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

Amanda K. Broz*[,1], Gus Waneka*[,1], Zhiqiang Wu[†,*,1], Daniel B. Sloan*[,2]

*Department of Biology, Colorado State University, Fort Collins, CO 80523

[†]Shenzhen 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

[1]These authors contributed equally to this work.

[2]Author for Correspondence:
Daniel B. Sloan: dan.sloan@colostate.edu
Colorado State University
1878 Campus Delivery
Fort Collins, CO 80523
970.491.2256

ABSTRACT

Although plant mitochondrial genomes typically show low rates of sequence evolution, the levels of sequence 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 *Silene*, an angiosperm genus that exhibits extreme heterogeneity in rates of mitochondrial sequence evolution among close relatives. Consistent with phylogenetic evidence, we found that *S. latifolia* maintains low mitochondrial variant frequencies that are comparable to 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* shows much higher variant frequencies, indicating an ongoing bout of elevated mutation rates. Moreover, we found an altered mutational spectrum in *S. conica* with a heavy bias towards AT→GC transitions (and to a lesser extent AT→CG transversions). 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 indicate that historical fluctuations in mutation rates are driving extreme variation in rates of plant mitochondrial sequence evolution.

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 (Drouin, et al. 2008), which contrasts with the rapid evolution of 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, Alverson, Chuckalovcak, et al. 2012; 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 approximately 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 distinctive 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 *S. noctiflora* and *S. conica* represent lineages (sections *Elisanthe* and *Conoimorpha*, respectively) that are within the same subgenus as section *Melandrium* but have highly divergent mtDNA sequence. These large differences among close relatives have enabled comparative approaches to investigate the evolutionary consequences of accelerated substitution rates for mitochondrial genome architecture (Sloan, Alverson, Chuckalovcak, et al. 2012), RNA editing (Sloan, et al. 2010), mitonuclear coevolution (Sloan, et al. 2014; Havird, et al. 2015; Havird, et al. 2017), and mitochondrial physiology (Havird, et al. 2019; Weaver, et al. 2020). Notably, accelerated species such as *S. noctiflora* and *S. conica* also exhibit massively expanded

59 mitochondrial genomes that have been fragmented into dozens of circularly mapping chromosomes
60 (Sloan, Alverson, Chuckalovcak, et al. 2012). However, the mechanisms that cause increased
61 mitochondrial substitution rates in some *Silene* species remain unclear (Havird, et al. 2017).
62 Previous studies on *Silene* and other angiosperms have generally concluded that elevated
63 rates of mitochondrial sequence evolution reflect increased mutation rates rather than changes in
64 selection pressure (Cho, et al. 2004; Parkinson, et al. 2005; Mower, et al. 2007; Sloan, et al. 2009;
65 Sloan, Alverson, Chuckalovcak, et al. 2012; Skippington, et al. 2015). Increases are especially
66 pronounced at synonymous sites, which are thought to experience very limited effects of selection in
67 plant mtDNA (Sloan and Taylor 2010; Wynn and Christensen 2015). Likewise, the ratio of
68 nonsynonymous to synonymous substitutions (d_N/d_S) is still low in accelerated species, indicating
69 that purifying selection on mitochondrial genes remains strong. A number of mechanisms have been
70 hypothesized to explain cases of increased mitochondrial mutation rates. However, thus far, all
71 inferences about variation in plant mitochondrial divergence are based on long-term patterns of
72 sequence divergence across species rather than direct detection and analysis of *de novo* mutations,
73 making it difficult to investigate the actual role of mutation.
74 Recent advances in high-fidelity DNA sequencing have improved our ability to distinguish
75 signal from noise and detect *de novo* mutations (Salk, et al. 2018; Sloan, et al. 2018). In particular,
76 Duplex Sequencing (Schmitt, et al. 2012; Kennedy, et al. 2014) has been used to obtain error rates
77 as low as ~2×10^{-8} per bp, allowing for accurate identification of mitochondrial variants that are still
78 present at low heteroplasmic frequencies within tissue samples (Kennedy, et al. 2013; Ahn, et al.
79 2016; Hoekstra, et al. 2016; Arbeithuber, et al. 2020; Wu, et al. 2020). Duplex Sequencing works by
80 attaching random barcodes to each original DNA fragment and independently sequencing the two
81 complementary DNA strands from that fragment multiple times each to produce a highly accurate
82 double-stranded consensus. To date, *Arabidopsis thaliana* is the only plant system to have been
83 analyzed with Duplex Sequencing (Wu, et al. 2020). Here, we apply Duplex Sequencing to analyze
84 *de novo* mitochondrial mutations in three *Silene* species previously inferred to have dramatically
85 different histories of mitochondrial mutation.
86
87
88 **MATERIALS AND METHODS**
89
90 **Plant lines and growth conditions**
91 A single family of full siblings was grown for each of three *Silene* species. Families were taken from
92 lines with previously sequenced mitochondrial and plastid genomes: *S. latifolia* UK2600, *S. noctiflora*
93 OSR, and *S. conica* ABR (Sloan, Alverson, Chuckalovcak, et al. 2012; Sloan, Alverson, Wu, et al.
94 2012). 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-hr/14-hr light/dark cycle). All three species were grown for approximately seven 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 approximately 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 (by differential centrifugation) and cpDNA (by sucrose gradient centrifugation) followed by construction of Duplex sequencing as described previously (Wu, et al. 2020). 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 positives. In particular, plant nuclear genomes harbor numerous insertions of mtDNA fragments (numts) that can differ in sequence from the mitochondrial genome because they accumulate mutations over time (Huang, et al. 2005; Noutsos, et al. 2005; Lough, et al. 2008; Hazkani-Covo, et al. 2010; Michalovová, et al. 2013). Therefore, contaminating nuclear DNA in mtDNA samples can mimic low-frequency *de novo* mitochondrial mutations. Because numts often closely resemble the mitochondrial sequence, they can be problematic to accurately assemble in nuclear genome sequencing projects even in high quality reference assemblies (Stupar, et al. 2001). As such, it can be difficult to identify and filter out numt-derived variants based only on reference genome assemblies. We have found that an effective alternative is to use raw reads from total-cellular sequencing datasets to generate a *k*-mer database (Wu, et al. 2020). The premise of this approach is that variants associated with numts 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.

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 alleles. Therefore, in order to

131 capture all numt alleles that might be segregating in the *S. latifolia* full-sib family, we generated total-
132 cellular samples for both parents that were crossed to generate the family. Sampling the male parent
133 in this *S. latifolia* cross also provides reads from its actual mitochondrial genome sequence, which
134 can identify potential false positives resulting from low-frequency paternal transmission of mtDNA.
135 Such "paternal leakage" has been identified in outcrossing *Silene* species (McCauley, et al. 2005;
136 Bentley, et al. 2010).
137 Illumina libraries were generated from each total-cellular DNA sample using the New
138 England Biolabs NEB FS II Ultra kit with approximately 100 ng of input DNA, a 20 min fragmentation
139 time, and eight cycles of PCR. The *S. latifolia* and *S. noctiflora* libraries were multiplexed and
140 sequenced on the same NovaSeq 6000 lane as the above Duplex Sequencing libraries. The *S.*
141 *conica* library was sequenced separately with 2×150 bp reads on an Illumina HiSeq 4000 lane. The
142 resulting raw reads were used to generate databases of *k*-mer counts (*k* = 39 bp) with KMC v3.0.0
143 (Kokot, et al. 2017).
144
145 **Duplex Sequencing data analysis and variant calling**
146 Duplex Sequencing datasets were processed with a previously published pipeline
147 (https://github.com/dbsloan/duplexseq) (Wu, et al. 2020). This pipeline uses Duplex Sequencing
148 barcodes (i.e., unique molecular identifiers) to group raw reads into families corresponding to the two
149 complementary strands of an original DNA fragment, requiring a minimum of three reads for each
150 strand. After trimming barcodes and linker sequences, the consensus sequences for each double-
151 stranded family were mapped to the reference mitochondrial and plastid genomes for the
152 corresponding species. Because of known sequencing artefacts associated with end repair and
153 adapter ligation (Kennedy, et al. 2014), the 10 bp at the end of each read were excluded when
154 identifying variants and calculating sequencing coverage.
155 Reads that contained a single mismatch relative to the reference mitochondrial sequences were
156 used to identify single nucleotide variants (SNVs). Variants with a *k-mer* count of 10 or greater in the
157 corresponding *k*-mer database were excluded as likely numts. This *k*-mer filtering also ensured
158 elimination of false positives due to paternal leakage from the *S. latifolia* male parent or errors in the
159 published reference sequence. The latter is an important concern because the published *S. conica*
160 mitochondrial genome sequence is a draft assembly due to its extreme repetitiveness (Sloan,
161 Alverson, Chuckalovcak, et al. 2012). Accordingly, we also excluded variants if the corresponding
162 reference sequence was not detected in the *k*-mer database to account for sites with ambiguities
163 (Ns) or possible errors in the reference. Variants were also filtered using the pipeline's
164 --recomb_check option, which identifies SNVs that can be explained by recombination between non-
165 identical repeat sequences within the mitochondrial genome rather than *de novo* point mutations
166 (Davila, et al. 2011). To calculate SNV frequencies for each library, the number of reads containing

167 an SNV (after filtering) was divided by the total bp of duplex consensus sequence mapped to the
168 genome. Differences in SNV frequency among species were tested with a one-way ANOVA
169 implemented, using the aov function in R v3.6.3 followed by *post hoc* pairwise comparisons with the
170 TukeyHSD function.
171 We are restricting our analysis in this study to the detection of mitochondrial SNVs because
172 the extremely high copy number of plastid DNA in total-cellular libraries makes applying a *k*-mer
173 filtering strategy to exclude variants associated with nuclear insertions of plastid DNA (nupts) less
174 reliable. The average coverage of the plastid genome in our total-cellular libraries was over 10,000×.
175 As such, even rare sequencing errors had the potential to occur repeatedly and exhibit sizeable
176 counts in our *k*-mer database. However, we were able to use the cpDNA Duplex Sequencing
177 libraries to increase our mitochondrial genome coverage because they contained a substantial
178 amount of contaminating mtDNA reads resulting from incomplete enrichment of cpDNA. Importantly,
179 being able to use mitochondrial reads from the cpDNA library for *S. conica* biological replicate 2 was
180 key because sequencing of the mtDNA library for that replicate failed (see Results).
181
182 **Analysis of organelle genome copy number**
183 To estimate the copy number of mitochondrial and plastid genomes based on the total-cellular
184 sequencing datasets, raw reads were trimmed with cutadapt v1.16 (Martin 2011) to remove adapter
185 sequences with an error tolerance of 0.15 and to trim low-quality ends with a q20 threshold. Read
186 pairs with a minimum length of 50 bp each after trimming were retained. Trimmed reads were then
187 mapped to reference mitochondrial and plastid genomes for the corresponding species using bowtie
188 v2.2.23 (Langmead and Salzberg 2012), and position-specific coverage data were extracted from
189 the resulting alignment (BAM/SAM) files. Average coverage was summarized for 2-kb windows tiled
190 across each organelle genome. To estimate the number of mitochondrial and plastid genome per
191 nuclear genome copy, we assumed that the remaining unmapped reads were all nuclear. We used
192 the total length of these unmapped reads divided by the nuclear genome size of the corresponding
193 species to estimate the average nuclear genome coverage and obtain ratios of organelle to nuclear
194 coverage. Nuclear genome size estimates of 2.67, 2.78, and 0.93 Gb were used for *S. latifolia*, *S.*
195 *noctiflora*, and *S. conica,* respectively (Williams, et al. 2020).
196
197 **Data Availability**
198 All duplex sequencing and shotgun Illumina sequencing reads were deposited to the NCBI
199 Sequence Read Archive (SRA) under BioProject PRJNA682809 (Table S1).
200
201
202 **RESULTS**

203

204 **Duplex Sequencing yield and read mapping**

205 Each of the Duplex Sequencing libraries produced between 88 M and 142 M read pairs (Table S1),
206 with the exception of the *S. conica* mtDNA library for biological replicate 2, which was not sequenced
207 due to an apparent loading error. Nevertheless, we were still able to calculate mitochondrial SNV
208 frequencies for *S. conica* replicate 2 by taking advantage of contaminating mitochondrial reads in the
209 enriched cpDNA library for that replicate. The three species differed in their degree of enrichment in
210 the mtDNA libraries. For both *S. noctiflora* and *S. conica*, an average of 79% of the sequences in the
211 mtDNA libraries could be mapped to the reference mitochondrial genome, whereas only 30% of the
212 sequences in the *S. latifolia* mitochondrial libraries mapped, indicating a much lower level of mtDNA
213 enrichment (Table S1). For all three species, between 64% and 67% of sequences from the cpDNA
214 libraries mapped to the plastid genome, with a substantial amount of contaminating sequences
215 mapping to the mitochondrial genome (8%, 27%, and 31% for *S. latifolia*, *S. noctiflora*, and *S.*
216 *conica*, respectively). To make use of as much mitochondrial data as possible, we combined all
217 mitochondrial-derived sequences from both sample types. After collapsing raw reads into duplex
218 consensus sequences and mapping them to the reference genome, we obtained between 84.6 and
219 382.0 Mb of mapped mitochondrial sequence data for each replicate (Table S1).

220

221 **Three *Silene* species differ in their frequencies of mitochondrial SNVs**

222 Using the variant calls from the duplex consensus sequences (File S1), we calculated the frequency
223 of mitochondrial SNVs per mapped bp for each *Silene* replicate and compared those values to
224 previously published estimates from *A. thaliana* (Wu, et al. 2020). After applying filtering criteria to
225 exclude false positives associated with numts, low-frequency paternal transmission, chimeric
226 recombination products, or errors in the reference sequence (see Materials and Methods), we found
227 significant variation in (log-transformed) SNV frequency among species (one-way ANOVA, $p = 5.3 \times$
228 10^{-6}; Figure 2, Table S2).

229 The three biological replicates of *S. latifolia* had a mean mitochondrial SNV frequency of 1.7
230 $\times 10^{-7}$ per bp. *Silene latifolia* was selected for this study because it exhibited very little mitochondrial
231 rate acceleration in previous phylogenetic analyses, largely retaining the slow rate of sequence
232 evolution that is characteristic of most plant mitochondrial genomes (Mower, et al. 2007; Sloan, et al.
233 2009). Accordingly, the *S. latifolia* estimate was very similar to our previous estimate of 1.8×10^{-7} per
234 bp for the mitochondrial SNV frequency in wild type *A. thaliana* Col-0 (Wu, et al. 2020), another plant
235 species with a typically low rate of mitochondrial sequence evolution (Mower, et al. 2007). In contrast
236 to the low historical substitution rates in *S. latifolia* and *Arabidopsis*, *S. noctiflora* exhibits highly
237 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 6.1×10^{-8} per bp, which was approximately 3-fold lower than in *S. latifolia* or *A. thaliana* (Tukey's HSD *post hoc* test, $p = 0.01$ 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 1.7×10^{-6} per bp, which was 9-fold 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).

Silene conica was also distinct in that a large proportion of the identified 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 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.

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. 2020) 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 (Table S2). 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 (Figure S1, Table S2). 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; Table S2). 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 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 *Silene 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 and S2-4, Table S3). 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 conservative because many plants experience endoreduplication and produce 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 0.4 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, which is more consistent with estimates from other plants (Preuten, et al. 2010; Oldenburg, et al. 2013; Shen, et al. 2019).

DISCUSSION

The challenges of detecting *de novo* mutations in plant mitochondrial 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 as *de novo* mitochondrial 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-associated variants. Moreover, our mitochondrial and plastid DNA preparations for this study only reached moderate levels of enrichment, leaving substantial amounts of contaminating nuclear DNA (Table S1). Our approach to avoid erroneous numt 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.

308 In particular, there is a risk that *k*-mer filtering may exclude true heteroplasmies if they are
309 shared between the total-cellular sample used to generate the *k*-mer database and the family used
310 for mtDNA Duplex Sequencing. We reduced the risk of this in *S. conica* and *S. noctiflora* by using
311 individuals separated by at least three generations for constructing the *k*-mer database. Therefore,
312 low-frequency heteroplasmies would have to have been maintained across multiple generations to
313 lead to improper exclusion of true mitochondrial variants. Although transmission of heteroplasmies
314 across generations certainly occurs (McCauley 2013; Zhang, et al. 2018; Mandel, et al. 2020), the
315 segregational loss of low-frequency variants should greatly reduce the magnitude of this problem. In
316 contrast, for the outcrossing species *S. latifolia*, we used both parents of the full-sib family to
317 construct total-cellular *k*-mer databases in order to screen for possible numts in all contributing
318 nuclear haplotypes. As such, variants that were heteroplasmic in the *S. latifolia* mother and
319 transmitted to her offspring might have been improperly filtered because of their presence in the
320 total-cellular *k*-mer database, resulting in a potential downward bias on our estimate of the overall
321 frequency of SNVs in *S. latifolia*.
322 Despite the uncertainty that this introduces into SNV frequency estimates, we feel that the
323 major qualitative conclusions of this study are robust to the challenges of numt artefacts. For
324 example, the finding that *S. noctiflora* appears to have "reverted" to a low SNV frequency is not
325 sensitive to *k*-mer filtering. Even if no *k*-mer filtering whatsoever were performed for *S. noctiflora*, it
326 would still exhibit an SNV frequency lower than the filtered values from the other three species
327 (Table S2). Likewise, even if we did not perform any numt filtering on the *S. latifolia* SNVs (which
328 would almost certainly lead to a substantial overestimation of true mitochondrial mutations), the SNV
329 frequency for *S. latifolia* would still not reach the highly elevated levels in *S. conica*. Therefore, the
330 key distinctions among the three species appear to hold even if some caution is warranted in
331 interpreting the specific frequency estimate in *S. latifolia*.
332
333 **Heteroplasmic SNV frequencies support a history of mitochondrial mutation rate acceleration**
334 **and reversion in *Silene***
335 The high frequency of mitochondrial SNVs captured by Duplex Sequencing in *S. conica* tissue
336 (Figure 2) is consistent with previous interpretations that increased mutation rates are driving
337 accelerated mitochondrial genome evolution in this and other atypical plant species (Cho, et al.
338 2004; Parkinson, et al. 2005; Mower, et al. 2007; Sloan, et al. 2009; Sloan, Alverson, Chuckalovcak,
339 et al. 2012; Skippington, et al. 2015). This view has become the consensus because accelerations
340 are evident at relatively neutral positions like synonymous sites over phylogenetic timescales, but *de*
341 *novo* mitochondrial mutations have never been directly investigated in these high-rate plant lineages
342 until now. The huge increase in AT→GC transitions that dominated the *S. conica* mutation spectrum
343 (Figure 2) coincides with the most common type of misincorporation observed in steady-state kinetic

344 analysis of the *Arabidopsis* mitochondrial DNA polymerases, which appear to be prone to
345 misincorporating Gs opposite Ts (Ayala-García, et al. 2018). Therefore, it is possible that the
346 increased mitochondrial substitution rate and biased mutation spectrum in *S. conica* reflects a
347 reduced ability to repair mismatches created by polymerase misincorporations. A disproportionate
348 increase in AT→GC transitions was also observed in *Arabidopsis* mutants lacking a functional copy
349 of *MSH1* (Wu, et al. 2020), a gene that may be involved in repair of mismatches via homologous
350 recombination (Christensen 2014; Wynn, et al. 2020). An intact and transcribed copy of *MSH1* is
351 retained in *S. conica* (Havird, et al. 2017), but its function and expression patterns have not been
352 investigated. Alternatively, GC-biased gene conversion has also been hypothesized as a mechanism
353 to explain skewed substitution patterns in some plant mitochondrial genomes (Liu, et al. 2020).
354 The extreme bias towards AT→GC transitions in *S. conica* SNVs (Figure 2) is not entirely
355 consistent with longer-term patterns of mitochondrial sequence divergence in this species. The
356 genome-wide GC content in *S. conica* (43.1%) is only slightly higher than in congeners like *S.*
357 *latifolia* (42.6%), *S. noctiflora* (40.8%), and *S. vulgaris* (41.8%) (Sloan, Alverson, Chuckalovcak, et
358 al. 2012). An analysis that used plastid DNA insertions in mitochondrial genomes as neutral markers
359 did find that *S. conica* had unusually high transition:transversion ratio compared to other
360 angiosperms (Sloan and Wu 2014), which is consistent with the Duplex Sequencing results.
361 However, it did not detect the strong GC-bias that we observed in the current study. These
362 discrepancies raise two obvious possibilities. First, the observed SNVs in this study may not reflect
363 the spectrum of heritable mutations because they are measured from rosette tissue and include *de*
364 *novo* mutations in vegetative tissues that are not transmitted to future generations. However, we
365 suspect that this is unlikely to be the primary explanation. A large proportion of the *S. conica* SNVs
366 were shared across more than one biological replicate, implying that they were inherited from a
367 heteroplasmic mother and thus transmitted across generations. Furthermore, these shared variants
368 were even more skewed towards AT→GC transitions than variants that were only detected in a
369 single replicate (Table S4). Second, it is possible that the mutation spectrum is unstable and has
370 changed over time such that the "snapshot" from Duplex Sequencing of heteroplasmic SNVs does
371 not match the average spectrum over the past millions of years in this lineage. A recent analysis of
372 another genus with accelerated mitochondrial sequence divergence (*Ajuga*) found large increases in
373 GC content (Liu, et al. 2020), suggesting that bouts of accelerated and GC-biased evolution can
374 occur in angiosperm mitochondrial genomes.
375 In contrast to the findings in *S. conica*, we did not observe elevated SNV frequencies in *S.*
376 *noctiflora* (Figure 2) despite a comparable history of accelerated sequence evolution (Figure 1). The
377 low SNV frequencies in *S. noctiflora* suggest that it has reverted to lower mutation rates after a past
378 episode of acceleration. This type of reversion has been inferred based on phylogenetic evidence in
379 other accelerated lineages such as *Plantago* and *Pelargonium* (Cho, et al. 2004; Parkinson, et al.

380 2005). We also have previously speculated that *S. noctiflora* no longer has a high mitochondrial
381 mutation rate based on its reduced rate of sequence divergence from close relatives *S. turkestanica*
382 and *S. undulata* and its extremely low amount of intraspecific sequence polymorphism (Sloan, et al.
383 2009; Sloan, Alverson, Chuckalovcak, et al. 2012; Wu, et al. 2015; Wu and Sloan 2019). However, if
384 a mutation rate reversion has occurred in this lineage, it may not have simply reversed the process
385 that led to the initial rate increase. Notably, *S. noctiflora* had a different mutation spectrum than the
386 two species that have maintained low mitochondrial substitution rates throughout their histories (*A.*
387 *thaliana* and *S. latifolia*). It also retains a mitochondrial genome that is radically altered in size,
388 structure, and apparent recombinational activity (Sloan, Alverson, Chuckalovcak, et al. 2012).
389 Therefore, the mechanisms responsible for mitochondrial DNA replication and maintenance in *S.*
390 *noctiflora* may still be quite different than in typical angiosperms despite the apparent reversion to
391 ancestral-like rates in this species.
392 The above interpretations are largely based on the premise that the abundance of
393 heteroplasmic SNVs is correlated with the mutation rate. Although, this is probably a reasonable
394 assumption, the amount of heteroplasmic genetic variation that is maintained will also depend on the
395 (effective) population size of mitochondrial genome copies. Therefore, we cannot rule out the
396 possibility that some of the observed differences in SNV frequency among species could be related
397 to variation in traits such as the extent of the mtDNA transmission bottleneck during reproduction
398 (Stewart and Chinnery 2015; Zhang, et al. 2018; Johnston 2019). Likewise, analysis of mitochondrial
399 SNVs in *Arabidopsis* leaf tissue has indicated that variant frequencies may be affected by features
400 such as plant age and development (Wynn, et al. 2020). Therefore, future studies to separate effects
401 of mutation and population size will be useful.
402
403 **Mitochondrial genome copy number and recombinational repair**
404 One unexpected finding from total-cellular shotgun sequencing was the remarkably low copy number
405 of the mitochondrial genome is *S. conica* (Figure 3). The ratio of mitochondrial to nuclear genome
406 copies implies that there is only about one mitochondrial genome copy per cell, under the
407 assumption that most cells are diploid. However, this would depend on the extent of
408 endoreduplication in *S. conica*. Species with smaller nuclear genomes have been found to undergo
409 a greater amount of endoreduplication on average (Barow and Meister 2003), so it is possible that
410 the ratio of organellar to nuclear genome copies is skewed downward in *S. conica* for this reason.
411 Although there is evidence that plant cells can have far more mitochondria than mitochondrial
412 genome copies (Preuten, et al. 2010; Shen, et al. 2019), it is difficult to imagine how mitochondrial
413 function could be maintained throughout development with only around a single mitochondrial
414 genome copy per cell. The sequencing dataset used to generate copy-number estimates was
415 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 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 total-cellular sequencing dataset, it is possible that such values will still be unusually low compared to most plants and other eukaryotes. *Silene conica* is notable for having one of the largest and most fragmented mitochondrial genomes ever identified (Sloan, Alverson, Chuckalovcak, et al. 2012). 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. 2020). 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. This interpretation 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 unrepaired. 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.

ACKNOWLEDGEMENTS

We thank Justin Havird for helpful discussion and providing *S. conica* full-sib seeds and Jessica Warren for assistance with figure preparation. This work was supported by a grant from the NIH (R01 GM118046) and an NSF graduate fellowship (DGE-1450032).

REFERENCES

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 Research 76:4569-4578.

Arbeithuber B, Hester J, Cremona MA, Stoler N, Zaidi A, Higgins B, Anthony K, Chiaromonte F, Diaz FJ, Makova KD. 2020. Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. PLoS biology 18:e3000745.

Ayala-García VM, Baruch-Torres N, García-Medel PL, Brieba LG. 2018. Plant organellar DNA polymerases paralogs exhibit dissimilar nucleotide incorporation fidelity. The FEBS journal 285:4005-4018.

Barba-Montoya J, dos Reis M, Schneider H, Donoghue PC, Yang Z. 2018. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. New Phytologist 218:819-834.

Barow M, Meister A. 2003. Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size. Plant, Cell & Environment 26:571-584.

Bendich AJ. 2013. DNA abandonment and the mechanisms of uniparental inheritance of mitochondria and chloroplasts. Chromosome research 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. Proceedings of the National Academy of Sciences 76:1967-1971.

Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. 2020. DNA repair and the stability of the plant mitochondrial genome. International Journal of Molecular Sciences 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. Proceedings of the National Academy of Sciences of the United States of America 101:17741-17746.

Christensen AC. 2014. Genes and junk in plant mitochondria-repair mechanisms and selection. Genome Biology and Evolution 6:1448-1453.

Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu YZ, Weigel D, Mackenzie SA. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC biology 9:64.

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Molecular phylogenetics and evolution 49:827-831.

Golczyk H, Greiner S, Wanner G, Weihe A, Bock R, Börner T, Herrmann RG. 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, Börner T, Herrmann RG. 2020. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. The Plant Journal 102:730-746.

Havird JC, Noe GR, Link L, Torres A, Logan DC, Sloan DB, Chicco AJ. 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 Biology and Evolution 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.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genetics 6:e1000834.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genetics 6:e1001115.

Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genetics 6:e1001107.
Hoekstra JG, Hipp MJ, Montine TJ, Kennedy SR. 2016. Mitochondrial DNA mutations increase in early stage Alzheimer disease and are inconsistent with oxidative damage. Annals of neurology 80:301-306.
Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiology 138:1723-1733.
Jafari F, Zarre S, Gholipour A, Eggens F, Rabeler RK, Oxelman B. 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. Frontiers in Cell and Developmental Biology 7:294.
Joubes J, Chevalier C. 2000. Endoreduplication in higher plants. In. The Plant Cell Cycle: Springer. p. 191-201.
Kennedy SR, Salk JJ, Schmitt MW, Loeb LA. 2013. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. PLoS Genetics 9:e1003794.
Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, Prindle MJ, Kuong KJ, Shen JC, Risques RA, et al. 2014. Detecting ultralow-frequency mutations by Duplex Sequencing. Nature Protocols 9:2586-2606.
Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant Journal 46:85-94.
Kokot M, Długosz 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. Nature 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 Biology and Evolution 8:2896-2913.
Liu F, Fan W, Yang JB, Xiang CL, Mower JP, Li DZ, Zhu A. 2020. Episodic and guanine–cytosine-biased bursts of intragenomic and interspecific synonymous divergence in Ajugoideae (Lamiaceae) mitogenomes. New Phytologist 228:1107-1114.
Lough AN, Roark LM, Kato A, Ream TS, Lamb JC, Birchler JA, Newton KJ. 2008. Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. Genetics 178:47-55.
Mandel JR, Ramsey AJ, Holley JM, Scott VA, Mody D, Abbot P. 2020. Disentangling complex inheritance patterns of plant organellar genomes: an example from carrot. JOURNAL OF HEREDITY In Press.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10-12.
McCauley DE. 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytologist 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 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 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 Evolutionary Biology 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 Research 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, Adams KL, Cho Y, Parkinson CL, Qiu Y-L, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proceedings of the National Academy of Sciences 97:6960-6966.
Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. Journal of Molecular Evolution 28:87-97.
Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song K, Young ND, DePamphilis CW, Palmer JD. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evolutionary Biology 5:73.
Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T. 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant Journal 64:948-959.
Rautenberg A, Sloan DB, Aldén V, Oxelman B. 2012. Phylogenetic relationships of Silene multinervia and Silene section Conoimorpha (Caryophyllaceae). Systematic Botany 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 biology 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 Biology 9:3.
Salk JJ, Schmitt MW, Loeb LA. 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nature Reviews Genetics 19:269-285.
Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. 2012. Detection of ultra-rare mutations by next-generation sequencing. Proceedings of the National Academy of Sciences of the United States of America 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. Horticulture Research 6:95.
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 Biology 17:49.
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. Proceedings of the National Academy of Sciences In Press.
Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS biology 10:e1001241.
Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus Silene. Genome Biology and Evolution 4:294-306.
Sloan DB, Broz AK, Sharbrough J, Wu Z. 2018. Detecting rare mutations and DNA damage with sequencing-based methods. Trends in Biotechnology 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 Evolutionary Biology 9:260.

Sloan DB, Taylor DR. 2010. Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. Journal of Molecular Evolution 70:479-491.
Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Molecular Biology and Evolution 31:673-682.
Sloan DB, Wu Z. 2014. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. Genome Biology and Evolution 6:3210-3221.
Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proceedings of the National Academy of Sciences of the United States of America 112:10177-10184.
Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nature Reviews Genetics 16:530-542.
Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J. 2001. Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. Proceedings of the National Academy of Sciences of the United States of America 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 102:101-112.
Weaver RJ, Carrion G, Nix R, Maeda GP, Rabinowitz S, Iverson EN, Thueson K, Havird JC. 2020. High mitochondrial mutation rates in Silene are associated with nuclear-mediated changes in mitochondrial physiology. Biology Letters 16:20200450.
Williams AM, Itgen MW, Lambert A, Broz AK, Mueller RL, Sloan DB. 2020. Long-read transcriptome and other genomic resources for the angiosperm Silene noctiflora. bioRxiv.
Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences 84:9054-9058.
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. Proceedings of the National Academy of Sciences 112:10185-10191.
Wu Z, Sloan DB. 2019. Recombination and intraspecific polymorphism for the presence and absence of entire chromosomes in mitochondrial genomes. Heredity 122:647-659.
Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020. MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proceedings of the National Academy of Sciences In Press.
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? Journal of Molecular Evolution 81:131-135.
Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586-1591.
Zervas A, Petersen G, Seberg O. 2019. Mitochondrial genome evolution in parasitic plants. BMC Evolutionary Biology 19:87.
Zhang H, Burr SP, Chinnery PF. 2018. The mitochondrial DNA genetic bottleneck: inheritance and beyond. Essays in Biochemistry 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 Phytologist 209:1747-1756.

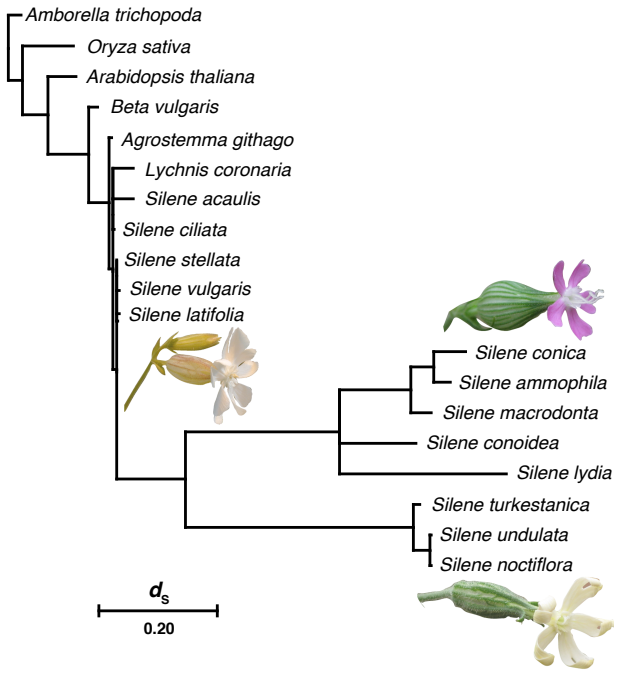

Figure 1. Mitochondrial substitution rate variation among *Silene* species. Branch lengths are scaled based on synonymous substitutions per site (d_S) in a concatenation of three mitochondrial protein-coding genes (*atp1*, *cox3*, and *nad9*) used in previous analyses (Sloan, et al. 2009; Rautenberg, et al. 2012). Branch lengths were estimated with PAML v4.9j (Yang 2007), using a constrained topology. Note that the two accelerated clades (section Conoimorpha [represented by *S. conica*] and section Elisanthe [represented by *S. noctiflora*]) were set as sister groups for this analysis, but there is only weak and inconsistent evidence for that relationship (Rautenberg, et al. 2012; Havird, et al. 2017; Jafari, et al. 2020). Images are shown for each of the three focal species in this study (*S. latifolia*, *S. conica*, and *S. noctiflora*).

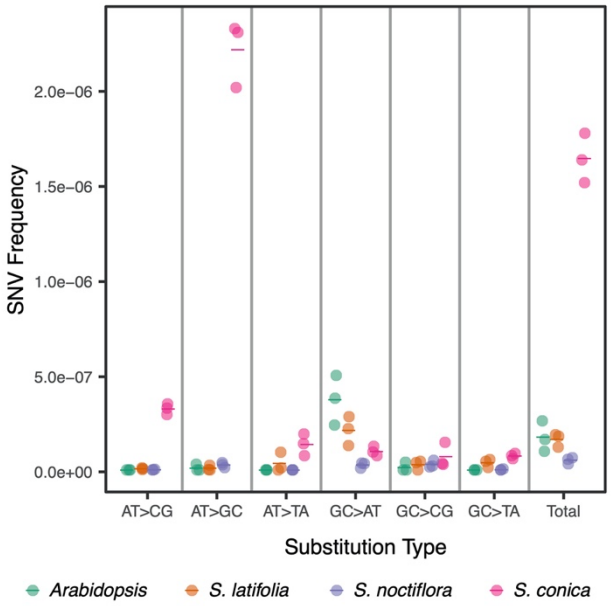

Figure 2. Variation in mitochondrial SNV frequency and spectrum across *Arabidopsis* and *Silene* species. SNV frequencies are calculated as the total number of observed single-nucleotide mismatches in duplex consensus sequence datasets dividided by the total base-pairs of mitochondrial genome coverage in those datasets. Three biological replicates (circles) are shown for each species, with the mean of those replicates indicated with a horizontal line. The *Arabidopsis* data points are taken from Wu et al. (2020). The same data are plotted on a log scale in Figure S1.

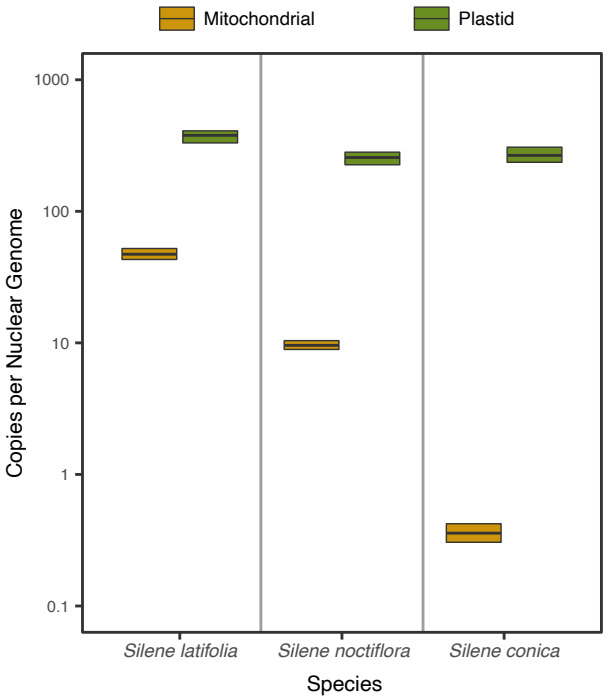

Figure 3. Variation in mitochondrial genome copy number among *Silene* species. Average mitochondrial, plastid, and nuclear genome copy number were estimated from total-cellular Illuimina shotgun sequencing of leaf tissue from each species. Boxplots show median and interquartile ranges for the ratio of organelle genome copy number to nuclear genome copy number based on scanning the organelle genomes in 2-kb windows.

SUPPLEMENTARY MATERIAL

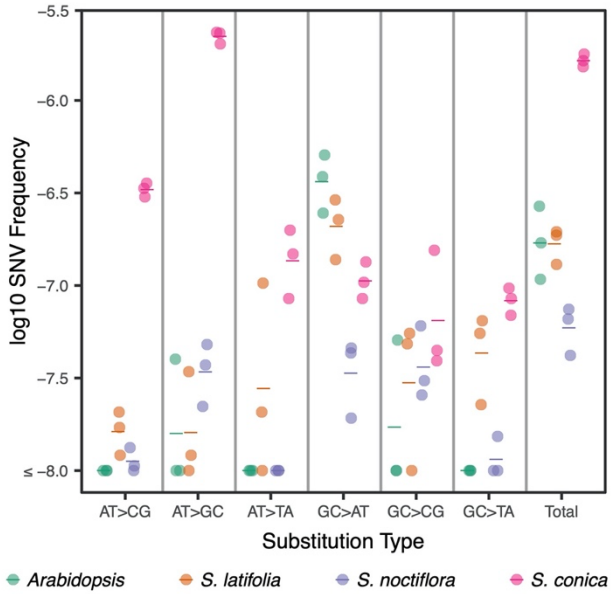

Figure S1. Variation in mitochondrial SNV frequency and spectrum across *Arabidopsis* and *Silene* species. The same data shown in Figure 2 are plotted on a log scale here.

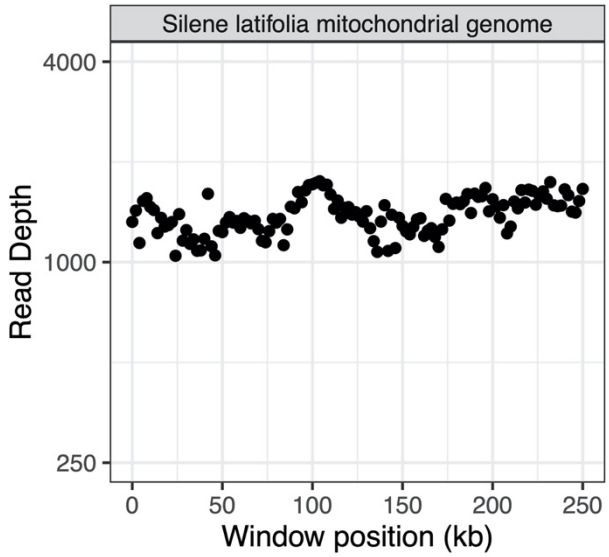

Figure S2. Summary of mitochondrial genome coverage in the *S. latifolia* UK2600 total-cellular sequencing library. Coverage estimated based on reads mapping with a maximum of one mismatch and no indels in 2-kb windows.

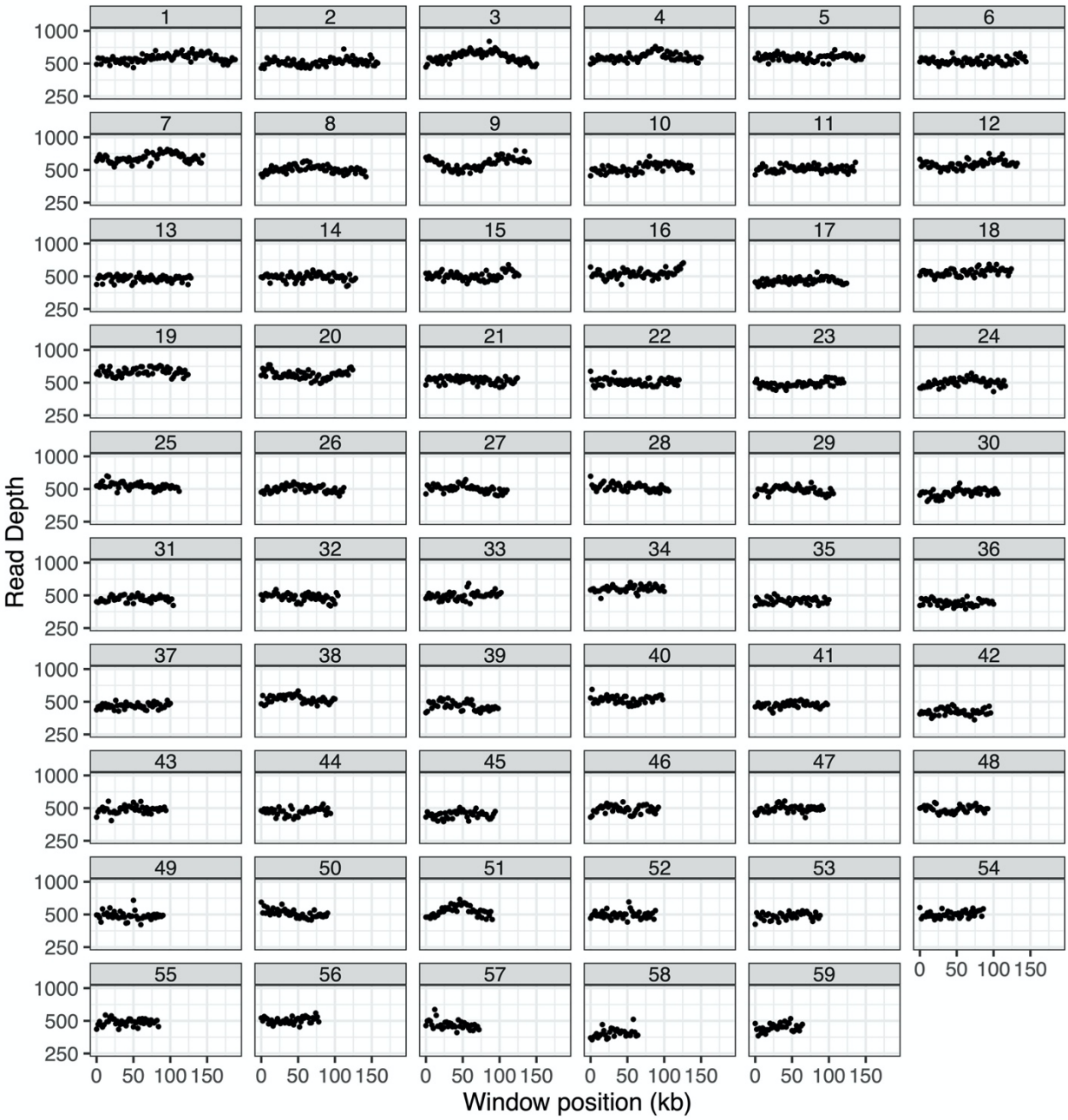

Figure S3. Summary of mitochondrial genome coverage in the *S. noctiflora* OSR total-cellular sequencing library. Coverage estimated based on reads mapping with a maximum of one mismatch and no indels in 2-kb windows. Each panel represents a different chromosome within the multichromosomal genome.

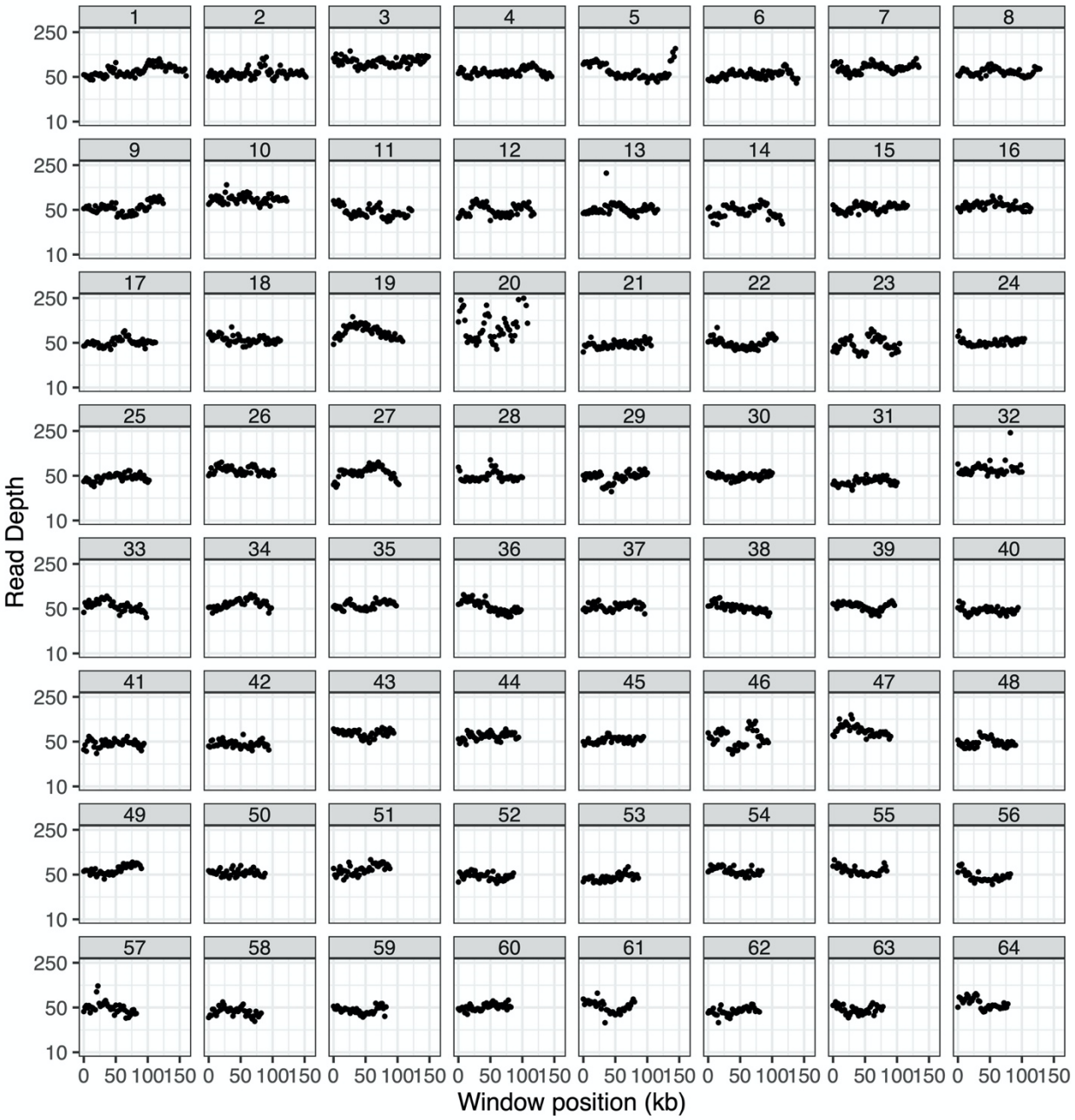

Figure S4a. Summary of coverage for mitochondrial chromosomes 1 through 64 in the *S. conica* ABR total-cellular sequencing library. Coverage estimated based on reads mapping with a maximum of one mismatch and no indels in 2-kb windows. Note that four of the data points on chromosome 20 and two of the data points on chromosome 32 exceeded a coverage of 250× and are not shown to improve readability of the plots.

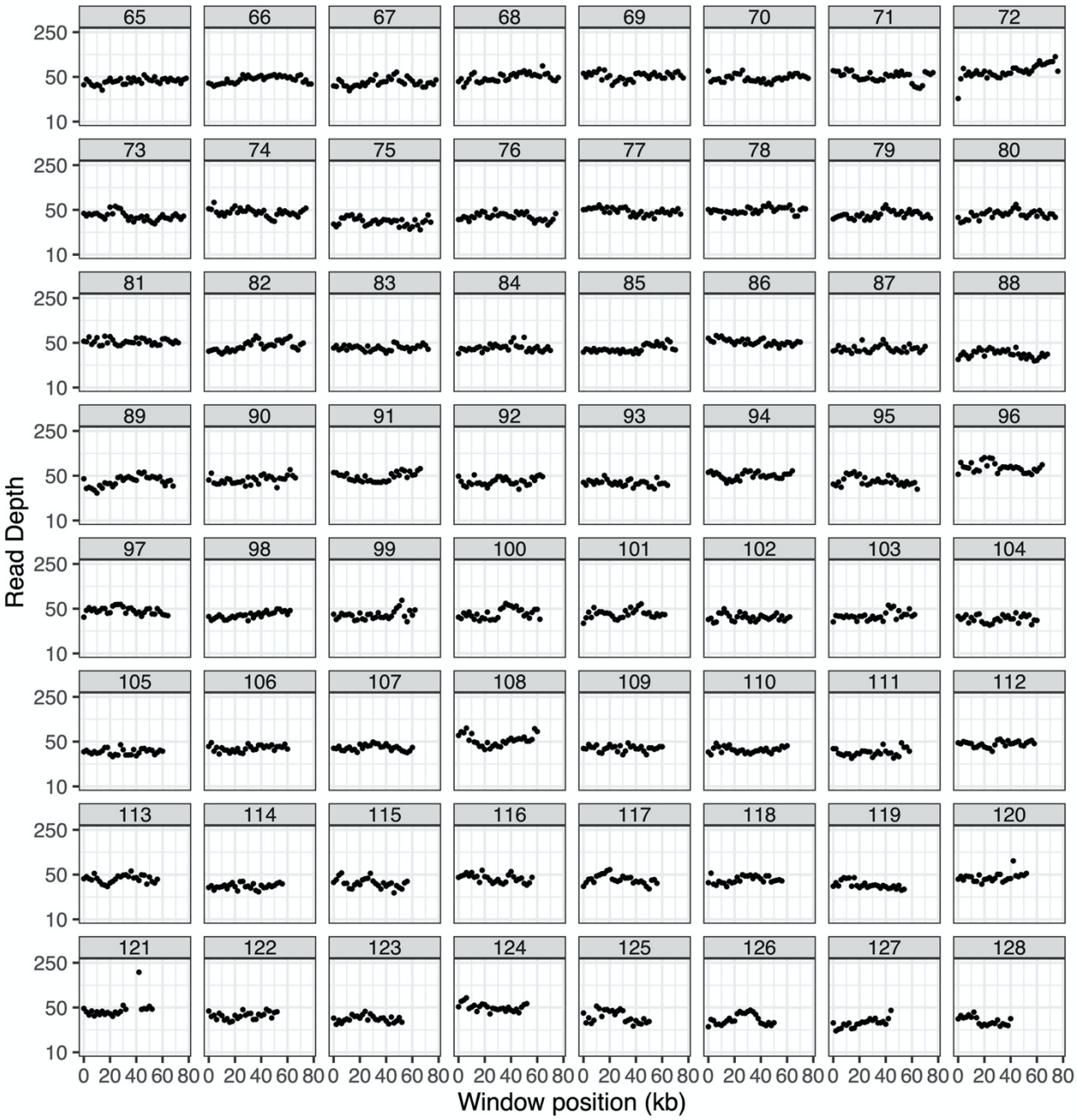

Figure S4b. Summary of coverage for mitochondrial chromosomes 65 through 128 in the *S. conica* ABR total-cellular sequencing library. Coverage estimated based on reads mapping with a maximum of one mismatch and no indels in 2-kb windows. Note that four of the data points on chromosome 121 exceeded a coverage of 250× and are not shown to improve readability of the plots.

Table S1. Summary of library sequencing and mapping

Library	Library Type	SRA Accession	Raw Read Pairs	Mito Coverage (bp)	Plastid Coverage (bp)	Mito Mapping (%)	Plastid Mapping (%)	Unmapped (%)
Silene conica chloroplast 1	Chloroplast duplex	SAMN17011097	1.15E+08	9.25E+07	2.03E+08	29.49%	64.88%	5.63%
Silene conica chloroplast 2	Chloroplast duplex	SAMN17011098	1.15E+08	1.33E+08	2.72E+08	31.48%	64.13%	4.39%
Silene conica chloroplast 3	Chloroplast duplex	SAMN17011099	1.42E+08	1.37E+08	2.70E+08	31.72%	62.55%	5.73%
Silene conica mito 1	Mitochondrial duplex	SAMN17011100	1.01E+08	2.18E+08	1.77E+07	80.87%	6.58%	12.55%
Silene conica mito 3	Mitochondrial duplex	SAMN17011101	1.14E+08	2.16E+08	2.01E+07	77.21%	7.16%	15.63%
Silene latifolia chloroplast 1	Chloroplast duplex	SAMN17011085	1.33E+08	2.89E+07	2.83E+08	7.39%	72.35%	20.26%
Silene latifolia chloroplast 2	Chloroplast duplex	SAMN17011086	1.28E+08	3.54E+07	2.41E+08	8.82%	60.10%	31.08%
Silene latifolia chloroplast 3	Chloroplast duplex	SAMN17011087	1.01E+08	3.68E+07	2.73E+08	9.13%	67.73%	23.14%
Silene latifolia mito 1	Mitochondrial duplex	SAMN17011088	1.13E+08	1.16E+08	2.67E+07	38.26%	8.83%	52.91%
Silene latifolia mito 2	Mitochondrial duplex	SAMN17011089	9.79E+07	4.93E+07	1.63E+07	24.87%	8.21%	66.92%
Silene latifolia mito 3	Mitochondrial duplex	SAMN17011090	8.83E+07	6.56E+07	2.23E+07	27.94%	9.48%	62.58%
Silene noctiflora chloroplast 1	Chloroplast duplex	SAMN17011091	1.37E+08	1.06E+08	2.55E+08	27.63%	66.35%	6.02%
Silene noctiflora chloroplast 2	Chloroplast duplex	SAMN17011092	1.25E+08	9.39E+07	2.20E+08	28.18%	66.12%	5.70%
Silene noctiflora chloroplast 3	Chloroplast duplex	SAMN17011093	1.37E+08	1.25E+08	3.19E+08	25.69%	65.36%	8.95%
Silene noctiflora mito 1	Mitochondrial duplex	SAMN17011094	1.32E+08	2.76E+08	1.56E+07	83.03%	4.68%	12.29%
Silene noctiflora mito 2	Mitochondrial duplex	SAMN17011095	1.06E+08	1.88E+08	1.22E+07	81.54%	5.28%	13.18%
Silene noctiflora mito 3	Mitochondrial duplex	SAMN17011096	1.03E+08	1.93E+08	2.39E+07	71.53%	8.84%	19.63%
Silene conica ABR	Total cellular shotgun	SAMN17011105	4.98E+08	3.53E+08	1.68E+09	0.49%	2.31%	97.20%
Silene noctiflora OSR	Total cellular shotgun	SAMN17011104	5.76E+08	3.45E+09	2.10E+09	2.51%	1.53%	95.96%
Silene latifolia UK2600	Total cellular shotgun	SAMN17011102	2.81E+08	5.95E+08	5.51E+09	0.50%	4.64%	94.86%
Silene latifolia Kew 32982	Total cellular shotgun	SAMN17011103	3.04E+08					

Table S2. *Silene* mitochondrial SNV count and frequency data with and without *k*-mer filtering

			SNV Counts							SNV Frequencies						
Species	Rep	k-mer Filt	A>C	A>G	A>T	C>A	C>G	C>T	Total	A>C	A>G	A>T	C>A	C>G	C>T	Total
Silene conica	1	10	53	356	26	13	6	18	472	3.0E-07	2.0E-06	1.5E-07	9.7E-08	4.5E-08	1.3E-07	1.5E-06
Silene conica	2	10	27	176	15	4	9	6	237	3.6E-07	2.3E-06	2.0E-07	6.9E-08	1.6E-07	1.0E-07	1.8E-06
Silene conica	3	10	67	462	17	13	6	13	578	3.3E-07	2.3E-06	8.5E-08	8.5E-08	3.9E-08	8.5E-08	1.6E-06
Silene conica	1	None	150	483	58	49	12	217	969	8.5E-07	2.7E-06	3.3E-07	3.6E-07	8.9E-08	1.6E-06	3.1E-06
Silene conica	2	None	64	238	24	20	12	98	456	8.5E-07	3.2E-06	3.2E-07	3.5E-07	2.1E-07	1.7E-06	3.4E-06
Silene conica	3	None	158	610	50	59	17	231	1125	7.9E-07	3.0E-06	2.5E-07	3.9E-07	1.1E-07	1.5E-06	3.2E-06
Silene latifolia	1	10	1	1	0	4	3	18	27	5.7E-09	5.7E-09	0.0E+00	3.0E-08	2.2E-08	1.3E-07	8.7E-08
Silene latifolia	2	10	1	0	1	2	2	5	11	1.3E-08	0.0E+00	1.3E-08	3.5E-08	3.5E-08	8.6E-08	8.2E-08
Silene latifolia	3	10	1	2	6	1	0	10	20	5.0E-09	1.0E-08	3.0E-08	6.5E-09	0.0E+00	6.5E-08	5.7E-08
Silene latifolia	1	None	9	8	6	17	5	42	87	1.1E-07	9.7E-08	7.3E-08	2.7E-07	8.1E-08	6.8E-07	6.0E-07
Silene latifolia	2	None	5	8	14	17	6	58	108	1.0E-07	1.7E-07	2.9E-07	4.7E-07	1.7E-07	1.6E-06	1.3E-06
Silene latifolia	3	None	9	16	18	19	10	52	124	1.5E-07	2.7E-07	3.1E-07	4.3E-07	2.3E-07	1.2E-06	1.2E-06
Silene noctiflora	1	10	3	5	1	0	4	3	16	1.3E-08	2.2E-08	4.4E-09	0.0E+00	2.6E-08	1.9E-08	4.2E-08
Silene noctiflora	2	10	1	8	0	0	7	5	21	6.0E-09	4.8E-08	0.0E+00	0.0E+00	6.1E-08	4.3E-08	7.4E-08
Silene noctiflora	3	10	2	7	0	2	4	6	21	1.1E-08	3.7E-08	0.0E+00	1.5E-08	3.1E-08	4.6E-08	6.6E-08
Silene noctiflora	1	None	3	8	3	4	6	12	36	1.3E-08	3.5E-08	1.3E-08	2.6E-08	3.8E-08	7.7E-08	9.4E-08
Silene noctiflora	2	None	2	11	1	2	13	9	38	1.2E-08	6.6E-08	6.0E-09	1.7E-08	1.1E-07	7.8E-08	1.3E-07
Silene noctiflora	3	None	5	13	2	11	9	11	51	2.7E-08	6.9E-08	1.1E-08	8.4E-08	6.9E-08	8.4E-08	1.6E-07

Table S3. Genome copy number estimates

Species	Nuclear (=Unmapped) Coverage (bp)	Mito Coverage (bp)	Plastid Coverage (bp)	Nuclear Genome Size (bp)	Mito Genome Size (bp)	Plastid Genome Size (bp)	Nuclear Cov.	Mito Cov.	Plastid Cov.	Mito per Nuclear	Plastid per Nuclear
Silene latifolia	7.58E+10	3.53E+08	1.68E+09	2.67E+09	2.53E+05	1.52E+05	29.22	1394.70	11047.66	47.72	378.02
Silene noctiflora	1.41E+11	3.45E+09	2.10E+09	2.78E+09	6.73E+06	1.52E+05	52.70	512.25	13838.01	9.72	262.59
Silene conica	1.20E+11	5.95E+08	5.51E+09	9.30E+08	1.13E+07	1.47E+05	136.49	52.55	37486.69	0.38	274.64

Table S4. Detection of SNVs in multiple biological replicates

Species	SNV Detection	A>C	A>G	A>T	C>A	C>G	C>T	Total
Silene conica	Shared among replicates	17	100	5	3	1	2	128
Silene conica	Unique to one replicate	32	157	28	19	19	21	276
Silene latifolia	Shared among replicates**	0	0	0	0	0	0	0
Silene latifolia	Unique to one replicate	3	3	7	7	5	33	58
Silene noctiflora	Shared among replicates	0	0	0	0	0	0	0
Silene noctiflora	Unique to one replicate	6	20	1	2	15	14	58

**The use of parentals for *k*-mer filtering in may have biased against detection of shared SNVs in *S. latifolia* (see main text).