Sorting of plant mitochondrial and plastid heteroplasmy is extremely rapid and depends on MSH1 activity

Amanda K Broz[1], Alexandra Keene[1], Matheus Fernandes Gyorfy[1], Mychaela Hodous[1], Iain G Johnston[2,3], Daniel B Sloan[1].

[1]Department of Biology, Colorado State University, Fort Collins, CO, USA
[2]Department of Mathematics, University of Bergen, Bergen, Norway
[3]Computational Biology Unit, University of Bergen, Bergen, Norway

Author for correspondence:
Amanda K Broz
Email: amandabroz@gmail.com

Abstract:

Because mitochondrial and plastid genomes exist at high cellular copy numbers, new mutations result in the co-occurrence of multiple alleles within a cell or individual, a phenomenon known as heteroplasmy. The extent to which heteroplasmies are transmitted across generations or eliminated through a sorting process involving successive genetic "bottlenecks" is not well understood in plants, in part because their low mutation rates make these variants so infrequent. We previously found that disruption of *MutS Homolog 1* (*MSH1*), a nuclear gene involved in plant organellar DNA repair, results in numerous de novo point mutations, providing an unprecedented opportunity to explore the inheritance of these variants. Here, we used droplet digital PCR to quantitatively track the inheritance of single nucleotide variants (SNVs) in mitochondrial and plastid genomes of both wild-type Arabidopsis and *msh1* mutants. We found that heteroplasmic sorting was rapid for both organelles, greatly exceeding the rates at which mitochondrial heteroplasmies are eliminated in animals. In *msh1* mutants, plastid SNVs usually reached fixation or were lost within a single generation, indicating an effective transmission bottleneck size (N) of ~1. In contrast, mitochondrial heteroplasmies often persisted across generations but still exhibited a relatively narrow transmission bottleneck ($N \approx 4$). Restoring MSH1 function in mitochondria increased the rate of heteroplasmic sorting ($N \approx 1.3$), potentially due to its hypothesized role in promoting gene conversion as a mechanism of DNA repair, which is expected to homogenize genome copies within a cell. We also found that heteroplasmic sorting favored GC base pairs. Therefore, recombinational repair and gene conversion in plant organellar genomes may accelerate the elimination of heteroplasmies and bias the outcome of this sorting process.

Introduction:

In plants, the genetic system is housed in three different compartments: the nucleus, the mitochondria and the plastids. The products of mitochondrial and plastid genomes perform functions critical for cellular metabolism, including oxidative phosphorylation and photosynthesis. Because organellar genomes are present at high cellular copy numbers, multiple alleles can coexist within a cell, a situation known as heteroplasmy. These variants can create opportunities for selfish competition within cells [1-4] and are of great interest because they are often associated with human disease phenotypes, including inherited disorders due to germline transmission and age-related disorders due to heteroplasmies in somatic tissues [5-7]. Because of their important health consequences, the dynamics by which de novo mitochondrial point mutations in mammals spread from initially low frequencies to eventually reach fixation (homoplasmy) within a cell have been investigated in detail. Mammalian mitochondrial genomes undergo physical and genetic bottlenecks that increase variance in heteroplasmic alleles among cells, providing a basis for selection [8-10]. The relative size of these bottlenecks can be calculated by comparing variance in heteroplasmic frequencies between mother and progeny [11, 12], and effective mitochondrial bottleneck size (i.e., modelling the heteroplasmy variance generated between generations as a single sampling event) ranges from ~10-30 segregating units in humans [13-16].

In contrast to the fairly detailed understanding of mitochondrial heteroplasmy in animal systems, there are fundamental gaps in our knowledge of how mitochondrial and plastid mutational variants sort out in plants. Studies of heteroplasmic frequency in plants have mostly been done with species that show biparental organelle inheritance [17-20], presumably because the exceedingly low point mutation rates in plant organellar genomes [21-23] limit the supply of *de novo* mutations to study. In addition, spontaneous organellar mutations that lead to visible phenotypes in plants (i.e., plastid mutations causing chlorosis or variegation) tend to be severe, and heteroplasmic lines can often only be maintained by vegetative propagation [24, 25]. These factors have hampered the study of heteroplasmy dynamics in plants, including the extent to which heteroplasmies are transmitted across generations.

The process by which heteroplasmies arise and spread is expected to differ markedly from the accumulation of mutations in the nuclear genome because of three distinguishing characteristics: mode of inheritance, copy number per cell and mutation rate. Organellar genomes show non-Mendelian inheritance, typically characterized by maternal transmission as opposed to the biparental inheritance of nuclear genomes in sexual organisms. Strict uniparental inheritance does not allow for the generation of novel combinations of alleles

through recombination and has historically been expected to result in a buildup of deleterious mutations, through a process known as Muller's ratchet [26, 27]. However, there is accumulating evidence that biparental inheritance of organellar genomes (sometimes referred to as paternal leakage) is more common than previously appreciated [19, 20, 28, 29]. Even infrequent biparental inheritance of organellar DNA could represent a pathway for recombination, slowing or preventing the 'mutational meltdown' associated with Muller's ratchet [30]. Alternatively, recent theory posits that the bottlenecking associated with uniparental inheritance of organellar genomes may actually provide a benefit by increasing cell-to-cell variability and improving the efficiency of selection at higher organizational levels [31-34]. This benefit may explain why uniparental inheritance of organellar genomes has been retained across eukaryotic lineages.

The number of genome copies per cell is another major difference between the nucleus and organelles that impacts the spread of new mutations. In contrast to the single nuclear genome copy that is passed to the next generation in each gamete, cells can contain numerous mitochondria and plastids, and genome copies within each organelle can reach high numbers. As such, a mutation arising in organellar DNA initially constitutes only a single member of a larger population of non-mutant genome copies in the cell. The size of this population is expected to influence the subsequent genetic dynamics. In Arabidopsis, only a few plastids are present in meristematic cells, while over 50 plastids can be found in cells of mature leaves, and the number of plastid genome copies per cell ranges from around 80 in meristematic tissue to >3,000 in mature leaves [35]. Mitochondrial genomes are present at ~50-100 copies in both egg cells and leaf cells of Arabidopsis [36-38]. However, in both cell types, the number of mitochondria per cell exceeds 300 suggesting that many or most mitochondria do not contain a full mitochondrial genome copy [37]. The number of organellar genome copies in female gametes of plants varies between species but is typically low, not exceeding 100 mitochondrial genome copies [38]. In comparison, a single mouse oocyte contains over 200,000 mitochondrial genomes, the result of proliferation from ~200 mitochondrial DNA copies found in progenitor germ cells of the embryo [8].

Mutation rates also differ among cellular compartments. In land plants, mitochondria have the lowest mutation rate (as inferred from synonymous substitutions), followed by plastids and then nuclei at a ratio of approximately 1:3:10 [21-23]. By comparison, mammalian mitochondrial sequence mutation rates greatly exceed those in the nucleus [23]. Although plant mitochondria and plastids have low rates of point mutation, their genomes undergo frequent recombination between repeated sequences, resulting in populations of alternative structures

[39-41]. Thus, reversible structural heteroplasmies can exist in plant mitochondria and have been associated with cytoplasmic male sterility and other types of phenotypic variation [39, 42, 43].

Our previous work found that the nuclear-encoded protein MutS Homolog 1 (MSH1) reduces plant organellar mutation rates [44], in addition to its previously characterized role in suppression of ectopic recombination [45, 46]. MSH1 is part of the larger MutS family of genes involved in mismatch repair and has a unique architecture that includes both mismatch recognition and endonuclease domains [47]. It has been hypothesized that MSH1 initiates double strand breaks (DSBs) at mismatched or damaged bases to facilitate repair through homologous recombination [44, 48, 49]. The high ploidies frequently associated with organelles provide multiple genome copies for homologous recombination which plays a major role in DNA replication and repair processes in plant organellar genomes [40, 50]. Gene conversion is a common outcome of DSBs and homologous recombination, resulting in homogenization of genome copies within a cell [40, 50]. It has been hypothesized that gene conversion could act as an alternative mechanism to increase cell-to-cell variance in heteroplasmic frequencies in eukaryotic lineages such as plants that may lack a physical bottleneck associated with germline development [33]. Thus, we would predict the action of *MSH1* and other genes involved in homologous recombination may accelerate the sorting of heteroplasmies.

We previously identified numerous mitochondrial and plastid heteroplasmies in Arabidopsis *msh1* mutant lines and showed that some of these could be transmitted through meristematic and reproductive tissues to subsequent generations [44]. Having this unique genetic material provided us the opportunity to study the dynamics of *de novo* heteroplasmies in plants, both within individuals and across generations in *msh1* mutants and wild type backgrounds. We find that heteroplasmic sorting is rapid in plants – particularly in plastids – and that MSH1 function in mitochondria increases the speed of heteroplasmic sorting. For heteroplasmic variants, we also found that GC base pairs preferentially increased in frequency over AT base pairs. These results imply that gene conversion contributes to a high rate of heteroplasmic sorting in plants and potentially biases the outcome of this sorting process.

Results:

Identification of heteroplasmic variants in msh1 mutants

The *msh1* mutant background afforded us the opportunity to explore the dynamics of heteroplasmy in Arabidopsis. We selected ten high frequency single nucleotide variants (SNVs) resulting from de novo mutations that were identified in organellar DNA pools of *msh1* mutant

125 plants (Table 1). Each SNV was associated with a specific *msh1* family line. All of the SNVs
126 were transitions (GC→AT or AT→GC mutations), and the majority were in intergenic regions.
127 The frequencies for each SNV measured by allele-specific droplet digital PCR (ddPCR) were
128 repeatable and similar to those obtained from sequencing read counts (Table 1).
129 We used this ddPCR assay to identify heteroplasmy in individual plants by screening
130 young leaves from the siblings and/or progeny of the plants used in the initial identification of
131 SNVs [44]. After adjusting for organellar genome copy number and numts (see Methods), only
132 four of the ten SNVs were found to be heteroplasmic in any of the screened individuals; the
133 other six variants had either reached fixation or were not detectable in the individuals screened
134 (Table 1, Table S1). Therefore, we proceeded with these four heteroplasmic SNVs (two
135 mitochondrial and two plastid) for all subsequent experiments and analyses. The number of
136 heteroplasmic individuals identified was higher for the mitochondrial SNVs tested (46.5% of
137 individuals) versus those from the plastid (3.7%) (Table S1), which is why a greater number of
138 markers and total plants were screened for plastid versus mitochondrial SNVs.

Table 1. Characteristics of SNV targets.

SNV Target (nucleotide position)	Reference Nucleotide	Variant Nucleotide	Genetic Context	Variant Frequency (sequencing)*	Variant Frequency (ddPCR)*	Heteroplasmy Identified
Plastid (26553)	A	G	intergenic	0.076	0.0926 (0.0080)	Yes
Plastid (29562)	T	C	intergenic	0.032	0.026	No
Plastid (36873)	A	G	intergenic	0.059	0.032 (0.0024)	Yes
Plastid (48483)	A	G	intergenic	0.028	0.033	No
Plastid (61599)	A	G	intergenic	0.068	0.090	No
Plastid (72934)	A	G	*psbB* - CDS	0.023	0.017	No
Plastid (118559)	T	C	*ndhG* - CDS	0.014	0.015 (0.0082)	No
Mito (91017)	A	G	intergenic	0.147	0.180	Yes
Mito (143184)	T	C	*rrn26* - rRNA	0.014	0.012	No
Mito (334038)	C	T	intergenic	0.033	0.019 (0.0018)	Yes

*Variant frequencies are calculated as the count of variant alleles out of the total copy number in the sample. Variants were initially identified from Duplex Sequencing of purified organellar DNA from a pool of ~60 F3 plants [44]. These same samples were used to estimate variant frequency with ddPCR. Standard deviation for ddPCR experiments is shown in parentheses for the SNVs used in subsequent studies, and seven to ten technical replicates were used for calculation of standard deviations.

139 ***Intergenerational heteroplasmic sorting occurs more rapidly in plastids than***
140 ***mitochondria in an msh1 mutant background***
141 To determine the frequency with which heteroplasmies are transmitted across
142 generations, we used *msh1* individuals that were each heteroplasmic for one of the four

identified SNVs as mothers to generate selfed progeny. We expected that variant frequencies in the mother would influence the distribution of heteroplasmy in the progeny, so we used mothers with a wide range of starting allele frequencies. Both mitochondria and plastid SNVs showed rapid sorting over a single generation, as many progeny were fixed for either the wild type or alternative (SNV) allele (Figure 1, Table S2). This trend was particularly striking in plastids: very few heteroplasmic progeny were identified for plastid SNVs, regardless of the heteroplasmic frequency in the mother. For mitochondrial SNVs, progeny showed a tighter and more continuous distribution of heteroplasmy roughly distributed around the allele frequency of the mother, with many progeny retaining a heteroplasmic state.

The distribution of heteroplasmy in progeny derived from a heteroplasmic mother can be used to calculate an effective transmission bottleneck size (N). Although this measure has often been (incorrectly) equated with the number of organellar genomes transmitted to the next generation, it is more appropriately thought of as a relative metric to compare the cumulative biological sampling variance in genome transmission throughout development and across generations [11, 12]. A common way of estimating bottleneck size is by taking the reciprocal normalized sample variance, $\hat{N}_{sv} \simeq \bar{h}\,(1-\bar{h})/s^2$, as an estimate for the number of effective segregating units [11], where $\bar{h}$ is the sample mean heteroplasmy and s^2 is the sample variance in heteroplasmic frequencies for a set of tissue or progeny samples. However, the use of this statistic is problematic for moderate samples of highly segregated cases, where the sample variance approach yields estimates and uncertainties that are incompatible with the concept of segregating units (for example, inferring $\hat{N}_{sv} < 1$ with confidence intervals that include zero) [51]. We took two approaches here to estimate bottleneck size. First, for comparison with previous work, we used the above equation to calculate $\hat{N}_{sv}$, but we employed the population variance (without Bessel's correction) for s^2. This approach yields a slightly biased estimate but one that respects heteroplasmy constraints ($\hat{N}_{sv} \geq 1$). More rigorously, we also used a maximum likelihood approach based on the Kimura distribution [51] that captures these constraints in order to identify the most likely value ($\hat{N}$) of the bottleneck parameter as well as the confidence interval (CI) of this measure given a full set of observations. We also used this Kimura-based approach to perform hypothesis testing (see Methods). Both values ($\hat{N}_{sv}$ and $\hat{N}$) are reported in tables for comparison, but we refer to the more rigorous Kimura estimates in the text.

Using data from seven plastid families and five mitochondrial families we found that, over a single generation, plastids typically have a smaller transmission bottleneck size than mitochondria (Table 2). Plastid transmission bottlenecks were extreme, approximating a value of one ($\hat{N}$ = 1.06, 95% CIs: 1.03-1.13), in accordance with observations that the vast majority of

177 offspring were homoplasmic for one allele or the other (Figure 1). The average mitochondrial
178 transmission bottleneck size in *msh1* mutants ($\hat{N}$ = 4.20, 95% CIs: 4.15-4.25) was significantly
179 larger than in plastids ($p = 3.2 \times 10^{-55}$, likelihood ratio test, Table 2). Therefore, the number of
180 effective genome copies passed throughout development and from mother to progeny is larger
181 for mitochondria than for plastids. However, the transmission bottleneck values for both
182 organelles reflect relatively rapid heteroplasmic sorting.

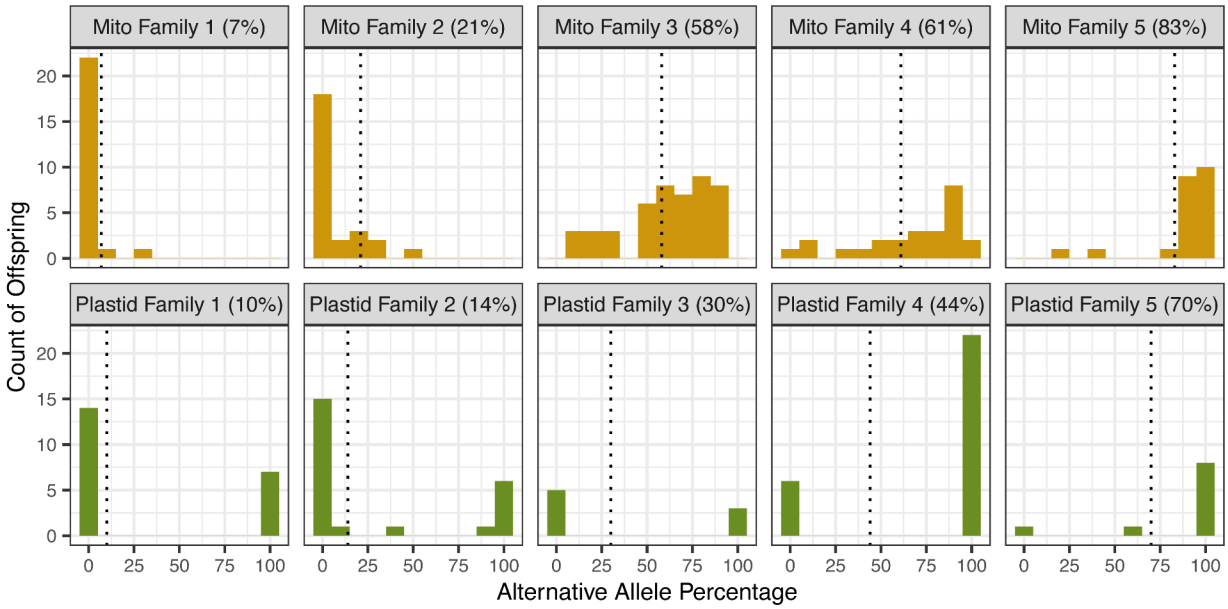

Figure 1. Distribution of heteroplasmy across generations in an *msh1* mutant background. Heteroplasmy (alternative allele frequency) was evaluated using ddPCR in progeny of maternal lines with different levels of heteroplasmy, indicated by dotted vertical lines and numbers in parentheses on each graph. Histograms indicate the number of individuals showing different levels of heteroplasmy. Mitochondrial families (orange) are depicted in the top row, plastid families (green) are shown on the bottom row. For the sake of comparison with the five mitochondrial families, data for plastid families 6 and 7 in not shown but is included in Table 2.

Table 2. Effective transmission bottlenecks in organellar genomes in the *msh1* mutant background calculated for plastid and mitochondrial markers over single generations in *msh1* mutant plants. Confidence intervals for Kimura estimates are noted in parentheses below calculated bottleneck value. Likelihood ratio test (see methods) show that plastid bottlenecks are significantly smaller than mitochondrial bottlenecks ($p = 3.2 \times 10^{-55}$).

Family name	SNV target	Maternal sample % alternative allele	Progeny mean % alternative allele ($\bar{h}$)	Number progeny (n)	Variance in heteroplasmy (s^2)	Bottleneck size estimate from s^2 ($\hat{N}_{sv}$)	Bottleneck size estimate from Kimura fit ($\hat{N}$, 95% CIs)
Mito Family 1	Mito 91017	7	2	24	0.0052	3.54	2.81 (1.38-9.71)
Mito Family 2	Mito 334038	21	8	26	0.0186	3.79	3.18 (2.04-5.54)
Mito Family 3	Mito 91017	58	61	47	0.0573	4.14	4.68 (3.52-6.37)
Mito Family 4	Mito 91017	61	68	25	0.0822	2.65	3.33 (3.11-3.57)
Mito Family 5	Mito 91017	83	89	22	0.0356	2.69	6.14 (5.69-6.62)
						Mito mean 3.36	Joint mito 4.20 (4.15-4.25)
Plastid Family 1	Plastid 26553	10	33	21	0.2222	1.00	1 (1-1)
Plastid Family 2	Plastid 26553	14	31	24	0.1934	1.10	1.11 (1.03-1.35)
Plastid Family 3	Plastid 36873	30	38	8	0.2344	1.00	1 (1-1)
Plastid Family 4	Plastid 26553	44	79	28	0.1684	1.00	1 (1-1)
Plastid Family 5	Plastid 26553	70	86	10	0.0986	1.25	1.15 (1.02-2.25)
Plastid Family 6	Plastid 26553	7	16	20	0.1034	1.29	1.22 (1.06-1.76)
Plastid Family 7	Plastid 36873	7	0	20	0	NA	NA
						Plastid mean 1.07	Joint plastid 1.06 (1.03-1.13)

Heteroplasmic sorting in vegetative and floral tissues occurs more rapidly in plastids than mitochondria in an msh1 mutant background

Although reproductive tissue is one location where heteroplasmic sorting can happen, it can also occur within cells and tissues as they grow. This may be particularly important in organisms like plants that do not exhibit early germline specification [52, 53], a hypothesis that is supported by recent theory predicting ongoing segregation in plant tissues [33]. We sampled multiple leaf and inflorescence tissues from selected progeny of *msh1* heteroplasmic mothers and quantified heteroplasmic rates (Figure 2, Table S3). Two plastid family lines (six individuals) initially selected for analysis showed no within-plant heteroplasmy and were either fixed for the wild-type or alternative SNV allele. The other family line showed high variance in heteroplasmy rates across tissues, indicative of rapid within-plant sorting. Using rates of heteroplasmy in the tissue samples from each plant, we found that within-plant bottleneck sizes were significantly

195 smaller on average for plastids ($\hat{N}$ = 5.92, 95% CIs: 3.85-9.51) than for mitochondria ($\hat{N}$ = 12.00,
196 CIs: 11.95-12.05, $p = 3.4 \times 10^{-29}$, likelihood ratio test; Table 3). Within-plant bottleneck size was
197 larger than the transmission bottleneck size seen between generations (plastids, $p = 2.4227 \times$
198 10^{-29}; mitochondria, $p = 1.6 \times 10^{-30}$, likelihood ratio tests), suggesting that sorting during
199 vegetative growth indeed contributes to heteroplasmic variance, but the full reproductive cycle
200 and transmission to the next generation involves a tighter effective bottleneck.

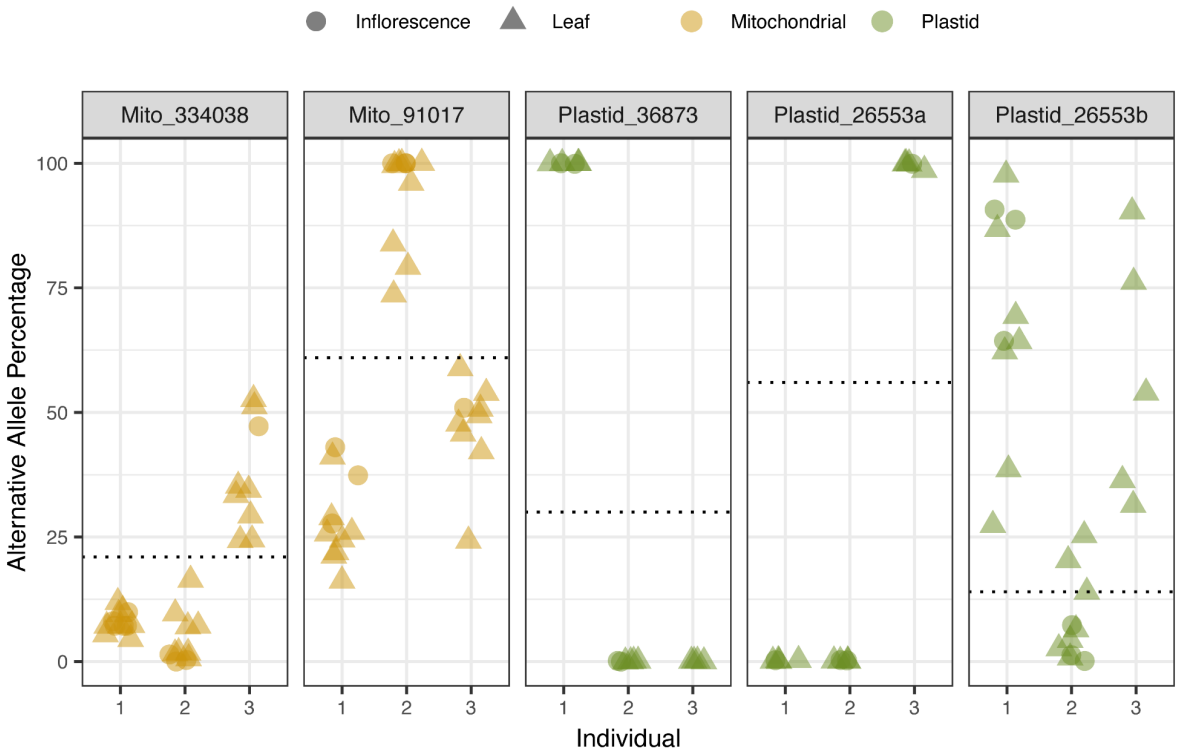

Figure 2. Distribution of heteroplasmy within individual plants in an *msh1* background. Progeny from mothers with varying levels of heteroplasmy (experiment shown in Figure 1) were selected for further analysis. Levels of heteroplasmy (alternative allele percentage) for leaf and inflorescence tissues were determined using ddPCR.

Table 3. Within-plant bottlenecks in organellar genome transmission in *msh1* mutant backgrounds. Confidence intervals for Kimura estimates are noted in parentheses below calculated bottleneck value. Not applicable (NA) indicates that there was no variation within the tissues tested. Likelihood ratio test (see methods) show that within plant plastid bottlenecks are significantly smaller than mitochondria bottlenecks ($p = 3.4 \times 10^{-29}$).

Individual name	SNV target	Mean within-plant % alternative allele ($\bar{h}$)	Number tissue samples (n)	Variance in heteroplasmy (s^2)	Within-plant Bottleneck size estimate from s^2 ($\hat{N}_{sv}$)	Within-plant bottleneck size from Kimura fit ($\hat{N}$, 95% CIs)
Mt 3 - 2	Mito 334038	4	11	0.0024	16.87	18.82 (9.75-37.28)
Mt 3 - 1	Mito 334038	8	13	0.0004	192.46	195.83 (190.48-201.32)
Mt 9 - 1	Mito 91017	29	9	0.0105	22.19	23.06 (18.5-28.81)
Mt 3 - 3	Mito 334038	37	11	0.0066	30.88	32.28 (29.65-35.15)
Mt 9 - 3	Mito 91017	47	9	0.0085	29.27	28.17 (22.17-35.88)
Mt 9 - 2	Mito 91017	94	11	0.0090	6.37	4.17 (2.08-10.33)
					Mito mean 49.72	Joint mito 12.00 (11.95-12.05)
Pt 2a - 1	Plastid 26553	0	5	0	NA	NA
Pt 2a - 2	Plastid 26553	0	6	0	NA	NA
Pt 3 - 2	Plastid 36873	0	6	0	NA	NA
Pt 3 - 3	Plastid 36873	0	4	0	NA	NA
Pt 2b - 2	Plastid 26553	8	10	0.0069	10.91	12.48 (6.75-23.94)
Pt 2b - 3	Plastid 26553	58	5	0.0513	4.76	5.18 (2.28-14.65)
Pt 2b - 1	Plastid 26553	69	10	0.0475	4.50	5.22 (2.93-10.22)
Pt 2a - 3	Plastid 26553	100	5	0	NA	NA
Pt 3 - 1	Plastid 36873	100	6	0	NA	NA
					Plastid mean 6.72	Joint plastid 5.92 (3.85-9.51)

MSH1 activity accelerates heteroplasmic sorting in mitochondria

Because MSH1 is hypothesized to introduce DSBs and promote recombinational repair [44, 48, 49], we predicted that a functional copy of the *MSH1* gene would speed up heteroplasmic sorting by homogenizing genome copies through gene conversion, as predicted by theoretical modeling [33]. To test this hypothesis, we transferred heteroplasmic variants to a wild-type background by crossing heteroplasmic *msh1* female plants with wild-type males and analyzing heteroplasmy levels in both F1 and F2 (selfed) progeny. All plants were genotyped at the *MSH1* locus, and only individuals that were heterozygous or homozygous wild type were included in

209 the heteroplasmy analysis (the *msh1* mutation is recessive). Due to the extremely low number
210 of individuals that were heteroplasmic for plastid SNVs and the rapid plastid heteroplasmic
211 sorting rates, this backcrossing method was only successful in generating lines to study
212 mitochondrial heteroplasmy. For mitochondrial SNVs in the wild-type background, we saw
213 extremely rapid sorting of heteroplasmies (Figure 3, Table S2, Table S4), akin to our
214 observations in plastids under *msh1* mutant backgrounds (Figure 1). The average mitochondrial
215 transmission bottleneck size for wild-type plants ($\hat{N}$ = 1.33, 95% CIs: 1.20-1.54) was significantly
216 reduced from that in the *msh1* mutant background ($\hat{N}$ = 4.20, 95% CIs: 4.15-4.25, $p = 8.1 \times 10^{-}$
217 ^{49}, likelihood ratio test; Table 2, Table 4). The standing number of organellar genomes per
218 nuclear genome copy in leaf tissue did not differ significantly between *msh1* and wild type
219 backgrounds (Figure S1), suggesting that the differences we identified in heteroplasmic sorting
220 were more likely due to differences in MSH1 activity rather than changes in the physical number
221 of organellar DNA copies per cell.

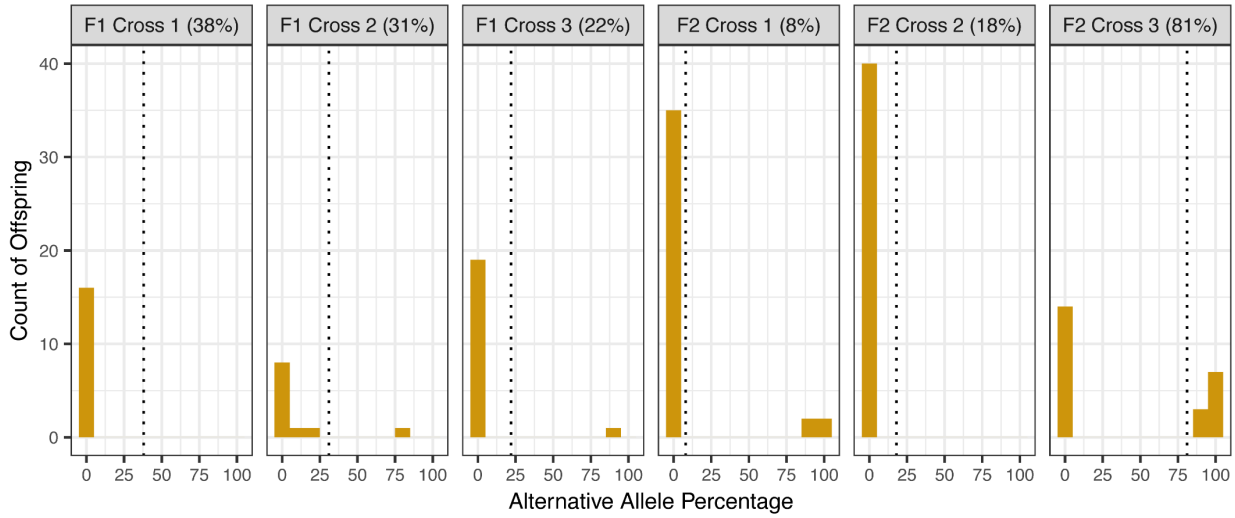

Figure 3. Distribution of mitochondrial heteroplasmy across generations in a wild-type background. Mitochondrial heteroplasmies were backcrossed into a wild-type *MSH1* background. Heteroplasmy (alternative allele frequency) was evaluated using ddPCR in progeny of maternal lines with different levels of heteroplasmy, indicated by vertical dotted lines and numbers in parentheses on each graph. Histograms indicate the number of individuals showing different levels of heteroplasmy.

Table 4. Effective transmission bottlenecks in organellar genomes in wild type background. Effective transmission bottleneck size was calculated for mitochondrial markers over single generations in wild type plants. Confidence intervals for Kimura estimates are noted in parentheses below calculated bottleneck value. Not applicable (NA) indicates that there was no variation within the progeny – they were all fixed for one allele or the other. Note that bottleneck size estimates for F1 cross 1 are imprecise because they are based on a family in which only one of 17 progeny remained heteroplasmic and that individual had a low allele frequency (2%). The large point estimates for that family weigh heavily when taking a mean of $\hat{N}_{sv}$ estimates but do not bias the joint maximum-likelihood estimate of $\hat{N}$.

Cross type	SNV target	Maternal sample % alternative allele	Progeny mean % alternative allele ($\bar{h}$)	Number progeny (n)	Variance in heteroplasmy (s^2)	Bottleneck size estimate from s^2 ($\hat{N}_{sv}$)	Bottleneck size estimate from Kimura fit ($\hat{N}$, 95% CIs)
F1 cross 1	Mito 334038	38	0.14	17	0.00003	46.13	18.18 (2.45-204.07)
F1 cross 2	Mito 334038	31	8	10	0.05867	1.63	1.94 (1.26-4.45)
F1 cross 3	Mito 91017	22	5	20	0.03708	1.15	1.58 (1.13-3.67)
F2 cross 1	Mito 334038	8	18	32	0.13963	1.05	1.26 (1.11-1.62)
F2 cross 2	Mito 334038	18	0	37	0.00000	NA	NA
F2 cross 3	Mito 334038	81	36	24	0.21511	1.07	1.24 (1.11-1.54)
						Mean 10.21	Joint estimate 1.33 (1.20-1.54)

GC-biased inheritance in organellar genomes

The fact that the SNVs used in our analysis were GC→AT or AT→GC transitions created the opportunity to search for GC or AT bias in plant organellar genome inheritance. Using the SNV data from our heteroplasmic mothers and progeny, we found evidence of a GC bias during heteroplasmic sorting in both plastids and mitochondria (Figure 4, Table S5). The frequency of the GC allele increased in the progeny relative to the mother in 14 of 17 families (two-sided binomial test, $p = 0.0127$). Although modest increases in the frequency of the GC allele were found in most lines, five showed that mean progeny GC frequency increased over 20% compared to the mother (Figure 4B). The average increase in frequency for the GC allele was 14.1%, which differed significantly from zero (two-sided t-test, $p = 0.0039$).

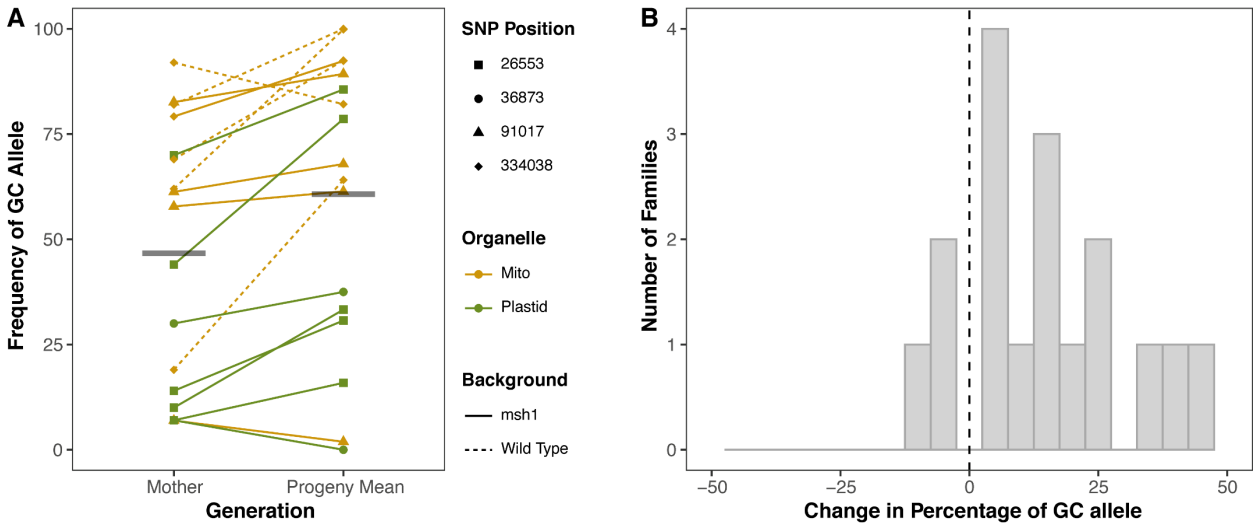

Figure 4. **Changes in GC frequency between generations.** A) Norm of reaction plot showing the frequency of GC alleles (SNV) in mothers and progeny (mean value). Means for each group are shown with horizontal gray bars. B) Histogram showing the change in percentage of GC alleles for tested family lines. A two-sided t-test shows bias toward GC alleles, $p = 0.0039$.

Discussion:

Potential causes of rapid heteroplasmic sorting in plant organelles

In Arabidopsis, we found that both plastid and mitochondrial heteroplasmies sorted out to homoplasmy within one to a few generations and exhibited tight effective bottlenecks, regardless of whether they were found in wild type or *msh1* mutant backgrounds (*msh1* plastid $\hat{N} \approx 1$, *msh1* mitochondria $\hat{N} \approx 4$, wild-type mitochondria $\hat{N} \approx 1.3$). In contrast, effective mitochondrial germline bottlenecks are estimated to be substantially larger in most animal systems: ~5-10 in *Daphnia* [54], ~10-30 in humans [13-16], ~30 in *Drosophila* [55], ~60 in *Caenorhabditis elegans* [56], ~80 in salmon [57], 170 in zebrafish [58] and 5-100 in mice [9, 51]. Thus, in animals, heteroplasmy is often retained over multiple generations.

Germline bottlenecks are well established as drivers of heteroplasmic variance in animals, but the mechanism is still not completely understood [8, 9, 11, 15]. The bottleneck is thought to be due, at least in part, to the physical reduction of mitochondrial DNA copies during female germline development [59, 60]. However, other lines of evidence suggest that selection [61-63], mitochondrial dynamics [64], and preferential genome amplification [65, 66] can also play roles in reducing the effective population size of mitochondrial genomes.

In plants, where the germline is segregated later in development [52, 53, 67], the way in which heteroplasmic variance increases is even less clear. However, the relatively modest number of mitochondrial and plastid genome copies in meristematic and reproductive tissues is

251 a likely contributor to rapid heteroplasmic sorting. In plants, all aboveground tissues, including
252 reproductive organs, are derived from the shoot apical meristem (SAM). Arabidopsis SAM cells
253 contain only ~80 copies of plastid DNA housed within 4-10 proplastids [35]. The number of
254 mitochondrial genome copies in SAM cells is not well established, but these values are typically
255 <100 in both vegetative tissues and egg cells [36-38]. In addition, the frequent recombination
256 associated with plant organellar genomes [40, 50] could act as an effective bottleneck by
257 homogenizing genomes through gene conversion without requiring a physical reduction in
258 organelle DNA copy number, as predicted by recent theory [33]. The rapid heteroplasmic sorting
259 in plant organelles, as well as our finding that MSH1 activity further accelerates the rate of
260 sorting (see below), supports this hypothesis.
261 Although the rate of heteroplasmic sorting has not been studied extensively in plants,
262 there are supporting lines of evidence that this process may be rapid in other angiosperms. For
263 example, controlled crosses of *Silene vulgaris* and *Daucus carota* have shown that
264 mitochondrial heteroplasmy is maternally transmitted at only low levels across generations [17,
265 18]. Mitochondrial variants produced by repeat-mediated recombination have been observed to
266 rapidly rise from mean cellular copy numbers under one to high frequencies or even
267 homoplasmy across cells, through a process called substoichiometric shifting (SSS) [39, 42].
268 Because of the role of *MSH1* in recombination surveillance, these structural variants arise more
269 frequently in *msh1* mutants [46], but the causes of rapid SSS have remained less clear. Studies
270 of variegation mutants derived from biparental plastid inheritance show that sorting occurs
271 rapidly at the level of whole organelles and is frequently complete within a single generation [12,
272 25, 68]. When plastid DNA is modified using a transgenic approach, homoplasmy is typically
273 reached after a few rounds of antibiotic selection [69, 70]. The rapid sorting of transgenic plastid
274 mutations has also been observed in the absence of antibiotics [71]. It is currently unclear
275 whether rapid sorting of plastid transgenes occurs in all plants as it has been extremely difficult
276 to obtain homoplasmic transgenic lines of various species, particularly monocots [72-74].
277 However, this may be due to the general recalcitrance of these species to plastid transformation
278 or inefficient selection that makes variants difficult to detect [72, 73]. The approach we have
279 used here would be valuable to test heteroplasmic sorting in other species because most
280 intergenic SNVs are unlikely to be under strong selective pressure, as opposed to the entire
281 genes that are introduced with transgenic approaches or the large structural rearrangements
282 associated with SSS. The rapid sorting we have identified could readily explain these diverse
283 instances of segregation and occasional amplification of rare variants in organellar genomes.
284

285 ***Differences in heteroplasmic sorting between mitochondria and plastids***

286 In the *msh1* mutant background, we found that plastids had tighter bottlenecks than
287 mitochondria, both across generations ($\hat{N}$ = ~1 versus ~4, respectively) and within individuals ($\hat{N}$
288 = ~6 versus ~12, respectively). This result may seem surprising because the much greater
289 relative copy number of plastid versus mitochondrial DNA typically seen in most aboveground
290 tissues [35, 37] would be expected to result in lower levels of heteroplasmic variance in plastids
291 (i.e., a wider bottleneck size). However, as noted above, mitochondria and plastids both
292 possess low genome copy numbers in the SAM and/or reproductive tissues [35-38], presenting
293 similar potential for physical bottlenecks. One key difference between the two types of
294 organelles is that mitochondria can experience both full and transient fusion events over the
295 plant life cycle allowing for genetic exchange, whereas plastids rarely if ever fuse [12, 75-79].
296 Mitochondrial fusion is particularly prevalent in the SAM, where it is estimated that 80% of the
297 mitochondrial volume is fused into a dynamic tentaculate cage-like structure, creating the
298 opportunity for sharing of genome copies between formerly distinct mitochondrial compartments
299 [76]. Here, gene conversion is predicted to homogenize the mitochondrial genomes within a cell
300 leading to increased heteroplasmic variance between cells [33, 76]. Thus, all else being equal,
301 existing theory would predict that fusion results in more rapid heteroplasmic sorting in
302 mitochondria than plastids – the opposite of what we observed. This suggests that other
303 differences in organelle biology can impact the rate of variability in organellar DNA populations.

304 Additional factors that are expected to influence the speed of heteroplasmic sorting, but
305 the extent to which they differ between plastids and mitochondria is often unclear. These
306 processes include the number of organelles per cell, rates of organellar DNA replication and
307 degradation, rates of organelle turnover, the physical partitioning of organellar DNA during
308 organelle replication, and the partitioning of whole organelles during cell division [12]. In
309 addition, plastids and mitochondria house distinct versions of many DNA replication,
310 recombination and repair proteins [50], which could influence relative rates of gene conversion.
311 It is also possible that dual-targeted proteins involved in these pathways, such as MSH1, may
312 differentially impact gene conversion rates between organelles. All of these factors likely vary
313 based on cell type and developmental stage, and they may work in combination. Models of
314 heteroplasmic sorting have tried to integrate many of these factors, but much of the relevant
315 biological data required for parameterization is still lacking, particularly in plant systems [9, 11,
316 15, 33, 80].
317
318

319 ***The role of MSH1 in heteroplasmic sorting***

320 We found that mitochondrial heteroplasmic sorting was faster in wild type plants than in
321 *msh1* mutants ($\hat{N}$ = ~1.3 versus ~4 respectively), with the majority of wild type progeny reaching
322 fixation for either the reference or alternative allele within a single generation. This result lends
323 support to the hypothesized repair mechanism in which MSH1 identifies mismatches and
324 initiates DSBs followed by template-based recombinational repair [44, 48, 49]. Under this model,
325 MSH1 would increase rates of gene conversion, thereby homogenizing genome copies within
326 cells and increasing variance among cells [33]. Somewhat confusingly, MSH1 is primarily known
327 as a recombination suppressor because it performs organellar genome surveillance, preventing
328 *illegitimate* recombination between small repeats that can result in genome rearrangements and
329 instability [45, 46]. However, this role is not contradictory to the hypothesis that MSH1 activity
330 increases the overall rate of *homologous* recombination. Indeed, these patterns may well reflect
331 the same mechanism of action, in which MSH1 promotes homologous recombination by
332 introducing DSBs at any mismatched bases, regardless of whether they were generated by
333 strand invasion between short/imperfect repeats or by DNA replication errors.

334 Another way in which MSH1 could alter sorting dynamics is through changing the
335 physical interactions and fusion events between mitochondria. Leaf cells of *msh1* mutants show
336 increased mitochondrial connectivity over wild type [81], suggesting an increased capacity for
337 genome mixing and homogenization. Under existing models [11, 33], increased rates of fusion
338 are predicted to accelerate mitochondrial sorting, the opposite of what we see in *msh1* mutants.
339 However, if the lack of functional MSH1 results in lower rates of gene conversion, higher rates
340 of mitochondrial fusion may not be sufficient to increase cell to cell variance. This observation,
341 in combination with the finding that plastids sort faster than mitochondria even though they do
342 not undergo fusion, suggests that the relationship between fusion and heteroplasmic sorting
343 may be more complex than suggested by current models.

344

345 ***Evidence for GC-biased gene conversion in plant organelles***

346 Organellar genomes, including those of Arabidopsis [82, 83], are typically AT rich but
347 also exhibit a wide range in nucleotide composition [84]. The reasons for these biases are not
348 fully understood, but most organisms across the tree of life are believed to experience an AT-
349 biased mutation spectrum [85-87]. However, in nuclear genomes, this mutation bias may be
350 offset to varying extents by GC-biased gene conversion [88, 89]. The possibility for biased gene
351 conversion in plant organellar genomes is largely unexplored, and the few studies that have
352 investigated this in plastids have come to differing conclusions about GC versus AT bias [90-

353 92]. In the SNVs studied here, we found evidence for GC-biased gene conversion in both
354 plastids and mitochondria, as well as in *msh1* and wild-type backgrounds. Notably, we observed
355 this bias regardless of whether the mutant allele was AT (Mito 334038) or GC (Mito 91017,
356 Plastid 26553, and Plastid 36873). Thus, it appears to be directly related to nucleotide
357 composition and not a systematic preference for or against the reference allele. All four SNVs
358 that were tracked in this study were in intergenic regions (Table 1), reducing the likelihood that
359 the variants alter organellar or cellular function. However, a bias due to selection on functional
360 effects cannot be ruled out.
361 If the observed GC bias is driven by gene conversion, it would suggest that the GC allele
362 is favored during homologous recombination between heterogeneous DNA copies. In
363 mitochondria, gene conversion could happen within heterogeneous organelles, but we expect it
364 to be a particularly powerful actor during fusion in the SAM. In plastids, which do not undergo
365 fusion, gene conversion is only expected to take place within organelles (not between
366 organelles). This raises an important point about our sampling design. The SNVs we tracked
367 are inferred to have arisen in the F2 generation, but we sampled mothers from later generations
368 (F3 to F5 in plastids, Table S1). Therefore, we did not analyze the initial dynamics of plastid
369 variants at their inception when they were present at low frequency among the genome copies
370 within a single plastid. It is possible that some or even all of the plastids in our sampled mothers
371 had already reached homogeneity for one allele or another, meaning that heteroplasmy
372 dynamics were playing out among rather than within plastids. However, if the observed trend
373 towards increasing frequency of GC alleles is the result of gene conversion bias, it would
374 suggest that the mother plants in our study still contained plastids with copies of both alleles.
375 This may be the case, but our data suggest extremely rapid heteroplasmic sorting, which might
376 be expected if it were happening at the level of the small number of whole plastids within a cell.
377 On the other hand, if the sets of genomes within each plastid had already reached homogeneity,
378 it would suggest some unknown mechanism favoring GC alleles. These uncertainties
379 emphasize the importance of determining how the rate of sorting among the genome copies
380 within a plastid compares to the sorting process among the multiple plastids within a cell.
381
382 ***Conclusions***
383 The mechanisms of organellar genome maintenance and transmission are
384 fundamentally different in plants and animals, but very little is known about how this affects the
385 fate of variants arising from de novo mutations. We found that heteroplasmies in plant
386 organelles sort very rapidly which is likely due to a combination of low genome copy numbers in

387 germline/progenitor cells and the recombinational nature of plant genomes. Our work supports a
388 role for gene conversion as an important mechanism facilitating rapid sorting of heteroplasmic
389 variants in plants, which has been hypothesized to be a key mechanism for increasing variance
390 in eukaryotic systems without early germline sequestration [33]. Notably, the recombination
391 surveillance and DNA repair gene *MSH1*, which is absent from most eukaryotic systems
392 including animals [44], appears to play a key role in this process.
393
394
395 **Methods:**
396 ***Plant material***
397 Two homozygous *msh1* (At3g24320) mutant lines containing point mutations that result
398 in either a nonsense mutation (CS3372: *chm1-1*) or an aberrant splice site (CS3246: *chm1-2*)
399 were obtained from the Arabidopsis Biological Resource Center [46]. Crossing design and
400 techniques used for identification of de novo mutations are described previously [44]. Briefly,
401 homozygous mutants were used as males in crosses onto wild-type plants. The resultant F1
402 plants were allowed to self-pollinate, seed was planted and homozygous *msh1* lines (F2
403 families) were identified. The F2 lines were allowed to self-pollinate, and mitochondria and
404 chloroplast were isolated from their progeny (pools of F3 plants). Organellar DNA was extracted
405 and analyzed with Duplex Sequencing [93] to identify de novo mutations. High frequency SNVs
406 identified in F3 organellar DNA pools from *msh1* lines were used in this study (Table 1).
407 In initial experiments, seeds were vernalized in water for 3 days at 4 °C, planted directly
408 into 3-inch pots containing Pro-Mix BX media and grown on light shelves under short day
409 conditions. Heteroplasmic individuals were moved to long (16 h light) day conditions when they
410 began to bolt. Once siliques were ripe, all seeds were harvested in bulk. Seeds of heteroplasmic
411 mothers from lines selected for further analysis were sterilized and plated on MS-agar [94],
412 vernalized for 3 days at 4 °C and placed on light shelves to germinate. When seedlings had two
413 true leaves, they were transferred to 1-inch pots filled with Pro-Mix BX media and placed in a
414 growth chamber under short (10 h light) day conditions. Plants were transferred to long day
415 conditions upon bolting.
416
417 ***DNA extraction and heteroplasmy analysis***
418 Tissue samples were disrupted using the TissueLyser (Qiagen), and total cellular DNA
419 was extracted using the Qiagen Plant DNeasy kit. In cases where tissue was limiting (e.g.,
420 inflorescences), DNA was extracted by grinding tissue in 200 mM Tris-HCl pH 9.0, 250 mM

421 NaCl, 25 mM EDTA, and 1% SDS, followed by precipitation in isopropanol. gDNA was
422 quantified using Qubit (concentrations ranged from 0.5 - 30 ng/μL), and gDNA integrity of
423 selected samples was checked by agarose gel electrophoresis.
424 Heteroplasmy analysis was performed with allele-specific ddPCR assays essentially as
425 described in Wu et al. 2020 [44]. Primers and probes for ddPCR were designed to ten different
426 high-frequency SNV targets, seven in the plastid and three in the mitochondria (Table 1).
427 Primers (Table S6) were designed to amplify fragments of 130 – 250 bp, with the SNV in the
428 middle of the amplified sequence. Probes (Table S7) were designed to either the reference
429 sequence or the variant sequence with the target SNV in the center. All primers and probes
430 were synthesized by Integrative DNA Technologies.
431 ddPCR reactions were composed of: Bio-Rad ddPCR Super Mix for Probes (no dUTP),
432 250 nM final concentration of each (reference and variant) probe, 900 nM final concentration of
433 each primer (F and R), 1 μL of the restriction enzyme BglII (which is used to fragment template
434 DNA but is not predicted to cut within any of the amplified products), and 5 μL of an appropriate
435 dilution of DNA in a 20 μL total reaction. Droplet generation was performed using a Bio-Rad
436 QX200 Droplet Generator as per the manufacturer's instructions, and PCR was performed in a
437 Bio-Rad C1000 with a deep-well block under the following thermal cycling conditions: enzyme
438 activation at 95 °C for 10 min, 40 cycles of 94 °C for 30 sec, annealing/extension temperature
439 (see Table S6) for 1 min, and deactivation of the polymerase and restriction enzyme at 98 °C for
440 10 min – with a ramp speed of 2 °C per sec for all steps. Droplets were read on the Bio-Rad
441 QX200 Droplet Reader and analyzed using QuantaSoft Analysis Software (Bio-Rad).
442 Dilutions of mutant (*msh1*) organellar DNA [44] were used to verify the presence of the
443 SNV in the original mutant F3 organellar extractions alongside paired wild type samples to
444 check for probe specificity. Wild type and mutant organellar samples and no template controls
445 were used to determine appropriate annealing temperatures for each primer probe set to
446 minimize off target binding in wild type and increase separation of positive and negative droplets
447 in both channels. Positive and negative controls were always run alongside experimental
448 samples to ensure assay fidelity and verify appropriate settings for channel thresholds.
449 Background rates of the variant probe binding to wild type samples were typically very low <
450 0.05%. Experimental samples with variant calls falling at or under wild type values were set to
451 zero for further analyses.
452
453

454 ***Organellar genome copy number***

455 Evagreen ddPCR was performed, essentially as described previously [95] to determine
456 the number of mitochondrial and plastid genomes per nuclear genome copy. The reason for this
457 was two-fold. First, it was important to determine whether the *msh1* mutant background altered
458 the relative numbers of organellar genome copies. Secondly, numts (nuclear mitochondrial
459 DNA, i.e., copies of mitochondrial DNA inserted into the nuclear genome) are known to be
460 present in the *A. thaliana* nuclear genome [96-98] and we wanted to correct for this in our
461 heteroplasmy analysis.

462 Two primer sets were used per genome to determine copy number (Table S1). Nuclear
463 genome markers were single copy and located on chromosomes 1 and 2. Organellar markers
464 were designed to single-copy mitochondrial (*rps12* and cox2) and plastid (*clpP1* and *psaA*)
465 genes. For the nuclear genome, values from each primer set were averaged and used to
466 calculate the number of organellar genome copies per nuclear genome for each organellar
467 primer set. For plastids, values from each primer set were averaged to generate a final number
468 of plastid copies per nuclear genome. For mitochondria, both primer sets amplify numts, so the
469 values were adjusted based on the number of numt copies. The amplified regions of *cox2* and
470 *rps12* are present at one and three numt copies, respectively, in the Arabidopsis nuclear
471 genome [98]. These values were subtracted from the calculated values of mitochondrial copies
472 per genome before averaging to obtain a final value. Our mitochondrial SNVs of interest
473 (mt91017 and mt334038) are both present in three copies in the Arabidopsis nuclear genome
474 [98].

475 The number of organellar genomes per nuclear genome were determined for paired wild
476 type and *msh1* mutants that were grown in parallel. Leaves were sampled when plants were 8
477 weeks old. For tissue specific analyses of mitochondrial genomes, samples consisted of whole
478 inflorescences (n = 13) and 8-week-old leaf tissue: old leaves harvested from the base of the
479 rosette (n = 12), and young leaves harvested from the top of the rosette (n = 16). A smaller
480 subset of these was used to determine tissue specific amounts of plastid genomes. For each
481 tissue, an average value was calculated for mitochondrial and plastid genomes per nuclear
482 genome.

483 We used experimentally determined values of organellar genome copy number along
484 with the deduced number of numt copies for each mitochondrial SNV (Nu = 3 in both instances)
485 to correct our heteroplasmy values by computing a correction factor as described below:

486

487 Correction factor for numts in heteroplasmy data, $M = (O + Nu) / O$

Where O = organellar genomes per nuclear genome copy, and Nu = nuclear genome copies of the SNV. The alternative allele frequency for a sample was multiplied by the correction factor M.

Heteroplasmy sampling across generations

To understand the extent to which heteroplasmy is transmitted across generations, we identified heteroplasmic individuals in an initial screen and then analyzed the distribution of allele frequencies in their progeny using ddPCR. Initial screens to identify heteroplasmic mothers were conducted on F3 and/or F4 *msh1* mutant individuals. Leaves were sampled after 4-6 weeks of growth. Seeds from heteroplasmic individuals were sown and an initial leaf sample was taken from each offspring after 4 weeks of growth.

Heteroplasmy sampling within plants

To understand how heteroplasmy is distributed within individuals, we selected three plants from each heteroplasmic mother (described above) for further tissue sampling. Three fully expanded rosette leaves were harvested from each plant at 5 weeks of growth. At 8 weeks of growth, leaves from the top (young) and base of the rosette (old) were harvested. Once plants began to bolt, entire single inflorescences were harvested. Selected tissue samples (8-week-old leaves and inflorescences) from this experiment were used to determine organellar genome copy number in experiments described above. This initial experiment did not lead to the identification of individuals heteroplasmic for plastid markers, so additional plastid SNV lines were grown and tissues were sampled in a subsequent experiment of the same design.

Bottleneck calculations

We used two methods for estimating bottleneck size. First, we applied the common approach based on measures of heteroplasmic variance:

$$\hat{N}_{sv} = \bar{h}\,(1 - \bar{h}) / s^2$$

where s^2 is the variance of heteroplasmic frequency in progeny (or within a plant), $(1/n) \sum (h_i - \bar{h})^2$, and $\bar{h}$ and h_i are the sample mean and individual heteroplasmic frequencies of offspring (or tissues), respectively. In our intergenerational analysis, we used the offspring mean as an approximation of the maternal heteroplasmic frequency, even though we obtained experimental estimates of maternal heteroplasmy levels. This approach was chosen because we saw bias

522 towards GC alleles (Figure 4), suggesting that maternal values would not be an effective
523 estimate of average offspring values.
524 The second approach we took to estimate the bottleneck size was using a Kimura model
525 [51]. Here we maximized the joint likelihood of a set of heteroplasmy measurements under the
526 Kimura model, which takes two parameters: mean heteroplasmy (p) and bottleneck parameter
527 b. is related to the effective "bottleneck size" N by $N = 1/(1-b)$. Previous work has proposed
528 simply setting the population parameters to match the sample statistics p and b [51], but as this
529 does not in general yield the maximum likelihood parameter estimates, it can give misleading
530 results and does not support hypothesis testing. Instead, we used the kimura package in R
531 (https://github.com/lbozhilova/kimura) to compute likelihoods and optimization using Nelder-
532 Mead and Brent algorithms [99] to explicitly find the maximum likelihood parameters and the
533 Fisher information matrix, from which we derive 95% confidence intervals. For homoplasmic
534 cases, numerical issues challenged the Fisher approach, and bootstrap resampling with 200
535 resamples was instead used to estimate confidence intervals.
536 For hypothesis testing regarding the bottleneck size in two different groups of sample
537 sets, we considered two statistical models. First, each set of samples is generated from the
538 Kimura distribution with a set-specific p and a group-specific b. Second, each set of samples is
539 generated with a set-specific p and a b common to both groups. For example, consider a
540 comparison between mitochondrial and plastid intergenerational bottlenecks. In the first model,
541 each family would have its own p, mitochondrial families would have one b value, and plastid
542 families would have another b value. In the second model, each family would have its own p
543 and a b common to all families. We then maximize the joint likelihood over all observations for
544 the two models and conduct a likelihood ratio test with one degree of freedom, reflecting the
545 additional b parameter in the first model. When reporting bottleneck size across samples in a
546 given group (for example, across *msh1* mitochondrial families), we give the maximum likelihood
547 estimate ($\hat{N}$) and confidence intervals from this within-group inference. All code is freely
548 available at https://github.com/StochasticBiology/plant-odna-sorting/.
549
550 ***Impact of* MSH1 *on heteroplasmy transmission***
551 To determine whether MSH1 influences the spread of heteroplasmy across generations,
552 we backcrossed *msh1/msh1* mutant females heteroplasmic for an SNV (either mt334038 or
553 pt26553) to wild-type males. A minimum of 20 F1 plants were tested for heteroplasmy.
554 Heteroplasmic plants were self-pollinated, and then F2 seedlings were planted and screened for
555 heteroplasmy. Both F1 and F2 seedlings were genotyped at the *MSH1* locus as described in

Wu *et al.* 2020 [44]. F2 plants that were homozygous for the *msh1* mutant allele were removed from subsequent analyses.

Acknowledgements:

This work was supported by a grant from the National Institutes of Health (R01 GM118046 to DBS). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 805046 (EvoConBiO) to IGJ).

Author contributions:

AKB: Conceptualization, Investigation, Formal analysis, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review and Editing

LK: Investigation, Formal analysis, Validation, Writing – Review and Editing

MFG: Investigation, Formal analysis, Writing – Review and Editing

MH: Investigation, Writing – Review and Editing

IGJ: Formal analysis, Funding Acquisition, Validation, Visualization, Writing – Review and Editing

DBS: Conceptualization, Formal analysis, Funding Acquisition, Visualization, Writing – Original Draft Preparation, Writing – Review and Editing

Supporting Information:

Figure S1. MSH1 does not significantly alter organellar genome copy number.

Figure S2. Organellar genome copy number varies within and between tissues.

Table S1. Distribution of SNVs over multiple generations of *msh1* mutant lines.

Table S2. Values of heteroplasmy for mother and progeny.

Table S3. Within plant heteroplasmy.

Table S4. Distribution of SNVs over two generations in wild type background.

Table S5. GC bias.

Table S6. Primers for ddPCR heteroplasmy assays and genome copy analysis.

Table S7. Allele-specific probes for ddPCR heteroplasmy assays.

Supporting information:

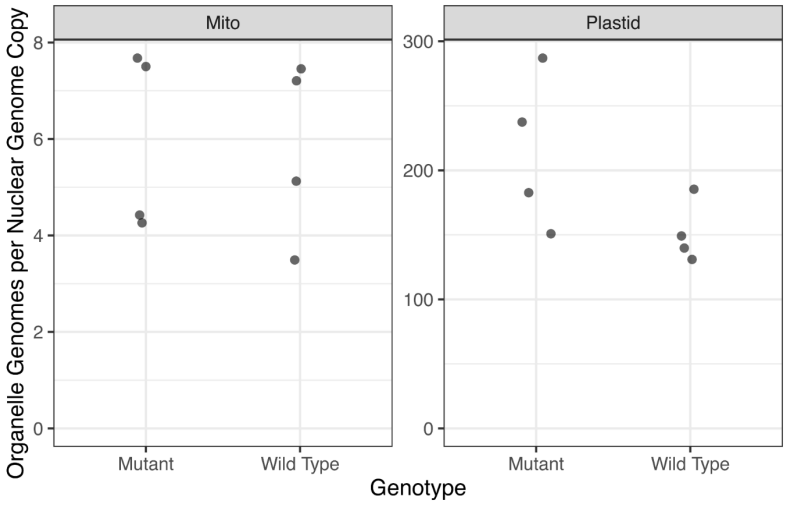

Figure S1. MSH1 does not significantly alter organellar genome copy number. The number of mitochondrial and plastid genome copies per nuclear genome copy was determined using ddPCR, corrected for numts and compared between *msh1* mutant lines (n = 4) and wild type plants (n = 4). Organellar genome copies between mutant and wild type lines were not significantly different (mitochondria, p = 0.910; plastid, p = 0.098).

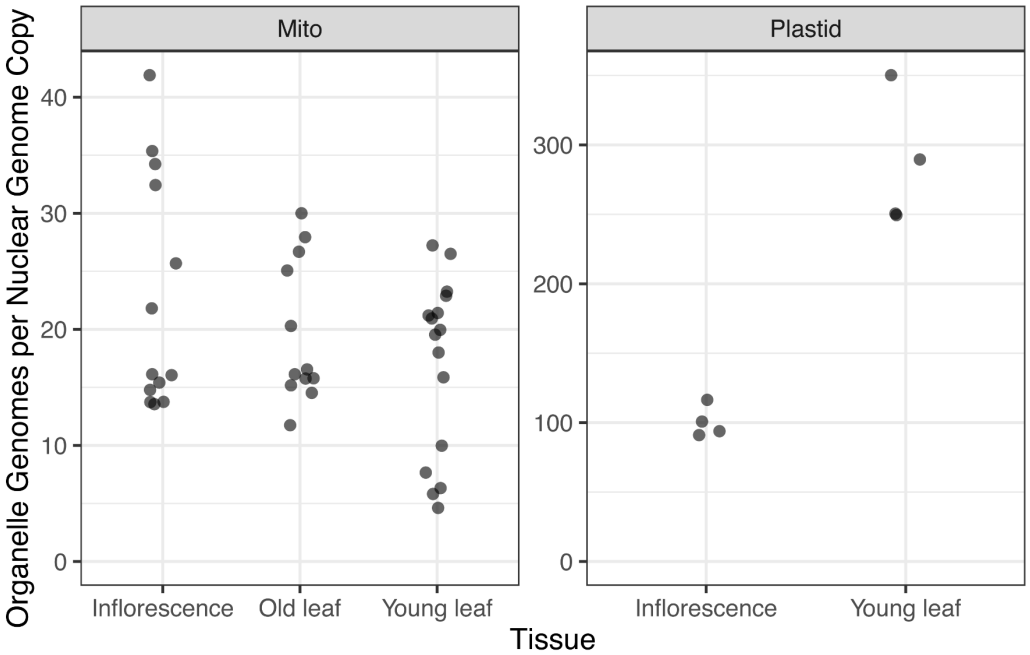

Figure S2. Organellar genome copy number varies within and between tissues. The number of mitochondrial and plastid genome copies per nuclear genome copy was determined using ddPCR and corrected for numts. For mitochondrial primers, tissue samples consisted of whole inflorescences (n = 13), old leaves harvested from the base of the rosette (n = 12), and young leaves harvested from the top of the rosette (n = 16). A subset of samples was chosen for analysis using plastid primers (inflorescences, n = 6; young leaf, n = 6). Plastid copies were significantly lower in inflorescences versus leaves (p = 0.00028); mitochondrial copies were not significantly different between tissue types. Mitochondrial values were adjusted for numt copies.

References:

1. Pereira CV, Gitschlag BL, Patel MR. Cellular mechanisms of mtDNA heteroplasmy dynamics. Critical reviews in biochemistry and molecular biology. 2021;56(5):510-25.
2. Havird JC, Forsythe ES, Williams AM, Werren JH, Dowling DK, Sloan DB. Selfish Mitonuclear Conflict. Curr Biol. 2019;29(11):R496-r511. Epub 2019/06/05. doi: 10.1016/j.cub.2019.03.020. PubMed PMID: 31163164; PubMed Central PMCID: PMCPMC6615485.
3. Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, Schöttler MA, et al. Chloroplast competition is controlled by lipid biosynthesis in evening primroses. Proceedings of the National Academy of Sciences. 2019;116(12):5665-74. doi: 10.1073/pnas.1811661116.
4. Røyrvik EC, Johnston IG. MtDNA sequence features associated with 'selfish genomes' predict tissue-specific segregation and reversion. Nucleic Acids Research. 2020;48(15):8290-301. doi: 10.1093/nar/gkaa622.
5. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet. 2015;16(9):530-42. Epub 2015/08/19. doi: 10.1038/nrg3966. PubMed PMID: 26281784.
6. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW. Mitochondrial DNA mutations and human disease. Biochim Biophys Acta. 2010;1797(2):113-28. Epub 2009/09/19. doi: 10.1016/j.bbabio.2009.09.005. PubMed PMID: 19761752.
7. Arbeithuber B, Hester J, Cremona MA, Stoler N, Zaidi A, Higgins B, et al. Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. PLoS Biol. 2020;18(7):e3000745. Epub 2020/07/16. doi: 10.1371/journal.pbio.3000745. PubMed PMID: 32667908; PubMed Central PMCID: PMCPMC7363077.
8. Stewart JB, Larsson NG. Keeping mtDNA in shape between generations. PLoS Genet. 2014;10(10):e1004670. Epub 2014/10/10. doi: 10.1371/journal.pgen.1004670. PubMed PMID: 25299061; PubMed Central PMCID: PMCPMC4191934.
9. Johnston IG, Burgstaller JP, Havlicek V, Kolbe T, Rulicke T, Brem G, et al. Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. Elife. 2015;4:e07464. Epub 2015/06/03. doi: 10.7554/eLife.07464. PubMed PMID: 26035426; PubMed Central PMCID: PMCPMC4486817.

10. Wei W, Tuna S, Keogh MJ, Smith KR, Aitman TJ, Beales PL, et al. Germline selection shapes human mitochondrial DNA diversity. Science. 2019;364(6442). Epub 2019/05/28. doi: 10.1126/science.aau6520. PubMed PMID: 31123110.
11. Johnston IG. Varied Mechanisms and Models for the Varying Mitochondrial Bottleneck. Front Cell Dev Biol. 2019;7:294. Epub 2019/12/12. doi: 10.3389/fcell.2019.00294. PubMed PMID: 31824946; PubMed Central PMCID: PMCPMC6879659.
12. Birky CW. The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms, and Models. Annu Rev Genetics. 2001;35:125-48.
13. Li M, Rothwell R, Vermaat M, Wachsmuth M, Schröder R, Laros JF, et al. Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. Genome Res. 2016;26(4):417-26. Epub 2016/02/27. doi: 10.1101/gr.203216.115. PubMed PMID: 26916109; PubMed Central PMCID: PMCPMC4817766.
14. Rebolledo-Jaramillo B, Su MS, Stoler N, McElhoe JA, Dickins B, Blankenberg D, et al. Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. Proc Natl Acad Sci U S A. 2014;111(43):15474-9. Epub 2014/10/15. doi: 10.1073/pnas.1409328111. PubMed PMID: 25313049; PubMed Central PMCID: PMCPMC4217420.
15. Wilton PR, Zaidi A, Makova K, Nielsen R. A Population Phylogenetic View of Mitochondrial Heteroplasmy. Genetics. 2018;208(3):1261-74. Epub 2018/01/19. doi: 10.1534/genetics.118.300711. PubMed PMID: 29343499; PubMed Central PMCID: PMCPMC5844335.
16. Zaidi AA, Wilton PR, Su MS, Paul IM, Arbeithuber B, Anthony K, et al. Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. Proc Natl Acad Sci U S A. 2019;116(50):25172-8. Epub 2019/11/24. doi: 10.1073/pnas.1906331116. PubMed PMID: 31757848; PubMed Central PMCID: PMCPMC6911200.
17. Bentley KE, Mandel JR, McCauley DE. Paternal leakage and heteroplasmy of mitochondrial genomes in Silene vulgaris: evidence from experimental crosses. Genetics. 2010;185(3):961-8. Epub 2010/04/28. doi: 10.1534/genetics.110.115360. PubMed PMID: 20421605; PubMed Central PMCID: PMCPMC2907211.
18. Mandel JR, Ramsey AJ, Holley JM, Scott VA, Mody D, Abbot P. Disentangling Complex Inheritance Patterns of Plant Organellar Genomes: An Example From Carrot. J Hered.

2020;111(6):531-8. Epub 2020/09/05. doi: 10.1093/jhered/esaa037. PubMed PMID: 32886780.

19. McCauley DE. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytol. 2013;200(4):966-77. Epub 2013/08/21. doi: 10.1111/nph.12431. PubMed PMID: 23952142.
20. Ramsey AJ, Mandel JR. When One Genome Is Not Enough: Organellar Heteroplasmy in Plants. Annual Plant Reviews online2019. p. 619-58.
21. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008;49(3):827-31. Epub 2008/10/08. doi: 10.1016/j.ympev.2008.09.009. PubMed PMID: 18838124.
22. Smith DR. Mutation rates in plastid genomes: they are lower than you might think. Genome Biol Evol. 2015;7(5):1227-34. Epub 2015/04/15. doi: 10.1093/gbe/evv069. PubMed PMID: 25869380; PubMed Central PMCID: PMCPMC4453064.
23. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci U S A. 1987;84(24):9054-8. Epub 1987/12/01. doi: 10.1073/pnas.84.24.9054. PubMed PMID: 3480529; PubMed Central PMCID: PMCPMC299690.
24. Massouh A, Schubert J, Yaneva-Roder L, Ulbricht-Jones ES, Zupok A, Johnson MT, et al. Spontaneous Chloroplast Mutants Mostly Occur by Replication Slippage and Show a Biased Pattern in the Plastome of Oenothera. Plant Cell. 2016;28(4):911-29. Epub 2016/04/08. doi: 10.1105/tpc.15.00879. PubMed PMID: 27053421; PubMed Central PMCID: PMCPMC4863383.
25. Greiner S. Plastome Mutants of Higher Plants. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria. Dordrecht: Springer Netherlands; 2012. p. 237-66.
26. Muller HJ. The relation of recombination to mutational advance. Mutat Res. 1964;106:2-9. Epub 1964/05/01. doi: 10.1016/0027-5107(64)90047-8. PubMed PMID: 14195748.
27. Felsenstein J. The evolutionary advantage of recombination. Genetics. 1974;78(2):737-56. Epub 1974/10/01. doi: 10.1093/genetics/78.2.737. PubMed PMID: 4448362; PubMed Central PMCID: PMCPMC1213231.
28. Zhang Q, Liu Y, Sodmergen. Examination of the Cytoplasmic DNA in Male Reproductive Cells to Determine the Potential for Cytoplasmic Inheritance in 295 Angiosperm Species. Plant and Cell Physiology. 2003;44(9):941-51. doi: 10.1093/pcp/pcg121.

29. Barr CM, Neiman M, Taylor DR. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 2005;168(1):39-50. Epub 2005/09/15. doi: 10.1111/j.1469-8137.2005.01492.x. PubMed PMID: 16159319.
30. Greiner S, Sobanski J, Bock R. Why are most organelle genomes transmitted maternally? Bioessays. 2015;37(1):80-94. Epub 2014/10/11. doi: 10.1002/bies.201400110. PubMed PMID: 25302405; PubMed Central PMCID: PMCPMC4305268.
31. Christie JR, Beekman M. Uniparental Inheritance Promotes Adaptive Evolution in Cytoplasmic Genomes. Mol Biol Evol. 2017;34(3):677-91. Epub 2016/12/28. doi: 10.1093/molbev/msw266. PubMed PMID: 28025277; PubMed Central PMCID: PMCPMC5896580.
32. Christie JR, Beekman M. Selective sweeps of mitochondrial DNA can drive the evolution of uniparental inheritance. Evolution. 2017;71(8):2090-9. Epub 2017/06/24. doi: 10.1111/evo.13291. PubMed PMID: 28640391.
33. Edwards DM, Royrvik EC, Chustecki JM, Giannakis K, Glastad RC, Radzvilavicius AL, et al. Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. PLoS Biol. 2021;19(4):e3001153. Epub 2021/04/24. doi: 10.1371/journal.pbio.3001153. PubMed PMID: 33891583; PubMed Central PMCID: PMCPMC8064548.
34. Radzvilavicius AL, Kokko H, Christie JR. Mitigating Mitochondrial Genome Erosion Without Recombination. Genetics. 2017;207(3):1079-88. Epub 2017/09/13. doi: 10.1534/genetics.117.300273. PubMed PMID: 28893855; PubMed Central PMCID: PMCPMC5676227.
35. Greiner S, Golczyk H, Malinova I, Pellizzer T, Bock R, Borner T, et al. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. Plant J. 2020;102(4):730-46. Epub 2019/12/20. doi: 10.1111/tpj.14658. PubMed PMID: 31856320.
36. Gao L, Guo X, Liu XQ, Zhang L, Huang J, Tan L, et al. Changes in mitochondrial DNA levels during early embryogenesis in Torenia fournieri and Arabidopsis thaliana. Plant J. 2018. Epub 2018/06/09. doi: 10.1111/tpj.13987. PubMed PMID: 29882345.
37. Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Borner T. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. Plant J. 2010;64(6):948-59. Epub 2010/12/15. doi: 10.1111/j.1365-313X.2010.04389.x. PubMed PMID: 21143676.

38. Wang DY, Zhang Q, Liu Y, Lin ZF, Zhang SX, Sun MX, et al. The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. Plant Cell. 2010;22(7):2402-16. Epub 2010/07/08. doi: 10.1105/tpc.109.071902. PubMed PMID: 20605854; PubMed Central PMCID: PMCPMC2929101.
39. Woloszynska M. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes--though this be madness, yet there's method in't. J Exp Bot. 2010;61(3):657-71. Epub 2009/12/10. doi: 10.1093/jxb/erp361. PubMed PMID: 19995826.
40. Gualberto JM, Newton KJ. Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. Annu Rev Plant Biol. 2017;68:225-52. Epub 2017/02/23. doi: 10.1146/annurev-arplant-043015-112232. PubMed PMID: 28226235.
41. Ruhlman TA, Zhang J, Blazier JC, Sabir JSM, Jansen RK. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot. 2017;104(4):559-72. Epub 2017/04/13. doi: 10.3732/ajb.1600453. PubMed PMID: 28400415.
42. Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA. Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. Plant Cell. 1998;10(7):1163-80. Epub 1998/07/21. doi: 10.1105/tpc.10.7.1163. PubMed PMID: 9668135; PubMed Central PMCID: PMCPMC144058.
43. Yang X, Mackenzie SA. Many Facets of Dynamic Plasticity in Plants. Cold Spring Harb Perspect Biol. 2019;11(10). Epub 2019/05/30. doi: 10.1101/cshperspect.a034629. PubMed PMID: 31138545; PubMed Central PMCID: PMCPMC6771370.
44. Wu Z, Waneka G, Broz AK, King CR, Sloan DB. MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proc Natl Acad Sci U S A. 2020;117(28):16448-55. Epub 2020/07/01. doi: 10.1073/pnas.2001998117. PubMed PMID: 32601224; PubMed Central PMCID: PMCPMC7368333.
45. Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, et al. Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol. 2011;9:64. Epub 2011/09/29. doi: 10.1186/1741-7007-9-64. PubMed PMID: 21951689; PubMed Central PMCID: PMCPMC3193812.
46. Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene

homologous to MutS. Proceedings of the National Academy of Sciences. 2003;100(10):5968-73. doi: 10.1073/pnas.1037651100.

47. Ogata H, Ray J, Toyoda K, Sandaa RA, Nagasaki K, Bratbak G, et al. Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. ISME J. 2011;5(7):1143-51. Epub 2011/01/21. doi: 10.1038/ismej.2010.210. PubMed PMID: 21248859; PubMed Central PMCID: PMCPMC3146287.

48. Ayala-García VM, Baruch-Torres N, García-Medel PL, Brieba LG. Plant organellar DNA polymerases paralogs exhibit dissimilar nucleotide incorporation fidelity. Febs j. 2018;285(21):4005-18. Epub 2018/08/29. doi: 10.1111/febs.14645. PubMed PMID: 30152200.

49. Christensen AC. Genes and junk in plant mitochondria-repair mechanisms and selection. Genome Biol Evol. 2014;6(6):1448-53. Epub 2014/06/07. doi: 10.1093/gbe/evu115. PubMed PMID: 24904012; PubMed Central PMCID: PMCPMC4079193.

50. Marechal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186(2):299-317. Epub 2010/02/26. doi: 10.1111/j.1469-8137.2010.03195.x. PubMed PMID: 20180912.

51. Wonnapinij P, Chinnery PF, Samuels DC. Previous estimates of mitochondrial DNA mutation level variance did not account for sampling error: comparing the mtDNA genetic bottleneck in mice and humans. Am J Hum Genet. 2010;86(4):540-50. Epub 2010/04/07. doi: 10.1016/j.ajhg.2010.02.023. PubMed PMID: 20362273; PubMed Central PMCID: PMCPMC2850432.

52. Lanfear R. Do plants have a segregated germline? PLoS Biol. 2018;16(5):e2005439. Epub 2018/05/17. doi: 10.1371/journal.pbio.2005439. PubMed PMID: 29768400; PubMed Central PMCID: PMCPMC5973621.

53. Radzvilavicius AL, Hadjivasiliou Z, Pomiankowski A, Lane N. Selection for Mitochondrial Quality Drives Evolution of the Germline. PLoS Biol. 2016;14(12):e2000410. Epub 2016/12/21. doi: 10.1371/journal.pbio.2000410. PubMed PMID: 27997535; PubMed Central PMCID: PMCPMC5172535.

54. Xu S, Schaack S, Seyfert A, Choi E, Lynch M, Cristescu ME. High Mutation Rates in the Mitochondrial Genomes of Daphnia pulex. Molecular Biology and Evolution. 2011;29(2):763-9. doi: 10.1093/molbev/msr243.

55. Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. PLoS Biol. 2008;6(8):e204. Epub 2008/08/22. doi: 10.1371/journal.pbio.0060204. PubMed PMID: 18715119; PubMed Central PMCID: PMCPMC2517619.
56. Konrad A, Thompson O, Waterston RH, Moerman DG, Keightley PD, Bergthorsson U, et al. Mitochondrial Mutation Rate, Spectrum and Heteroplasmy in Caenorhabditis elegans Spontaneous Mutation Accumulation Lines of Differing Population Size. Molecular Biology and Evolution. 2017;34(6):1319-34. doi: 10.1093/molbev/msx051.
57. Wolff JN, White DJ, Woodhams M, White HE, Gemmell NJ. The strength and timing of the mitochondrial bottleneck in salmon suggests a conserved mechanism in vertebrates. PLoS One. 2011;6(5):e20522. Epub 2011/06/10. doi: 10.1371/journal.pone.0020522. PubMed PMID: 21655224; PubMed Central PMCID: PMCPMC3105079.
58. Otten Auke BC, Theunissen Tom EJ, Derhaag Josien G, Lambrichs Ellen H, Boesten Iris BW, Winandy M, et al. Differences in Strength and Timing of the mtDNA Bottleneck between Zebrafish Germline and Non-germline Cells. Cell Reports. 2016;16(3):622-30. doi: https://doi.org/10.1016/j.celrep.2016.06.023.
59. Zhang H, Burr SP, Chinnery PF. The mitochondrial DNA genetic bottleneck: inheritance and beyond. Essays Biochem. 2018;62(3):225-34. Epub 2018/06/09. doi: 10.1042/ebc20170096. PubMed PMID: 29880721.
60. Cree LM, Samuels DC, de Sousa Lopes SC, Rajasimha HK, Wonnapinij P, Mann JR, et al. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nat Genet. 2008;40(2):249-54. Epub 2008/01/29. doi: 10.1038/ng.2007.63. PubMed PMID: 18223651.
61. Bilinski SM, Kloc M, Tworzydlo W. Selection of mitochondria in female germline cells: is Balbiani body implicated in this process? J Assist Reprod Genet. 2017;34(11):1405-12. Epub 2017/07/30. doi: 10.1007/s10815-017-1006-3. PubMed PMID: 28755153; PubMed Central PMCID: PMCPMC5699987.
62. Colnaghi M, Pomiankowski A, Lane N. The need for high-quality oocyte mitochondria at extreme ploidy dictates mammalian germline development. Elife. 2021;10. Epub 2021/07/20. doi: 10.7554/eLife.69344. PubMed PMID: 34279226; PubMed Central PMCID: PMCPMC8337077.
63. Zhang H, Esposito M, Pezet MG, Aryaman J, Wei W, Klimm F, et al. Mitochondrial DNA heteroplasmy is modulated during oocyte development propagating mutation transmission. Science Advances. 2021;7(50):eabi5657. doi: doi:10.1126/sciadv.abi5657.

64. Lieber T, Jeedigunta SP, Palozzi JM, Lehmann R, Hurd TR. Mitochondrial fragmentation drives selective removal of deleterious mtDNA in the germline. Nature. 2019;570(7761):380-4. doi: 10.1038/s41586-019-1213-4.
65. Cao L, Shitara H, Sugimoto M, Hayashi J, Abe K, Yonekawa H. New evidence confirms that the mitochondrial bottleneck is generated without reduction of mitochondrial DNA content in early primordial germ cells of mice. PLoS Genet. 2009;5(12):e1000756. Epub 2009/12/10. doi: 10.1371/journal.pgen.1000756. PubMed PMID: 19997484; PubMed Central PMCID: PMCPMC2777314.
66. Wai T, Teoli D, Shoubridge EA. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nature Genetics. 2008;40(12):1484-8. doi: 10.1038/ng.258.
67. Watson JM, Platzer A, Kazda A, Akimcheva S, Valuchova S, Nizhynska V, et al. Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. Proc Natl Acad Sci U S A. 2016;113(43):12226-31. Epub 2016/10/30. doi: 10.1073/pnas.1609686113. PubMed PMID: 27729523; PubMed Central PMCID: PMCPMC5087024.
68. Hagemann R. Plastid Genetics in Higher Plants. In: Herrmann RG, editor. Cell Organelles. Vienna: Springer Vienna; 1992. p. 65-96.
69. Lutz KA, Svab Z, Maliga P. Construction of marker-free transplastomic tobacco using the Cre-loxP site-specific recombination system. Nat Protoc. 2006;1(2):900-10. Epub 2007/04/05. doi: 10.1038/nprot.2006.118. PubMed PMID: 17406323.
70. Svab Z, Hajdukiewicz P, Maliga P. Stable transformation of plastids in higher plants. Proc Natl Acad Sci U S A. 1990;87(21):8526-30. Epub 1990/11/01. doi: 10.1073/pnas.87.21.8526. PubMed PMID: 11607112; PubMed Central PMCID: PMCPMC54989.
71. Lutz KA, Maliga P. Plastid genomes in a regenerating tobacco shoot derive from a small number of copies selected through a stochastic process. Plant J. 2008;56(6):975-83. Epub 2008/08/16. doi: 10.1111/j.1365-313X.2008.03655.x. PubMed PMID: 18702667.
72. Maliga P, Bock R. Plastid Biotechnology: Food, Fuel, and Medicine for the 21st Century. Plant Physiology. 2011;155(4):1501-10. doi: 10.1104/pp.110.170969.
73. Rascón-Cruz Q, González-Barriga CD, Iglesias-Figueroa BF, Trejo-Muñoz JC, Siqueiros-Cendón T, Sinagawa-García SR, et al. Plastid transformation: Advances and challenges for its implementation in agricultural crops. Electronic Journal of Biotechnology. 2021;51:95-109. doi: https://doi.org/10.1016/j.ejbt.2021.03.005.

74. Wang Y, Wei Z, Xing S. Stable plastid transformation of rice, a monocot cereal crop. Biochemical and Biophysical Research Communications. 2018;503(4):2376-9. doi: https://doi.org/10.1016/j.bbrc.2018.06.164.
75. Liu X, Weaver D, Shirihai O, Hajnóczky G. Mitochondrial 'kiss-and-run': interplay between mitochondrial motility and fusion–fission dynamics. The EMBO journal. 2009;28(20):3074-89.
76. Segui-Simarro JM, Coronado MJ, Staehelin LA. The mitochondrial cycle of Arabidopsis shoot apical meristem and leaf primordium meristematic cells is defined by a perinuclear tentaculate/cage-like mitochondrion. Plant Physiol. 2008;148(3):1380-93. Epub 2008/09/19. doi: 10.1104/pp.108.126953. PubMed PMID: 18799659; PubMed Central PMCID: PMCPMC2577259.
77. Rose RJ. Contribution of Massive Mitochondrial Fusion and Subsequent Fission in the Plant Life Cycle to the Integrity of the Mitochondrion and Its Genome. Int J Mol Sci. 2021;22(11). Epub 2021/06/03. doi: 10.3390/ijms22115429. PubMed PMID: 34063907; PubMed Central PMCID: PMCPMC8196625.
78. Segui-Simarro JM, Staehelin LA. Mitochondrial reticulation in shoot apical meristem cells of Arabidopsis provides a mechanism for homogenization of mtDNA prior to gamete formation. Plant Signal Behav. 2009;4(3):168-71. Epub 2009/09/02. doi: 10.4161/psb.4.3.7755. PubMed PMID: 19721742; PubMed Central PMCID: PMCPMC2652521.
79. Arimura SI. Fission and Fusion of Plant Mitochondria, and Genome Maintenance. Plant Physiol. 2018;176(1):152-61. Epub 2017/11/16. doi: 10.1104/pp.17.01025. PubMed PMID: 29138352; PubMed Central PMCID: PMCPMC5761811.
80. Johnston IG. Tension and Resolution: Dynamic, Evolving Populations of Organelle Genomes within Plant Cells. Mol Plant. 2019;12(6):764-83. Epub 2018/11/18. doi: 10.1016/j.molp.2018.11.002. PubMed PMID: 30445187.
81. Chustecki JM, Etherington RD, Gibbs DJ, Johnston IG. Altered collective mitochondrial dynamics in an *Arabidopsis msh1* mutant compromising organelle DNA maintenance. bioRxiv. 2021:2021.10.22.465420. doi: 10.1101/2021.10.22.465420.
82. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA research. 1999;6(5):283-90.
83. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nature Genetics. 1997;15(1):57-61. doi: 10.1038/ng0197-57.

84. Smith D. Updating Our View of Organelle Genome Nucleotide Landscape. Frontiers in Genetics. 2012;3. doi: 10.3389/fgene.2012.00175.
85. Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 2010;6(9):e1001115. Epub 2010/09/15. doi: 10.1371/journal.pgen.1001115. PubMed PMID: 20838599; PubMed Central PMCID: PMCPMC2936535.
86. Hildebrand F, Meyer A, Eyre-Walker A. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 2010;6(9):e1001107. Epub 2010/09/15. doi: 10.1371/journal.pgen.1001107. PubMed PMID: 20838593; PubMed Central PMCID: PMCPMC2936529.
87. Sloan DB, Wu Z. History of plastid DNA insertions reveals weak deletion and at mutation biases in angiosperm mitochondrial genomes. Genome Biol Evol. 2014;6(12):3210-21. Epub 2014/11/25. doi: 10.1093/gbe/evu253. PubMed PMID: 25416619; PubMed Central PMCID: PMCPMC4986453.
88. Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GA. Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol Evol. 2012;4(7):675-82. Epub 2012/05/26. doi: 10.1093/gbe/evs052. PubMed PMID: 22628461; PubMed Central PMCID: PMCPMC5635611.
89. Galtier N, Duret L. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 2007;23(6):273-7. Epub 2007/04/10. doi: 10.1016/j.tig.2007.03.011. PubMed PMID: 17418442.
90. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 2006;46(1):85-94. Epub 2006/03/24. doi: 10.1111/j.1365-313X.2006.02673.x. PubMed PMID: 16553897.
91. Wu C-S, Chaw S-M. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion. Genome Biology and Evolution. 2015;7(7):2000-9. doi: 10.1093/gbe/evv125.
92. Niu Z, Xue Q, Wang H, Xie X, Zhu S, Liu W, et al. Mutational Biases and GC-Biased Gene Conversion Affect GC Content in the Plastomes of Dendrobium Genus. International Journal of Molecular Sciences. 2017;18(11):2307. PubMed PMID: doi:10.3390/ijms18112307.
93. Kennedy SR, Schmitt MW, Fox EJ, Kohrn BF, Salk JJ, Ahn EH, et al. Detecting ultralow-frequency mutations by Duplex Sequencing. Nature protocols. 2014;9(11):2586-606.

94. Zhang X, Henriques R, Lin SS, Niu QW, Chua NH. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc. 2006;1(2):641-6. Epub 2007/04/05. doi: 10.1038/nprot.2006.97. PubMed PMID: 17406292.

95. Fernandes Gyorfy M, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, et al. Nuclear-cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. Plant J. 2021;108(1):219-30. Epub 2021/07/27. doi: 10.1111/tpj.15436. PubMed PMID: 34309123.

96. Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, et al. 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. 2001;98(9):5099. doi: 10.1073/pnas.091110398.

97. Lin X, Kaul S, Rounsley S, Shea TP, Benito MI, Town CD, et al. Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature. 1999;402(6763):761-8. Epub 2000/01/05. doi: 10.1038/45471. PubMed PMID: 10617197.

98. Fields PD, Waneka G, Naish M, Schatz MC, Henderson IR, Sloan DB. Complete sequence of a 641-kb insertion of mitochondrial DNA in the *Arabidopsis thaliana* nuclear genome. bioRxiv. 2022:2022.02.22.481460. doi: 10.1101/2022.02.22.481460.

99. Nocedal J, Wright SJ. Numerical optimization: Springer; 1999.