samples exhibited a small increase in the mitochondrial:nuclear ratio relative to our original sample but still fell well below typical observations for plant cells.

**Discussion**

The challenges of detecting de novo mutations in plant organelle genomes

High-fidelity techniques such as Duplex Sequencing (Schmitt et al. 2012) have been key innovations to address the challenge of detecting and quantifying rare mutations (Salk et al. 2018; Sloan et al. 2018), but some sources of false positives remain. The potential misidentification of numts and nupts as de novo mutations was a particular concern in this study. High-quality nuclear genome assemblies are not available for *Silene*, so it is not possible to use a reference genome to identify and filter numt- and nupt-associated variants. Moreover, our mitochondrial and plastid DNA preparations only reached moderate levels of enrichment, leading to substantial amounts of contaminating nuclear DNA (Supplementary Table S2). Our approach to avoid erroneous numt and nupt variant calls was based on filtering using a k-mer database derived from total-cellular sequencing (see “Materials
and methods"), but there are some concerns about balancing false positives and false negatives that should be considered.

In particular, there is a risk that k-mer filtering may exclude true heteroplasmies if they are shared between the total-cellular sample used to generate the k-mer database and the family used for Duplex Sequencing. We reduced the risk of this in S. conica and S. noctiflora by using individuals separated by at least three generations for constructing the k-mer database. Therefore, low-frequency heteroplasmies would have to have been maintained across multiple generations to lead to improper exclusion of true mitochondrial variants. Although transmission of heteroplasmies across generations certainly occurs (McCaulay 2013; Zhang et al. 2018; Mandel et al. 2020), the segregational loss of low-frequency variants should greatly reduce the magnitude of this problem. In contrast, for the outcrossing species S. latifolia, we used both parents of the full-sib family to construct total-cellular k-mer databases in order to screen for possible numts and nupts in all contributing nuclear haplotypes. As such, variants that were heteroplasmic in the S. latifolia mother and transmitted to her offspring might have been improperly filtered because of their presence in the total-cellular k-mer database, resulting in a potential downward bias on our estimate of the overall frequency of SNVs in S. latifolia.

Despite the uncertainty that this introduces into SNV frequency estimates, we feel that the major qualitative conclusions of this study are robust to the challenges of numt and nupt artifacts. For example, the finding that S. noctiflora appears to have "reverted" to a low SNV frequency is not sensitive to k-mer filtering. Even if no k-mer filtering whatsoever were performed for S. noctiflora, it would still exhibit an SNV frequency lower than the filtered values from the other three species (Supplementary Table S3). Likewise, even if we did not perform any numt filtering on the S. latifolia SNVs (which would almost certainly lead to a substantial overestimation of true mitochondrial mutations), the SNV frequency for S. latifolia would still not reach the highly elevated levels in S. conica. Therefore, the key distinctions among the three species in mitochondrial SNV frequency appear to hold even though some caution is warranted in interpreting the specific frequency estimate in S. latifolia. Furthermore, as noted in the "Results," the conclusion that plastid SNV frequencies remain low in S. conica is not sensitive to k-mer filtering, as removing this filtering step only produces a modest increase in the estimate for S. conica plastid mutations (Supplementary Table S4).

Heteroplasmic SNV frequencies support a history of mitochondrial mutation rate acceleration and reversion in Silene

The high frequency of mitochondrial SNVs captured by Duplex Sequencing in S. conica tissue (Figure 2) is consistent with previous interpretations that increased mutation rates are driving accelerated mitochondrial genome evolution in this and other atypical plant species (Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2009, 2012a; Skippington et al. 2015). This view has become the consensus because accelerations are evident at relatively neutral positions like synonymous sites over phylogenetic timescales, but de novo mitochondrial mutations have never been directly investigated in these high-rate plant lineages until now. The huge increase in AT→GC transitions that dominated the S. conica mutation spectrum (Figure 2) coincides with the most common type of misincorporation observed in steady-state kinetic analysis of the Arabidopsis mtDNA polymerases, which appear to be prone to misincorporating Gs opposite Ts (Ayala-Garcia et al. 2018). Therefore, it is possible that the increased mitochondrial substitution rate and biased mutation spectrum in S. conica reflect a reduced ability to repair mismatches created by polymerase misincorporations. A disproportionate increase in AT→GC transitions was also observed in Arabidopsis mutants lacking a functional copy of MSH1 (Wu et al. 2020a), a gene that may be involved in repair of mismatches via homologous recombination (Christensen 2014; Wynn et al. 2020).

An intact and transcribed copy of MSH1 is retained in S. conica (Havird et al. 2017), but its function and expression patterns have not been investigated. Alternatively, GC-biased gene conversion has also been hypothesized as a mechanism to explain skewed substitution patterns in some plant mitochondrial genomes (Liu et al. 2020).

The extreme bias towards AT→GC transitions in S. conica mitochondrial SNVs (Figure 2) is not entirely consistent with longer-term patterns of mitochondrial sequence divergence in this species. The genome-wide GC content in S. conica (43.1%) is only slightly higher than in congeners like S. latifolia (42.6%), S. noctiflora (40.8%), and S. vulgaris (41.8%; Sloan et al. 2012a). An analysis that used plastid DNA insertions in mitochondrial genomes as neutral markers did find that S. conica had unusually high transition:transversion ratio compared with other angiosperms (Sloan and Wu 2014), which is consistent with the Duplex Sequencing results. However, it did not detect the strong GC bias that we observed in this study.

These discrepancies between phylogenetic patterns and our duplex sequencing data raise two obvious possibilities. First, the observed SNVs in this study may not reflect the spectrum of heritable mutations because they are measured from rosette tissue and thus are expected to include some de novo mutations that occurred in vegetative tissues and were not inherited from the mother or transmitted to future generations. Our choice to sample whole rosettes (as opposed to more targeted "germline" tissue such as dissected meristems) reflected the practical need to obtain sufficient quantities of mtDNA and cpDNA for Duplex Sequencing library construction. Although it is important to recognize the observed SNVs include some mutations that are not heritable, we do not believe that this is likely to be the primary explanation for inferred differences in mitochondrial mutation spectra. A large proportion of the S. conica SNVs were shared across more than one biological replicate, implying that they were inherited from a heteroplasmic mother and thus transmitted across generations. Furthermore, these shared variants were even more skewed towards AT→GC transitions than variants that were only detected in a single replicate (Supplementary Table S7), suggesting that heritable mutations do indeed exhibit a very strong GC bias. Relatedly, the fact that S. conica had such a large number of shared SNVs compared with the other two Silene species (see "Results" and Supplementary Table S7) supports the conclusion that the higher overall SNV frequency in S. conica is not solely due to a higher mutation rate in vegetative tissue and indeed reflects an elevated rate of heritable mutations.

Second, it is possible that the mitochondrial mutation spectrum in S. conica is unstable and has changed over time such that the "snapshot" from Duplex Sequencing of heteroplasmic SNVs does not match the average spectrum over the past millions of years in this lineage. A recent analysis of another genus with accelerated mitochondrial sequence divergence (Ajuga) found large increases in GC content (Liu et al. 2020), suggesting that bouts of accelerated and GC-biased evolution can occur in angiosperm mitochondrial genomes.

In contrast to the findings in S. conica, we did not observe elevated SNV frequencies in S. noctiflora (Figure 2) despite a
comparable history of accelerated sequence evolution (Figure 1). The low SNV frequencies in S. noctiflora suggest that it has reverted to lower mutation rates after a past episode of acceleration. This type of reversion has been inferred based on phylogenetic evidence in other accelerated lineages such as Plantago and Pelargonium (Cho et al. 2004; Parkinson et al. 2005). We also have previously speculated that S. noctiflora no longer has a high mitochondrial mutation rate based on its reduced rate of sequence divergence from close relatives S. turkestanica and S. undulata and its extremely low amount of intraspecific sequence polymorphism (Sloan et al. 2009, 2012a; Wu et al. 2015; Wu and Sloan 2019).

However, if a mutation rate reversion has occurred in this lineage, it may not have simply reversed the process that led to the initial rate increase. Notably, S. noctiflora had a different mitochondrial mutation spectrum than the two species that have maintained low mitochondrial substitution rates throughout their histories (A. thaliana and S. latifolia). It also retains a mitochondrial genome that is radically altered in size, structure, and apparent recombinational activity (Sloan et al. 2012a). Therefore, the mechanisms responsible for mtDNA replication and maintenance in S. noctiflora may still be quite different than in typical angiosperms despite the apparent reversion to ancestral-like rates in this species.

The above interpretations are largely based on the premise that the abundance of heteroplasmic SNVs is correlated with the mutation rate. Although it is probably a reasonable assumption that these two features are correlated, the amount of heteroplasmatic genetic variation that is maintained will also depend on the (effective) population size of mitochondrial genome copies. Therefore, we cannot rule out the possibility that some of the observed differences in SNV frequency among species could be related to variation in traits such as the extent of the mtDNA transmission bottleneck during reproduction (Stewart and Chinnery 2015; Zhang et al. 2018; Johnston 2019). Likewise, analysis of mitochondrial SNVs in Arabidopsis leaf tissue has indicated that variant frequencies may be affected by features such as plant age and development (Wynn et al. 2020). Therefore, future studies to separate effects of mutation and population size will be useful.

One possibility is that heteroplasmic SNVs identified by Duplex Sequencing could then be tracked across generations with allele-specific ddPCR. Quantifying variance in levels of inherited heteroplasmies can serve as an effective means to quantify the effective number of transmitted genome copies (Johnston 2019). However, this may be more challenging with species such as S. latifolia and S. noctiflora where inherited heteroplasmies appear to be rare.

Mitochondrial genome copy number and recombinational repair

One unexpected finding from total-cellular shotgun sequencing was the remarkably low copy number of the mitochondrial genome in S. conica (Figure 3). The ratio of mitochondrial to nuclear genome copies in the analyzed leaf tissue samples implies that there are only about one to two mitochondrial genome copies per cell, under the assumption that most cells are diploid. However, this would depend on the extent of endoreduplication in S. conica. Species with smaller nuclear genomes have been found to undergo a greater amount of endoreduplication on average (Barow and Meister 2003), so it is possible that the ratio of organellar to nuclear genome copies is skewed downward in S. conica for this reason. Although there is evidence that plant cells can have far more mitochondria than mitochondrial genome copies (Preuten et al. 2010; Shen et al. 2019), it is difficult to imagine how mitochondrial function could be maintained throughout development with only one or two mitochondrial genome copies per cell. The sequencing and ddPCR datasets used to generate copy-number estimates were derived from mature rosette leaf tissue. Therefore, it is possible that this is a case of mitochondrial genome “abandonment” in tissue that is not destined for further growth or reproduction (Bendich 2013; Oldenburg et al. 2013; Wynn et al. 2020).

Previous studies have suggested that some plants undergo a major decline in plastid genome copy number in mature leaves (Shaver et al. 2006; Rowan et al. 2009), although this conclusion has been the subject to substantial criticism and debate (Golczyk et al. 2014; Greiner et al. 2020). We found that all three Silene species retained hundreds of plastid genome copies per cell, but the dramatic differences in mitochondrial genome copy number across species have intriguing implications. An important future direction will be to characterize variation in S. conica mitochondrial genome copy numbers throughout development, especially in meristematic and reproductive tissues.

Even under the likely scenario that other S. conica tissues harbor higher mitochondrial genome copy numbers than observed in our analysis, it is possible that such values will still be unusually low compared with most plants and other eukaryotes. Silene conica is distinct in having one of the largest and most fragmented mitochondrial genomes ever identified (Sloan et al. 2012a). Such genome size and architecture might pose a challenge for mtDNA maintenance in this species. Notably, mtDNA accounted for a similar proportion of total-cellular DNA in S. conica and S. latifolia despite the ~100-fold difference in mitochondrial genome copy number between these samples because the S. conica mitochondrial genome is ~45-fold larger than in S. latifolia, and the S. conica nuclear genome is ~3-fold smaller than in S. latifolia. Nevertheless, the low mitochondrial genome copy number in S. conica means that any given region of the mitochondrial genome, including key functional content such as protein-coding sequence, has an unusually low stoichiometry relative to the nucleus.

We hypothesize that low mitochondrial genome copy number may be a cause of the high mutation rates in S. conica. It is thought that the typically low mutation rates in plant organelle genomes can be attributed to accurate DNA repair via homologous recombination (Khakhlova and Bock 2006; Christensen 2014; Ayala-García et al. 2018; Chevigny et al. 2020; Wu et al. 2020a). As such, the ability to maintain low rates would be sensitive to the availability of templates for recombinational repair and, thus, the number of genome copies. Notably, we did not observe elevated SNV frequencies in the S. conica plastid genome (Figure 2), which appears to retain a typical copy number, unlike the S. conica mitochondrial genome (Figure 3). This hypothesized role of copy number in recombinational repair is consistent with the observation that nucleotide substitution rates are lower in large repeats than in single-copy regions within plant organelle genomes (Wolfe et al. 1987; Davila et al. 2011; Zhu et al. 2016).

Therefore, if the population of mitochondrial genome copies is too sparsely distributed in the cells of S. conica, it may be unable to make full use of recombinational repair and instead rely on less accurate forms of repair or leave some DNA damage and mismatches entirely unrepaird. Guo (2014) arrived at a similar hypothesis in dissertation research conducted with Jeffrey Mower after observing an unusually low mitochondrial genome copy number in Plantago media, another angiosperm with extremely elevated rates of mitochondrial sequence evolution (Cho et al. 2004). Therefore, it appears possible that correlated changes in mitochondrial mutation rate and genome copy number may have occurred many times independently in plants. Testing this hypothesized relationship between mitochondrial genome copy number and mutation rate should provide insight into the causes of extreme variation in rates of mitochondrial sequence evolution observed across angiosperms.
Data availability
All duplex sequencing and shotgun Illumina sequencing reads were deposited to the NCBI Sequence Read Archive (SRA) under BioProject PRJNA682809 (Supplementary Table S2). Supplementary material is available at figshare: https://doi.org/10.25386/genetics.13726183.

Acknowledgments
We thank Justin Havird for helpful discussion and providing S. conica full-sib seeds and Jessica Warren for assistance with DNA extraction and figure preparation. We also thank two anonymous reviewers for insightful comments that improved the article.

Funding
This work was supported by a grant from the NIH (R01 GM118046) and an NSF graduate fellowship (DGE-1450032).

Conflicts of interest
None declared.

Literature cited
Ahn EH, Lee SH, Kim JY, Chang C-C, Loeb LA. 2016. Decreased mitochondrial mutagenesis during transformation of human breast stem cells into tumorigenic cells. Cancer Res. 76:4569–4578.
Arbeitheuber B, Hester J, Cremona MA, Stoler N, Zaidi A, et al. 2020. Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. PLoS Biol. 18: e3000745.
Ayala-García VM, Baruch-Torres N, García-Medel PL, Brieba LG. 2012. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6: e1000834.
Bendich AJ. 2013. DNA abandonment and the mechanisms of uniparental inheritance of mitochondria and chloroplasts. Chromosome Res. 21:287–296.
Bentley KE, Mandel JR, McCauley DE. 2010. Paternal leakage and heteroplasmy of mitochondrial genomes in Silene vulgaris: evidence from experimental crosses. Genetics. 185:961–968.
Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci U S A. 76:1967–1971.
Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. 2020. DNA repair and the stability of the plant mitochondrial genome. Int J Mol Sci. 21:328.
Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci U S A. 101:17741–17746.
Christensen AC. 2014. Genes and junk in plant mitochondria-repair mechanisms and selection. Genome Biol Evol. 6:1448–1453.
Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, et al. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol. 9:64.
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.
Golczyk H, Greiner S, Wanner G, Weihe A, Bock R, et al. 2014. Chloroplast DNA in mature and senescing leaves: a reappraisal. Plant Cell. 26:847–854.
Greiner S, Golczyk H, Malinova I, Pellizzer T, Bock R, et al. 2020. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. Plant J. 102:730–746.
Guo W. 2014. Evolution of Organellar Genome Architecture in Seed Plants: The Role of Intracellular Gene Transfer, Recombination and Mutation. Lincoln, Nebraska: University of Nebraska. p. 107–144.
Havird JC, Noe GR, Link L, Torres A, Logan DC, et al. 2019. Do angiosperms with highly divergent mitochondrial genomes have altered mitochondrial function? Mitochondrion. 49:1–11.
Havird JC, Trapp P, Miller C, Bazos I, Sloan DB. 2017. Causes and consequences of rapidly evolving mtDNA in a plant lineage. Genome Biol Evol. 9:323–336.
Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. Evolution. 69:3069–3081.
Hakizimana-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6: e1000834.
Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6: e1001115.
Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 6: e1001107.
Hoekstra JR, Hipp MJ, Montine TJ, Kennedy SR. 2016. Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage. Ann Neurol. 80:301–306.
Huang CY, Grunheit N, Ahmadinejad N, Timmins JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol. 138:1723–1733.
Jafari F, Zarre S, Gholipour A, Eggens F, Rabeler RK, et al. 2020. A new taxonomic backbone for the infrageneric classification of the species-rich genus Silene (Caryophyllaceae). Taxon. 69:337–368.
Johnston IG. 2019. Varied mechanisms and models for the varying mitochondrial bottleneck. Front Cell Dev Biol. 7:294.
Joubes J, Chevalier C. 2000. Endoreduplication in higher plants. In: Dirk Inze, editor. The Plant Cell Cycle. Springer. p. 191–201.
Kennedy SR, Schmitt MW, Loeb LA. 2013. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. PLoS Genet. 9: e1003794.
Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, et al. 2014. Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc. 9:2586–2606.
Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46:85–94.
Kokot M, Dlugosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. Bioinformatics. 33:2759–2761.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357–359.
Lavrov DV, Pett W. 2016. Animal mitochondrial DNA as we don’t know it: MT-genome organization and evolution in non-bilaterian lineages. Genome Biol Evol. 8:2896–2913.
Liu F, Fan W, Yang J-B, Xiang C-L, Mower JP, et al. 2020. Episodic and guanine–cytosine-biased bursts of intragenomic and interspecific synonymous divergence in Ajugaoidae (Lamiaceae) mitogenomes. New Phytol. 228:1107–1114.

Lough AN, Roark LM, Kato A, Ream TS, Lamb JC, et al. 2008. Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. Genetics. 178:47–55.

Mandel JR, Ramsey AJ, Harley JM, Scott VA, Mody D, et al. 2020. Disentangling complex inheritance patterns of plant organellar genomes: an example from carrot. J Hered. 111:531–538.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embret J. 17:10–12.

McCauley DE. 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytol. 200:966–977.

McCauley DE, Bailey MF, Sherman NA, Darnell MZ. 2005. Evidence for paternal transmission and heteroplasmy in the mitochondrial genome of Silene vulgaris, a gynodioecious plant. Heredity (Edinb). 95:50–58.

Michalovová M, Vyskot B, Kejnovsky E. 2013. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity (Edinb). 111:314–320.

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.

Oldenburg DJ, Kumar RA, Bendich AJ. 2013. The amount and integrity of mtDNA in maize decline with development. Planta. 237:603–617.

Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol. 28:87–97.

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, et al. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci USA. 97:6960–6966.

Parkinson CL, Mower JP, Qiu YI, Shirk AJ, Song K, et al. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol. 5:73.

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, et al. 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant J. 64:948–959.

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

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The “fossilized” mitochondrial genome of Liriodendron tulipifera: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol. 11:29.

Rowan BA, Oldenburg DJ, Bendich AJ. 2009. A multiple-method approach reveals a declining amount of chloroplast DNA during development in Arabidopsis. BMC Plant Biol. 9:3.

Salk JJ, Schmitt MW, Loeb LA. 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat Rev Genet. 19:269–285.

Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, et al. 2012. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci USA. 109:14508–14513.

Shaver JM, Oldenburg DJ, Bendich AJ. 2006. Changes in chloroplast DNA during development in tobacco, Medicago truncatula, pea, and maize. Planta. 224:72–82.

Shen J, Zhang Y, Havey MJ, Shou W. 2019. Copy numbers of mitochondrial genes change during melon leaf development and are lower than the numbers of mitochondria. Hortic Res. 6:95.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant Viscum scurulideum is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci USA. 112:E3515–E3524.

Skippington E, Barkman TJ, Rice DW, Palmer JD. 2017. Comparative mitogenomics indicates respiratory competence in parasitic Viscum despite loss of complex I and extreme sequence divergence, and reveals horizontal gene transfer and remarkable variation in genome size. BMC Plant Biol. 17:49.

Sloan DB, Taylor DR. 2010. Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. J Mol Evol. 70:479–491.

Sloan DB, Wu Z. 2014. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. Genome Biol Evol. 6:3210–3221.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, 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, Broz AK, Sharbrough J, Wu Z. 2018. Detecting rare mutations and DNA damage with sequencing-based methods. Trends Biotechnol. 36:729–740.

Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. 2010. Extensive loss of RNA editing sites in rapidly evolving Silene mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics. 185:1369–1380.

Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe Sileneae (Caryophyllaceae). BMC Evol Biol. 9:260.

Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol Biol Evol. 31:673–682.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci USA. 112:10177–10184.

Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmacy: implications for human health and disease. Nat Rev Genet. 16:530–542.

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, et al. 2001. Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proc Natl Acad Sci USA. 98:5099–5103.

Teixeira S, Foerster K, Bernasconi G. 2009. Evidence for inbreeding depression and post-pollination selection against inbreeding in the dioecious plant Silene latifolia. Heredity (Edinb). 102:101–112.

Weaver RJ, Carrion G, Nix R, Madia GP, Robinowitz S, et al. 2020. High mitochondrial mutation rates in Silene are associated with nuclear-mediated changes in mitochondrial physiology. Biol Lett. 16:20200450.

Williams AM, Igen MW, Lambert A, Broz AK, Carter OG, Sloan DB. 2020. Long-read transcriptome and other genomic resources for the angiosperm Silene noctiflora. bioRxiv. doi: 10.1101/2020.08.09.243378.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 84:9054–9058.
Wu Z, Sloan DB. 2019. Recombination and intraspecific polymorphism for the presence and absence of entire chromosomes in mitochondrial genomes. Heredity (Edinb). 122:647–659.

Wu Z, Cuthbert JM, Taylor DR, Sloan DB. 2015. The massive mitochondrial genome of the angiosperm Silene noctiflora is evolving by gain or loss of entire chromosomes. Proc Natl Acad Sci U S A. 112:10185–10191.

Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020a. MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proc Natl Acad Sci U S A. 117:16448–16455.

Wu Z, Waneka G, Sloan DB. 2020b. The tempo and mode of angiosperm mitochondrial genome divergence inferred from intraspecific variation in Arabidopsis thaliana. G3 (Bethesda) 10:1077–1086.

Wynn E, Purfeerst E, Christensen A. 2020. Mitochondrial DNA repair in an Arabidopsis thaliana uracil N-glycosylase mutant. Plants. 9:261.

Wynn EL, Christensen AC. 2015. Are synonymous substitutions in flowering plant mitochondria neutral? J Mol Evol. 81:131–135.

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

Zervas A, Petersen G, Seberg O. 2019. Mitochondrial genome evolution in parasitic plants. BMC Evol Biol. 19:87.

Zhang H, Burr SP, Chinnery PF. 2018. The mitochondrial DNA genetic bottleneck: inheritance and beyond. Essays Biochem. 62:225–234.

Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209:1747–1756.

Communicating editor: J. Sekelsky