Dynamics of organelle DNA segregation in *Arabidopsis* development and reproduction revealed with tissue-specific heteroplasmy profiling and stochastic modelling

Amanda K Broz[1], Daniel B Sloan[1], Iain G Johnston[2,3,*]

[1] Department of Biology, Colorado State University, Fort Collins, CO 80523.

[2] Department of Mathematics, University of Bergen, Bergen, 5007, Norway.

[3] Computational Biology Unit, University of Bergen, Bergen, 5007, Norway.

[*] correspondence to iain.johnston@uib.no

Abstract

Organelle DNA (oDNA) in mitochondria and plastids is vital for plant (and eukaryotic) life. Selection against damaged oDNA is mediated in part by segregation – the sorting of different oDNA types into different cells in the germline. Plants segregate oDNA very rapidly, with oDNA recombination protein MutS Homolog 1 (MSH1), a key driver of this segregation, but in contrast to mammals, we have very limited knowledge of the dynamics of this segregation within plants and between generations. Here, we combine stochastic modelling with tissue-specific heteroplasmy measurements to reveal the trajectories of oDNA segregation in *Arabidopsis thaliana* development and reproduction. We obtain and use new experimental observations of oDNA through development to confirm and refine the predictions of the theory inferred from existing measurements. Ongoing segregation proceeds gradually but continually during plant development, with a more rapid increase between inflorescence formation and the establishment of the next generation. When MSH1 is compromised, we show that the majority of observed segregation could be achieved through partitioning at cell divisions. When MSH1 is functional, mtDNA segregation is far more rapid than can be achieved through cell divisions; we show that increased oDNA gene conversion is a plausible mechanism quantitatively explaining this acceleration. We also discuss the support for different models of the plant germline provided by these observations.

Introduction

Mitochondria and plastids are essential sites of energy transduction across eukaryotes. Originally independent organisms, they retain their own genomes (organelle DNA or oDNA; mtDNA and ptDNA respectively) encoding essential aspects of bioenergetic machinery in plants (and other eukaryotes) [Allen & Martin, 2016; Giannakis et al., 2022; Mohanta et al., 2020; Palmer et al., 2000; Clegg et al., 1994]. Plant cells typically contain populations that range from dozens to thousands of mtDNA and ptDNA molecules [Preuten et al., 2010; Greiner et al., 2020; Wang et al. 2010; Fernandes Gyorfy et

al., 2021], contained within their respective organelles [MacCauley, 2013; Woloszynska, 2010; Barr et al., 2005; Johnston, 2019a]. Due to their centrality in bioenergetic, metabolic, and other cellular processes, it is essential to preserve the integrity of oDNA genes. This preservation necessitates a way of dealing with oDNA mutations and ensuring faithful inheritance of oDNA between generations.

Mutations in oDNA can give rise to heteroplasmy – a mixture of several oDNA types within a cell [Wallace & Chalkia, 2013; Stewart & Chinnery, 2015]. Across eukaryotes, developmental and genetic processes exist to limit the inheritance of heteroplasmy [Edwards et al., 2021]. In several animals, mtDNA inheritance is shaped by the so-called developmental bottleneck [Johnston, 2019b; Stewart & Chinnery, 2015; Zhang et al., 2018]. Here, cell-to-cell variance in heteroplasmy is increased in the female germline, so that individual gametes have a wide range of heteroplasmy levels. Through this increase in variance – called segregation or “sorting out” – it is then possible for some gametes to inherit lower levels of damaging mutations than the mother’s average. If gametes with high levels of such mutations are removed by selection, the mutational burden passed to the next generation is limited.

Previous work has characterised inheritance and vegetative sorting of heteroplasmy in carrot [Mandel et al., 2020]. Here, little evidence was described for segregation during plant development, with most observations involving a loss of heteroplasmy between generations. The heteroplasmy levels involved in this study were typically extreme (around 1% frequency of the minor allele), meaning that such segregation would be very hard to detect; and one notable instance was recorded of a 31% heteroplasmic offspring arising from a <1% heteroplasmic mother and father, suggesting that a mechanism for substantial amplification of minor alleles may be present. The loss of heteroplasmy upon inheritance agrees with results in Silene [Bentley et al., 2010], where only 17% of offspring retained heteroplasmy that was present in their mother.

How plants limit the inheritance of these damaging mutations is less well understood [MacCauley, 2013; Woloszynska, 2010; Barr et al., 2005; Galtier, 2011]. Although the observation of within-plant segregation of oDNA-linked phenotypes dates back over a century (and led to the discovery of cytoplasmic inheritance) [Hagemann, 2010; Greiner 2012], the quantitative dynamics and mechanisms of this segregation remain unclear. Recent experimental evidence has shown that sorting out of plant mtDNA and ptDNA is extremely rapid compared to animals [Broz et al., 2022]. This work showed that this sorting depends on *MSH1*, a gene responsible for controlling recombination activity in organelle DNA [Abdelnoor et al., 2003]. Although the precise nature and mechanism of this control is yet to be determined [Arrieta-Montiel et al., 2009; Virdi et al., 2015; Christensen, 2014], *MSH1* is required to maintain a low mutational burden in plant oDNA [Wu et al., 2020], accelerates oDNA segregation [Broz et al., 2022], and supports oDNA gene conversion [Gualberto et al., 2014; Edwards et al., 2021]. Other recombination factors including members of the

RECA gene family also contribute to oDNA maintenance [Rowan et al., 2010; Maréchal & Brisson, 2010; Day & Madesis, 2007; Shedge et al., 2007; Miller-Messmer et al., 2012]. Theoretical work has explored the role of recombination processes in shaping plant oDNA [Atlan & Couvet, 1993; Albert et al., 1996], suggesting that gene conversion provides a strategy for oDNA segregation [Lonsdale et al., 1988; Khakhlova & Bock, 2006], with stochastic modelling showing that such segregation can occur without requiring a reduction in cellular oDNA copy number [Edwards et al., 2021]. This feature is potentially useful for plants, where, due to developmental dynamics, a germline cannot readily be sequestered and manipulated to impose a physical bottleneck. oDNA copy number in plant meristems is lower than in many animal cases [Edwards et al., 2021; Preuten et al. 2010; Wang et al. 2010; Greiner et al., 2020], but this reduction alone cannot account for the extent of segregation observed [Broz et al., 2022]. The developmental history of the plant germline differs dramatically from the animal case [Lanfear, 2018; Burian et al., 2016], and any understanding of how oDNA segregation proceeds during development necessitates an analysis approach that can both account for the developmental history underlying samples [Wilton et al., 2018; Stadler et al., 2021] and the uncertainty over different models of plant germline development [Lanfear, 2018; Kirk et al., 2013].

Here, we attempt to illuminate the dynamics and mechanisms by which plants perform this rapid sorting of oDNA heteroplasmy. We combine existing heteroplasmy measurements within and across plant generations with a stochastic phylodynamic model for cellular oDNA dynamics during plant development. We use Bayesian inference and model selection to reveal when and where cell-to-cell variability is generated; model selection and mathematical analysis reveals the likely physical mechanisms responsible for this segregation. We confirm the predictions of this model with new experimental observations, characterising the segregation dynamics of mtDNA and ptDNA within plants in unprecedented quantitative detail.

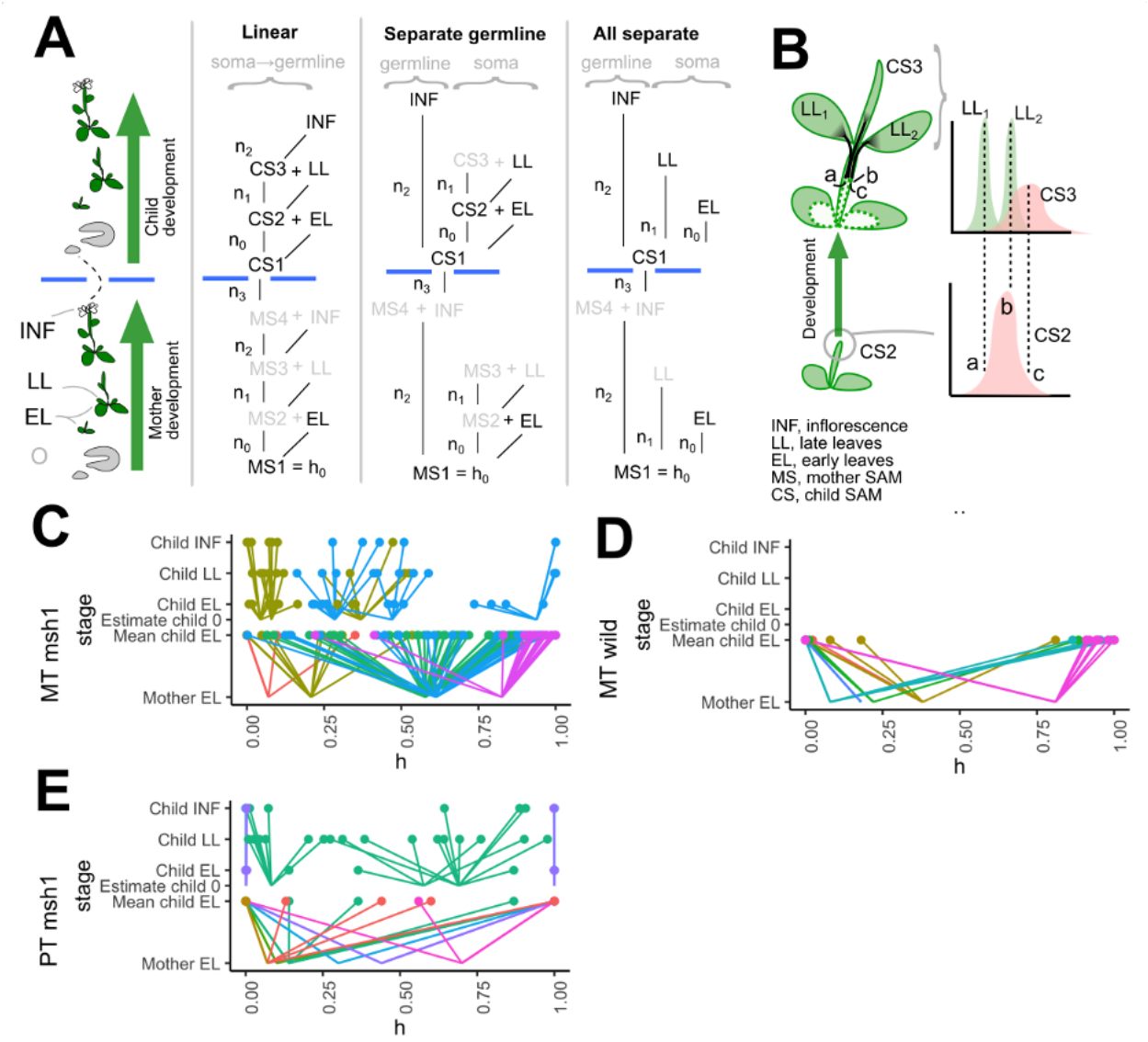


Figure 1. Models and data for heteroplasmy segregation in plant development. (A) Developmental models for heteroplasmy observations. MSi and CSi are the unobserved (latent) ancestral cells at different developmental stages in Mother and Child shoot apical meristem (SAM). The blue horizontal bars denote the generation of sex cells and establishment of a new generation. Greyed-out elements are unidentifiable given our observations and play no role in our model. n_i correspond to the number of effective segregation events (model cell divisions) at each developmental stage. (B) Example of heteroplasmy model within the linear developmental model in (A). The SAM at the CS2 stage includes cell with a distribution of heteroplasmy levels. In this example, three cells a, b, and c from this distribution, with different heteroplasmy levels, go on to be the ancestors of two late leaves (LL_1 and LL_2) and part of the future SAM at stage CS3. Segregation increases heteroplasmy variance as the descendants of a, b, and c develop, leading to new distributions. These may be sampled (the mean of LL_1 and LL_2 are recorded) or unseen (the CS3 distribution plays a latent role in our model). (C-E) Observed heteroplasmy data through development in different heteroplasmic plant families from Broz et al. [2022]: (C) mtDNA in mutant *msh1* background; (D) mtDNA in wildtype background; (E) ptDNA in mutant *msh1* background. Between-generation (upper) and within-plant (lower) observations are shown. The “fanning out” of individual sample heteroplasmies over time corresponds to increasing sample-to-sample variance.

Results

Developmental models for heteroplasmy within and across plant generations

To use heteroplasmy measurements through developmental history to infer the dynamics of oDNA segregation, we require a quantitative model connecting the behaviour of heteroplasmy at the different developmental and generational timepoints we observe [Wilton et al., 2018; Johnston et al., 2015; Burgstaller et al., 2018; Burian et al., 2016]. We analyzed bulk tissue samples, so cell-to-cell variability cannot be directly quantified; instead, we assume that the heteroplasmy mean in a tissue sample reflects the heteroplasmy of the single cell that was the developmental ancestor of the tissue [Burian et al., 2016; Furner & Pumfrey, 1992; Irish & Sussex, 1992]. This assumption allows for any amount of segregation to occur during the development of the tissue from the precursor cell but assumes there is no systematic shift due to selection for one oDNA type over another (compatible with evidence in this system [Broz et al., 2022] and others [Mandel et al., 2020]).

Given this picture, bulk heteroplasmy samples from different tissues are interpretable as readouts of single-cell heteroplasmy in the population of stem cell precursors to each tissue. For example, mean heteroplasmy samples from three leaves are interpreted as three single cell heteroplasmy values from the (earlier) population of stem cells that gave rise to those leaves. We can then construct a developmental model inspired by the “ontogenetic phylogeny” picture tracking the relationships between cells at different developmental stages [Wilton et al., 2018]. Here, the developmental history of a set of cells is accounted for by a “cell pedigree” or “lineage tree” [Stadler et al., 2021] describing the relationship between ancestral and descended cells. Wilton et al. [2018] used such a picture to infer rates of segregation and mutation through human development given cellular profiles of the presence of different heteroplasmic variants. We will follow this philosophy but instead work with plant development and the continuous heteroplasmy level as it varies through development. This model describes and links the distributions of heteroplasmy in the estimated stem cell populations through and between generations (Fig 1A-B; see Methods). We consider three different models, corresponding to no sequestered germline, separate germline and soma developmental lineages, and a separate developmental lineage for every tissue we consider [Lanfear, 2018] (Fig. 1A). Importantly, our model considers individual heteroplasmy measurements, not their summary statistics – the sample variance, for example, would in most of our cases have a high uncertainty due to the limited number of observations in each sample. Instead of fitting a model using coarse-grained, uncertain summary statistics, we use the full set of individual heteroplasmy measurements in a parametric framework to obtain high statistical power [Giannakis et al., 2023].

The amount of segregation occurring between each developmental period is quantified in our model as “effective segregation events”. This is the number n of binomial cell divisions (and associated

oDNA reamplifications) that would generate the observed heteroplasmy variance, with an effective population size N_e. We use this variable rather than a “bottleneck size” or “drift parameter” [Johnston, 2019b; Wonnapinij et al., 2008] because (a) it corresponds to a biological “null model” where variance is generated by cell divisions alone (see below); and (b) because it is a convenient additive quantity, so that the effective number of segregation events describing n_1 events followed by n_2 events is simply n_1+n_2. We assume, based on biological observations in the *Arabidopsis* germline (see Methods), that $N_e = 50$ for mtDNA [Wang et al., 2010; Preuten et al., 2010] and 7 for ptDNA (the latter corresponding to 7 genetically homogeneous organelles [Greiner et al., 2020; Scarcelli et al., 2016]). We adopt binomial cell divisions and reamplification as a convenient null model with some empirical support [Johnston et al., 2012; Johnston et al., 2015], although mtDNA partitioning in yeast has been observed to be controlled to a tighter extent [Jajoo et al., 2016].

To learn the likely mechanisms of oDNA segregation in real plants, we begin with the dataset from Broz et al. [2022], labelled by different developmental stages (Fig. 1C-E). These stages are early-emerging leaves (EL, fully expanded between 4-6 weeks of growth), late-emerging leaves (LL, upper rosette leaves that were fully expanded after 8 weeks of growth), and inflorescences (INF) (Fig. 1A; see Methods), reflecting tissues generated progressively later in development from the SAM. These data include observations of mtDNA heteroplasmy in wild type and *msh1* mutant backgrounds, and ptDNA heteroplasmy in the *msh1* mutant. All wild type lines measured were homoplasmic in ptDNA, likely due to the high rate of wild type segregation [Broz et al. 2022].

Generation of heteroplasmy variance across tissues and between generations

We first aim to infer the number of effective segregation events at each developmental stage in Fig. 1. We used reversible jump Markov chain Monte Carlo (RJMCMC) [Green, 1995; Dellaportas et al., 2002] with uniform priors over models and all parameters (see Methods) to infer the posterior probability associated with each of the three possible developmental histories in Fig. 1A. This approach produces posterior distributions on each parameter and model index, describing the probability of different mechanisms given the data [Kirk et al., 2013]. We validated this modelling and inference approach with a set of synthetic observations compatible with different mechanisms of variance generation through development and between generations, including cases distinguishing the likely presence of an early germline (Supplementary Fig. S1), and confirmed that inference results were stable across different MCMC chains (Supplementary Fig. S2). Because the statistical approach considers individual heteroplasmy observations rather than losing information with summary statistics like the sample variance, substantial statistical power is retained to infer parameters and select mechanisms [Giannakis et al., 2023].

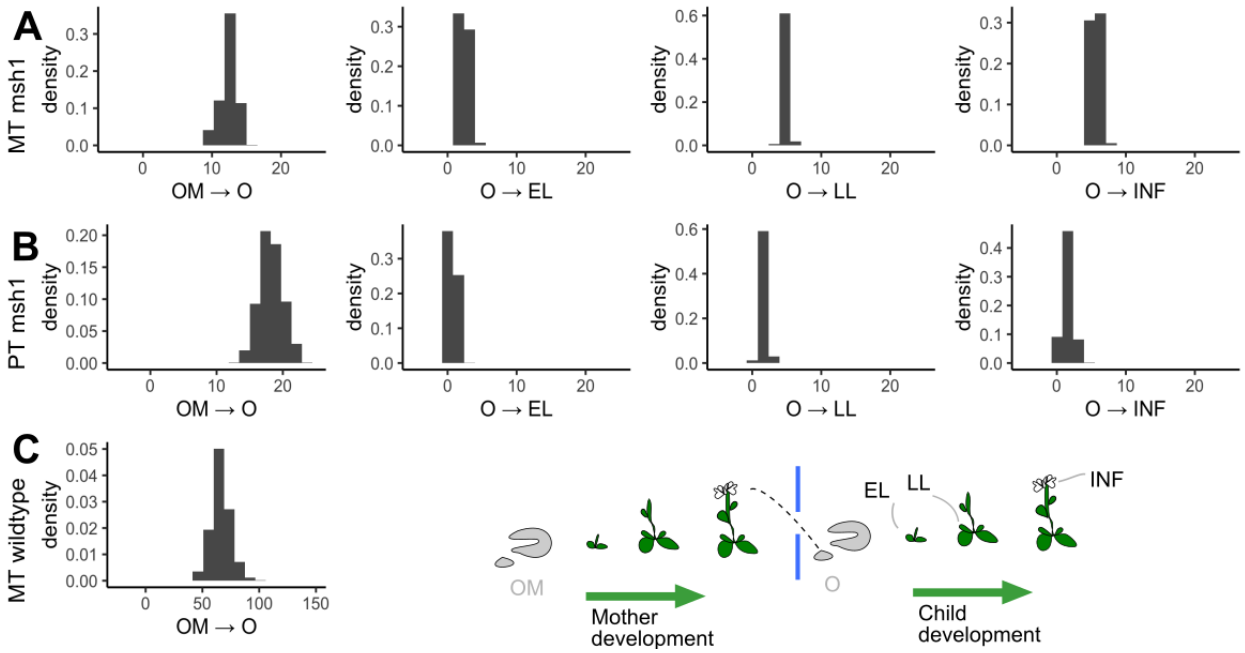


Figure 2. **Posteriors from inference process.** Posterior distributions, inferred across models, for the effective segregation events from a precursor state (O) to different tissue precursors (EL, early leaf; LL, late leaf; INF, inflorescence), and between generations (OM → O): (A) *msh1* mtDNA (N_e = 50), (B) *msh1* ptDNA (N_e = 7); (C) wildtype mtDNA (N_e = 50, different scale).

Fig. 2 shows the inferred posteriors for the number of effective segregation events at different stages of plant development and between generations, integrated over the different model structures in Fig. 1A. As above, this value is the number of binomial cell divisions that would be required to generate the observed heteroplasmy variance, given an effective population size of 50 mtDNAs or 7 ptDNAs per cell.

The amount of segregation occurring between generations (OM→O) is substantially greater than that occurring within a single plant up to the inflorescence stage (O→INF). In the *msh1* mutant, a total of between 9 and 15 events are inferred to occur for mtDNA and between 15 to 25 for ptDNA between generations. In the wildtype, between 50 and 100 events – on average around a seven-fold increase in segregation -- are inferred to occur between generations for mtDNA. These numbers correspond to normalised heteroplasmy variances V'(h) of 0.17-0.26 for *msh1* mtDNA, 0.90-0.98 for *msh1* ptDNA, and 0.64-0.87 for wildtype mtDNA; where the usual "bottleneck size" is 1/V'(h). In all cases, substantial segregation is inferred to occur between the bulk inflorescences of one generation and the early stem cells in the next. This could correspond to the generation of large cell-to-cell variability within the reproductive cells in an inflorescence, matching the generation of variance in female reproductive cells in mammalian systems.

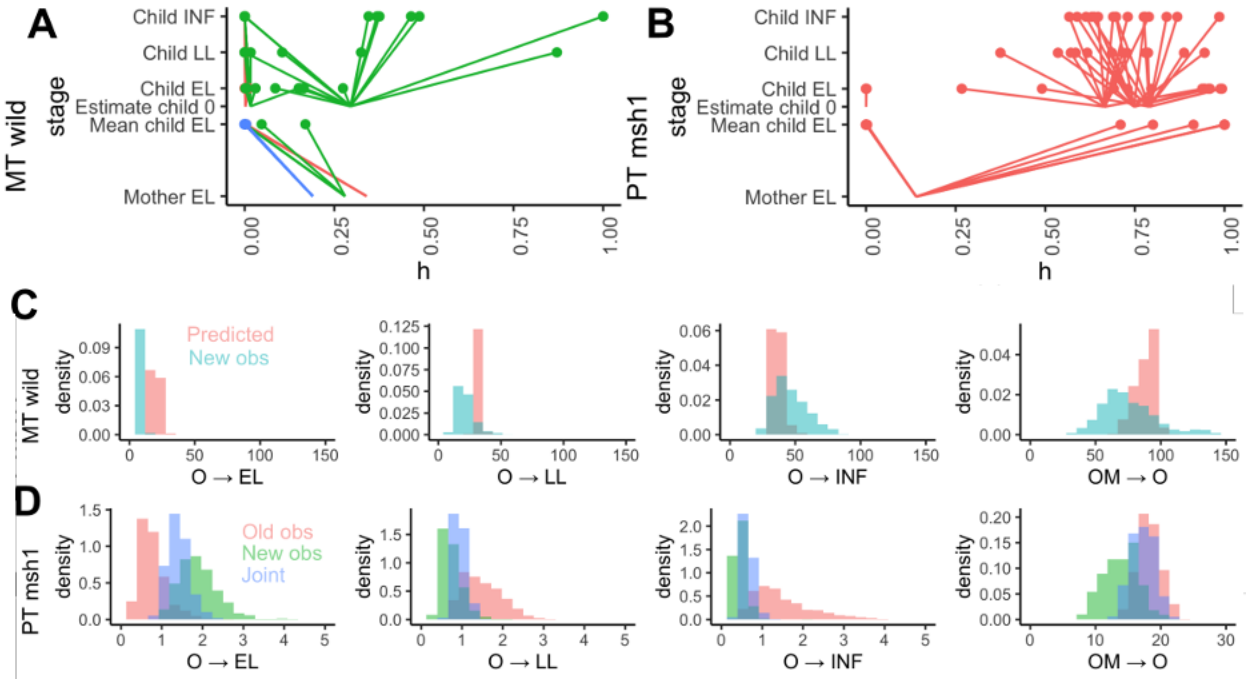


Figure 3. **New data and predicted segregation behaviour.** (A-B) New oDNA observations for (A) wildtype mtDNA and (B) *msh1* ptDNA, displayed as in Fig. 1C-E. (C) Within-plant segregation dynamics for wildtype mtDNA. Predictions (red) from scaling the *msh1* observations seven-fold to match between-generation observations; (blue) inferred effective segregation events from new data. (D) Segregation dynamics of *msh1* ptDNA; previous observations (red); new observations (green); and refined posteriors inferred from the joint dataset (blue).

Segregation differences in samples within a generation were less pronounced, with comparatively few variance-generating events inferred to occur up to the generation of early leaves (sampled at 4-5 weeks of growth), and few more inferred to occur up to late leaf generation (sampled at 8 weeks of growth). The means of each posterior show a roughly linear trend through within-plant development, with heteroplasmy variance increasing through developmental stages; but the extent of this increase is at most half the total segregation between generations.

Due to sampling limitations in Broz et al. [2022], no within-plant samples were generated for wildtype mtDNA, and *msh1* ptDNA sampling was also somewhat limited. Based on the seven-fold scaling of mtDNA segregation from the *msh1* mutant to the wildtype, we hypothesised that the amount of segregation at each within-plant developmental stage would also be scaled seven-fold. We next set out to test this prediction and to verify the results of the ptDNA inference with further experiments.

New heteroplasmy observations support and refine model predictions for segregation dynamics

224 To further illuminate the developmental dynamics of *Arabidopsis* heteroplasmy, we measured
225 mitochondrial heteroplasmy across developmental profiles in lines where MSH1 functionality was
226 recovered by back crossing to a wildtype male, while preserving the heteroplasmy that was present
227 in the female. The heteroplasmy dynamics in these lines are expected to reflect those in the wild
228 type (where heteroplasmy rarely arises because of low mutation rates and the rapid sorting). The
229 new observations are shown in Fig. 3A-B.

230 Matching our predictions, we found dramatically accelerated mtDNA segregation in the wildtype at
231 the late leaf and inflorescence stages, with the rates inferred from new observations compatible with
232 the seven-fold scaling predicted from the between-generations data (Fig. 3C). However, the extent
233 of wildtype mtDNA segregation prior to early leaf development was lower than this hypothesis
234 predicted – and more similar to the lower levels in the *msh1* mutant. This difference suggests a
235 refinement to our predicted picture -- that the increased segregation activity of MSH1 is mainly
236 manifest in later development, which in turn is in qualitative agreement with observed patterns of
237 MSH1 expression (Supplementary Fig. S3).

238 Our new ptDNA observations also matched the predictions inferred from previous data, with the
239 increased volume of observations substantially refining the estimates of variance-generating events
240 at different developmental stages (Fig. 3D). The new observations were always compatible with the
241 (more uncertain) inferred posteriors from the original measurements, and combined provide a tightly
242 defined estimate of segregation dynamics through development. Assuming as before an effective
243 population size $N_e = 7$, the number of variance-generating events is quite limited from early leaf to
244 late leaf to inflorescence, with an over ten-fold further increase in segregation following between
245 generations. It seems likely that this dramatic segregation between generations is due to a severe
246 physical bottleneck on ptDNA, perhaps involving the inheritance of only approximately one
247 homoplasmic organelle (see Discussion).

248 To ask whether within-generation segregation was a genuinely continuous process, we next
249 explored the probability that the magnitude of segregation increased sequentially through
250 developmental stages (for example, whether the amount of segregation experienced by late leaves
251 exceeded that experienced by early leaves). Here we found evidence for continuous segregation
252 through development when functional MSH1 is present, but more limited support when MSH1 was
253 compromised (Fig. 4A). When MSH1 is compromised, segregation patterns can be explained by all
254 segregation occurring in early development prior to early leaf sampling; with functional MSH1,
255 segregation proceeds continuously through development.

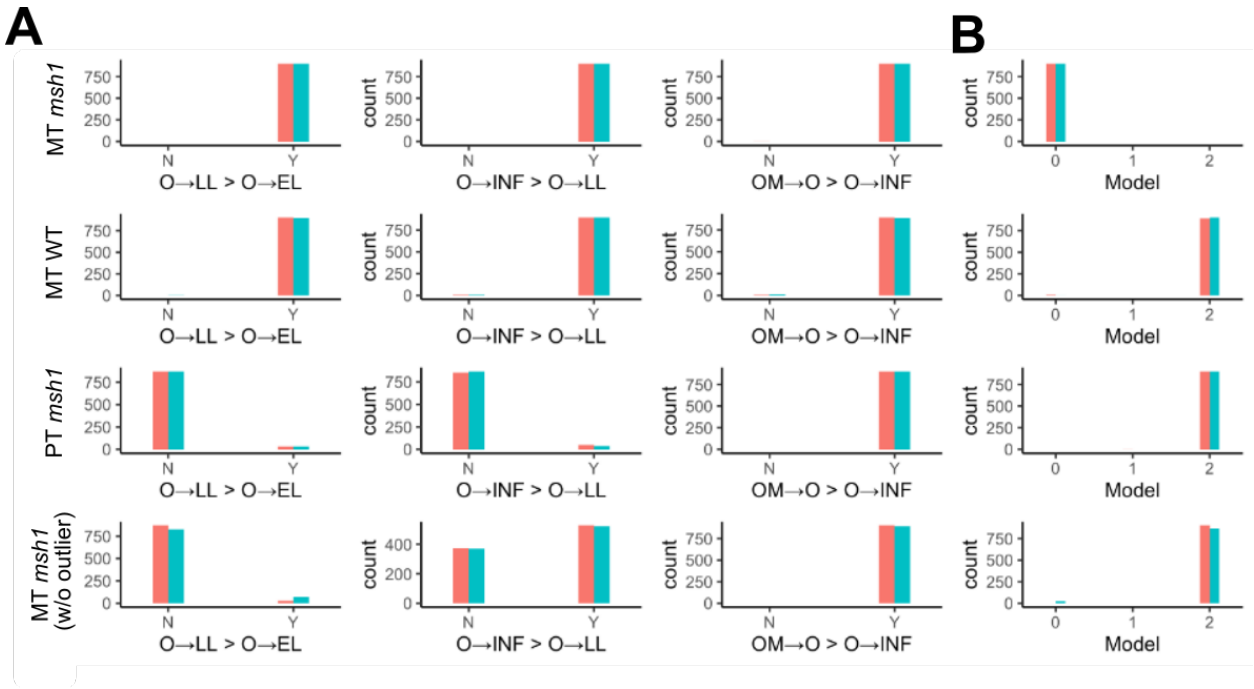


Figure 4. **Patterns and models of segregation through development inferred from combined heteroplasmy profiles.** (A) Evidence for progressive segregation through development. Posterior support for the magnitude of segregation at different developmental stages increasing sequentially – for example, O->LL > O->EL asks whether late leaves have experienced greater segregation than early leaves. (B) Posteriors on model index from reversible jump MCMC. Models 0-2 are respectively the linear germline, separate soma, all separate lineage models from Fig. 1. Rows correspond to different organelle-mutation combinations: ; the final row is the mtDNA *msh1* mutant with one potentially outlier lineage removed (see text). The two colours correspond to results from different RJMCMC simulation to demonstrate convergence (see also Supplementary Fig. S2).

Cell divisions account for oDNA variance in the *msh1* mutant, and gene conversion can account for additional wildtype segregation of mtDNA

Arabidopsis has been estimated to undergo around 34 germline cell divisions between generations [Watson et al., 2016]. In the *msh1* mutant, the number of inferred effective segregation events (averages around 12 for mtDNA and 20 for ptDNA) easily fall within what would be expected from this number of binomial cell divisions for cellular populations of N_e = 50 mtDNAs and N_e = 7 ptDNAs, meaning that the observed heteroplasmy variance could then be readily accounted for through random cell divisions and reamplification alone.

In the wildtype mtDNA, much more segregation is observed than can be accounted for by 34 cell divisions – the average number of inferred events is around 75. Several possibilities exist for the mechanism generating this additional variance. As hypothesised in mammalian systems, partitioning of oDNA clusters, increased random turnover of oDNA, and oDNA replication restricted to a subset of the cellular population can all increase heteroplasmy variance (reviewed in Johnston [2019b]).

However, given the clear difference between the wildtype and *msh1* mutant, we suggest that an MSH1-dependent process may be responsible for this increased segregation in *Arabidopsis*. Following Edwards et al. [2021], we propose that gene conversion may be this process – in the Discussion we consider alternative mechanisms. That reference characterised the contribution of gene conversion to V'(h) as $2(1-f)\,\kappa\, t$, where t is time, f is the proportion of mtDNA molecules in a fused state and thus physically capable of recombination, and κ is the rate of gene conversion between a pair of fused molecules per unit time. As the difference between V'(h) in *msh1* and wildtype mtDNA is roughly 0.5, this expression suggests that a rate of $\kappa = 0.007$ per cell division (corresponding to ~0.1 gene conversion events per mtDNA per cell division; see Methods) would be sufficient to generate the observed segregation patterns over ~34 cell divisions.

This approach employed a linear noise approximation that may be challenged by the substantial segregation magnitudes involved in this system. To check these results, we constructed a stochastic model for oDNA during development, including binomial cell divisions, random reamplification between divisions, and a variable rate of gene conversion in a population of $N_e = 50$ oDNA molecules (see Methods). We asked what rates of gene conversion were required to generate the observed V'(h) within ~34 cell divisions, finding support for a figure around 0.25 events per mtDNA per cell cycle (Supplementary Fig. S4). This combined model provides predictions for heteroplasmy distributions at any given stage of plant development (Supplementary Fig. S5). We should note that this gene conversion activity could be partitioned into more intense bursts in reduced developmental stages to achieve the same variance generation – as suggested by the new mtDNA observations in Fig. 3, where early meristem development appears not to generate as much segregation as later developmental stages. Such a partition of activity would agree with observed patterns of *MSH1* expression during plant development (Supplementary Fig. S3) and the observed physical behaviour of mitochondria, forming a reticulated network in the shoot apical meristem, with the potential to facilitate recombination between mtDNA molecules [Seguí-Simarro & Staehelin, 2009; Edwards et al., 2021].

Plant germline history

The posterior distributions we have presented are integrated over all the model structures in Fig. 1A, so that they reflect "universal" behaviour regardless of the support for the individual models. However, the RJMCMC process also quantifies this support for the different models of the plant germline. Interestingly, we initially observed some diversity in the posterior distributions over this model index (Fig. 4B). The mtDNA *msh1* data has strong support for the "linear germline" model, while the mtDNA wildtype and ptDNA *msh1* data provide strong support for the "all separate

lineages" model (Supplementary Fig. S2).

To interpret these findings, it helps to consider the behaviour of heteroplasmy statistics under the different models. Under separate lineages, samples from different developmental stage (EL, LL, INF) reflect the mean heteroplasmy of an early stem cell (CS1) and have independent variances. Under a linear germline, progressive sampling events form the precursor state for each stage. Differences in heteroplasmy mean can therefore arise due to this sampling, and the variance for each stage is "overlaid" on top of any such mean variability (Fig. 1B). Observations where mean heteroplasmy shifts between developmental stages are therefore more compatible with a linear germline model; limited or no shifts in mean heteroplasmy may select the separate lineages model as more flexible.

The mtDNA *msh1* data has one developmental lineage in particular that suggests a strong shift in mean heteroplasmy (top right of Fig. 1C), where an initial heteroplasmy around 0.85 gives rise to several homoplasmic late leaves and inflorescences. If this lineage is removed from the dataset, the results of inference fall more in line with the other systems (Fig. 4). If that lineage is regarded as an outlier corresponding to an accident of sampling – where, for example, other heteroplasmic inflorescences may have existed to bring the mean heteroplasmy back down – then all the remaining data support a model where segregation proceeds independently in different tissue types after diverging from a developmentally early source. There is thus at least some support for the heteroplasmy profiles in inflorescences and leaf tissue developing independently [Lanfear, 2018], although further characterisation of somatic heteroplasmy in wildtype lineages will help resolve this question.

Discussion

We have shown, with a combination of oDNA measurements from heteroplasmic plant lines and mathematical modelling, how oDNA segregation proceeds through plant development and between generations (Fig. 5). New experiments support the predictions of the inferred mathematical models; the models make further predictions about heteroplasmy distributions at any stage of plant development (Supplementary Fig. S5). We have shown that in the absence of MSH1 functionality, oDNA segregation can largely be accounted for by the physical process of binomial partitioning at cell divisions. Although other mechanisms likely support some gene conversion activity in the absence of MSH1, high rates of such activity are not required to explain observed segregation patterns in the mutant. By contrast, MSH1 functionality induces a seven- to ten-fold increase in segregation strength, leading to rapid shifts towards homoplasmy, which cannot be explained by cell divisions alone.

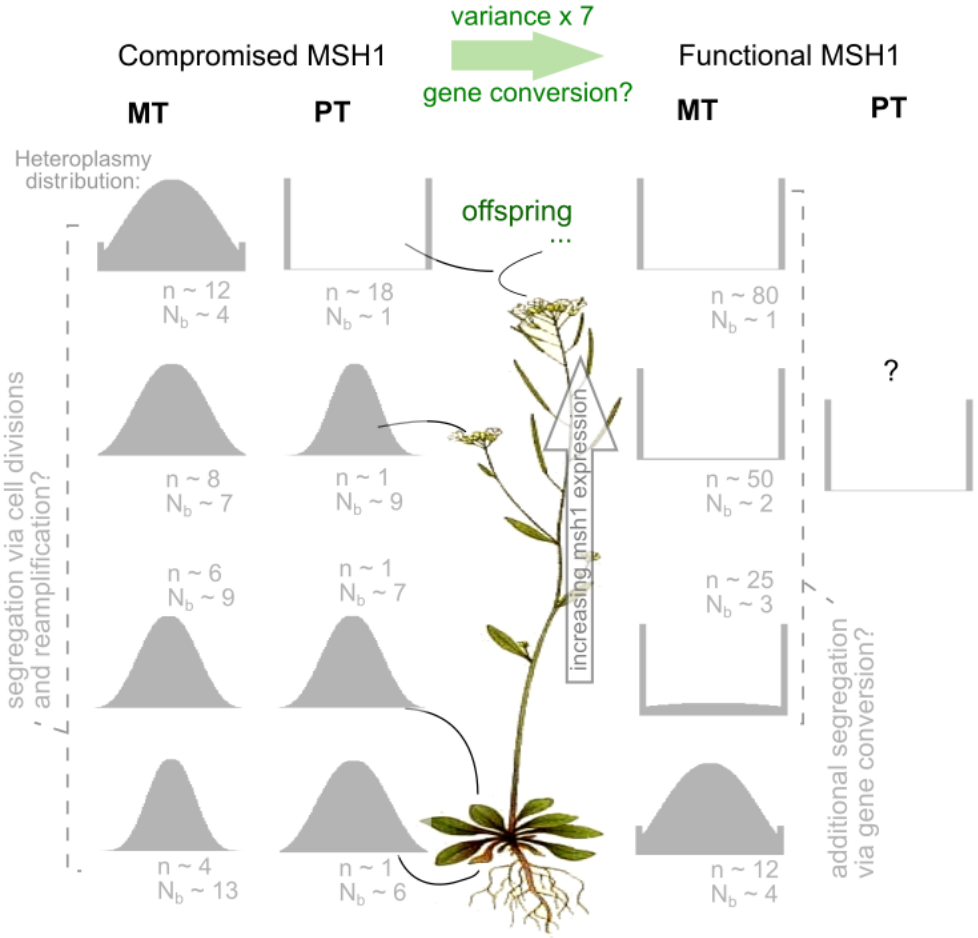


Figure 5**. Summary of inferred segregation dynamics within plants and between generations.** Illustrative distributions of heteroplasmy, corresponding to the inferred mean segregation magnitude (*n* segregating events, for N_e = 50 mtDNAs or N_e = 7 ptDNAs; and N_b, effective bottleneck size). Distributions at each developmental stage, and an initial heteroplasmy of 0.5, are shown for mtDNA (MT) and ptDNA (PT) in wildtype and *msh1* mutants (all wildtype PT observations are homoplasmic, so no inference is possible; see Discussion for hypotheses). Continuous segregation is supported by inference in all systems except PT; model selection suggests most support for a picture where separate developmental lineages are involved for each developmental stage.

We do not have measurements of heteroplasmic ptDNA on the wildtype background – all lines measured so far have been homoplasmic. The predictions of this theory for wildtype plastid heteroplasmy dynamics depend on the spatial arrangement of ptDNA information. If ptDNA within a single plastid is homoplasmic, and heteroplasmy arises from a mixture of internally homoplasmic organelles, then the effect of functional gene conversion will be limited. This is because each ptDNA will usually only be physically colocalised with an identical partner, leaving no capacity to change genetic identity. If, however, plastids are internally heteroplasmic, functional gene conversion may act to further speed up segregation. In this case, following observations for mtDNA, we would expect

roughly seven times as many effective cell divisions to take place (matching the mtDNA case), leading to an effective 150-200 cell divisions for the $N_e = 7$ case. This would lead to homoplasmy in all but a very small proportion of offspring (as observed).

The quantitative details of our model depend on some assumptions, including a binomial division – random reamplification model for oDNA at cell divisions, the Kimura model for oDNA heteroplasmy, and particular choices for effective population size of oDNAs. The choices we have made have support from the literature (see Methods), but are not expected to be universally true or perfectly precise single values. oDNA population sizes change through development (see Methods and references therein) and oDNA partitioning at cell divisions may be more or less tightly controlled than a binomial distribution [Jajoo et al., 2016; Johnston et al., 2015]. Our effective ptDNA population size is based on a picture where ptDNA populations inside individual plastids are homogeneous: this assumption may be challenged in the case of recent *de novo* mutations that have not yet fixed within an organelle. The results we report – the relative magnitudes of segregation at different developmental stages, the difference between wildtype and *msh1* lines, the role for gene conversion, and the agreement of new experiments with theoretical predictions – are robust with respect to different choices of these parameters. The specific numbers of segregating events we infer should be interpreted as effective quantities, reflecting biological reality if our parameter choices are accurate, otherwise requiring some scaling (see Methods and Supplementary Fig. S6) for a precise quantitative connection to other conditions.

The indirect evidence from our study, after removing a potential outlier lineage, supports a picture where different tissues, including the germline, have different developmental lineages after diverging from a developmental ancestor [Lanfear, 2018]. A previous study in carrot [Mandel et al. 2020] did not find shifts of mean heteroplasmy during development, although some individual observations suggested the capacity for large minor allele amplification; this picture would also be compatible with segregation (increasing variance) through independent developmental lineages. Regardless of the within-plant model, most of the between-generation segregation we observe occurs between the inflorescences of the mother and the early meristem of the offspring. For plastids in particular, it seems likely that this strong segregation may be in part due to a physical bottleneck, where a small number – perhaps just one in some cases – of homoplasmic organelles are inherited. For mitochondria, our observations support a picture where some segregation occurs progressively through development. The rate of this increase is limited in the *msh1* mutant and clearer in the wildtype, and its magnitude is smaller than the between-generation shift.

Substoichiometric shifting (SSS) involves the sudden amplification of a rare mtDNA type (a sublimon) to dominance [Abdelnoor et al., 2003; Arrieta-Montiel et al., 2001; Woloszynska, 2010]. The dynamics characterised here illustrate how this amplification may occur. Even if a sublimon is

present only rarely in SAM cells, if one of those cells becomes the precursor to a plant branch or organ, the sublimon can very naturally (and quickly) come to dominate that branch or organ (and hence offspring from it). Our work here quantifies how this shifting may occur across different organs in a plant, leading to inherited differences. In a similar vein, branch-to-branch differences in variegation caused by oDNA features have been recognised for over a century (initially laying the foundation for the understanding of cytoplasmic inheritance [Hagemann, 2010]). Such branch-to-branch differences are caused by the segregation of oDNA from an initially heteroplasmic state across different parts of the plant. The quantitative model we present links, for example, the unobservable initial inherited heteroplasmy to the proportion of different variegated phenotypes throughout the plant, by quantifying the extent of segregation through different periods of plant development.

Observations here and in Broz et al. [2022] point to MSH1 dramatically accelerating oDNA segregation. We have proposed that this acceleration may be due to gene conversion. However, the function and mechanism of action of MSH1 in plants remain debated. Evidence certainly points to its role in the control of oDNA recombination (often described as recombination surveillance [Abdelnoor et al., 2003; Shedge et al., 2007]). Its unusual structure - - including an endonuclease domain - - has led to the suggestion that it induces double stand breaks that then provide the substrates for gene conversion [Christensen, 2014]. The heteroplasmy measurements here strongly suggest that MSH1 acts to generate high cell-to-cell variance in oDNA heteroplasmy through plant development. Theory has suggested gene conversion as one plausible mechanism with desirable properties [Edwards et al., 2021]. However, it may be that MSH1 generates heteroplasmy variance via another mechanism. Depletion of oDNA copy number, for example, would impose a physical bottleneck on the population, both amplifying variability from divisions and inducing variability from subsampling the population. If MSH1 acts to deplete oDNA, these effects could be of comparable or greater importance in generating variability, depending on the quantities involved [Cree et al., 2008; Johnston et al., 2015]. Broz et al. [2022] showed that oDNA copy number was not significantly impacted in leaves of MSH1 versus wildtype plants, but it is unknown whether these results reflect oDNA levels in germline. If, in some way, MSH1 enforces replication of a subset of oDNA molecules as proposed by Wai et al. [2008] in a mammalian context, this mechanism could also explain the observed segregation. While the evidence points towards a more direct link between MSH1 and gene conversion [Wu et al., 2020; Broz et al., 2022], we cannot completely discard these hypotheses without measurements of copy number and oDNA replication activity. We were unable to find or acquire estimates for absolute rates of oDNA recombination in *Arabidopsis*; future estimates of these quantities will help provide further evidence for these mechanisms. It is noteworthy that *MSH1* expression is increased relative to other tissues in the meristem in *Arabidopsis* and other species

(Supplementary Fig. S3, [Edwards et al., 2021]), and that mitochondria physically fuse to a greater extent in the meristem cells [Seguí-Simarro & Staehelin, 2009; Edwards et al., 2021]. Physical colocalization of mitochondria is a prerequisite for mtDNA interaction and recombination [Logan, 2006; Arimura, 2018; Giannakis et al., 2022], and the collective dynamics of mitochondria are altered in the *msh1* mutant, potentially as a compensatory response to support more interaction [Chustecki et al., 2022; Chustecki et al., 2021]. Together, these observations suggest a linked physical and genetic axis of control acting to shape oDNA through plant generations.

Methods

Plant material and growth

The initial generation and selection of heteroplasmic plant lines is described in Broz et al. [2022]. Here, plants of the homozygous *msh1* (At3g24320) mutant line CS3372 (*chm1-2*) were used for analysis of plastid heteroplasmy. For mitochondrial heteroplasmy analysis in a wild type background, maternal lines of *msh1* CS3246 (*chm1-1*) were crossed with wildtype males to generate F1 progeny. These different *msh1* alleles were used because it was on these backgrounds that oDNA variants present at reasonable allele frequencies arose and were retained (see below); both have been reported to be full allelic knockouts [Broz et al., 2022]. All progeny were confirmed to be heterozygous for MSH1. Seeds of desired lines were vernalized in water at 4 °C for 3 days, sown in 3 inch pots containing Pro-Mix BX media and grown under short day conditions (10 h light / 14 h dark) on light racks with fluorescent bulbs (~150 μE m^{-2} s^{-1}) at ambient temperature (~25 °C). An initial fully expanded rosette leaf sample was taken at 4 weeks of growth to identify heteroplasmic individuals. Three additional leaves were sampled at 5 weeks of growth. These 4-5 week old leaf samples are considered “early leaf” (EL) for subsequent analyses. At 8 weeks, four additional leaf samples were taken. Two were harvested from the base of the rosette. These leaves were already fully expanded at 5 weeks and emerged from the SAM around the same time as the EL samples described. Thus, these are also considered "EL". Two additional fully expanded leaves were harvested at 8 weeks from the top of the rosette, emerging from the SAM at a later timepoint that ELs, and are considered as late leaf "LL" in the analysis. Inflorescence tissue (INF) was harvested after plants began to bolt.

Heteroplasmy measurements

DNA extraction and heteroplasmy analysis were performed as described previously [Broz et al. 2022]. Briefly, single nucleotide variants (SNVs) in oDNA of *msh1* mutant lines were identified by

sequencing [Wu et al. 2021] and ddPCR assays were designed to track these SNVs within plants and between generations. Allele specific primers and probes were designed to each SNV and droplet generation and reading was performed using Bio-Rad QX200 system. This study used the specific loci plastid 26553, mitochondria 91017 and mitochondria 334038, which were retained after screening the original set of heteroplasmic variants for those present at moderate allele frequencies. A correction factor was applied to mitochondrial data to account for the amplification of nuclear copies of the mitochondrial genome (numts) found in Arabidopsis. Specifically, the large numt on chromosome 2 is too similar to actual mtDNA to be distinguished with short reads or ddPCR markers. So we approximate the number of nuclear genome copies in the sample (which would inflate the number of apparent mitochondrial “wild type” alleles) and correct accordingly. All nupts have enough sequence divergence that nuclear and plastid copies can be unambiguously distinguished.

Developmental history models

First picture a fertilised zygote giving rise to an early population of stem cells. At some developmental time point this population will contain the single ancestral cell of all early leaf samples, as well as of cells that will continue to proliferate in the SAM. At a later time point, the new SAM population will contain the ancestor for all late leaf samples, as well as for further proliferating cells. At a still later time point, the new SAM population will contain the ancestral cell to all inflorescence samples. Inflorescences are interpreted as containing the egg cells for the next generation, in which the developmental outline above is repeated for each single fertilised zygote. Each tissue's heteroplasmy value is drawn from a distribution describing some amount of segregation acting on developing descendants of these ancestral stem cells, with relationships described via the “cell pedigrees” or “lineage trees” in Fig. 1A [Wilton et al., 2018; Stadler et al., 2021].

The developmental history of plant germlines is debated [Lanfear, 2018]. To compare hypotheses on plant germline behaviour, we also consider two additional alternative models. In Fig. 1B, the future germline is sequestered early in development and then develops in parallel to the somatic tissues. Here, the model is as above, except the inflorescence ancestral cell is drawn from the early stem cell population. In Fig. 1C, separate somatic lines also exist, so that the different organs all develop independently from an original early precursor. In theory, different germline histories – where soma and germline are sequestered at different developmental timepoints – will give rise to different correlations and variance structures in the oDNA populations in different tissue types. For example, if the germline develops independently of the soma, correlations between mean oDNA heteroplasmy

in somatic and inflorescence samples are less likely, and it may be possible for inflorescence oDNA to have lower variance than soma oDNA. If the germline shares a common developmental ancestry with the soma, correlations are more likely, and inflorescence variance will be at least as high as soma variance.

Inference of segregation dynamics

To assign a likelihood to our tissue observations given a developmental model, we need to (a) estimate the ancestral cell heteroplasmies and (b) estimate the probability of observing a tissue heteroplasmy given the ancestral value and some parameterised description of segregation [Burgstaller et al.,2014; Burgstaller et al., 2018]. For (a), we treat ancestral cell heteroplasmies as latent variables and integrate the likelihood over all possible values for each. For (b), we use the Kimura distribution [Wonnapinij et al., 2008; Kimura, 1955] to describe the probability of observing a given heteroplasmy in individual tissue samples, creating a stochastic model with a full likelihood function [Giannakis et al., 2022b, Broz et al., 2022]. We change variables from the “drift parameter” b to an effective number of variance-generating events $n = \log b / (1 - 1/N_e)$ (see below) to provide a convenient, additive parameter for serial segregation events. The corresponding likelihood is then used in a reversible jump Markov chain Monte Carlo (RJMCMC) framework [Green, 1995; Dellaportas et al., 2002] (see below) with uninformative uniform priors on initial heteroplasmies and division numbers and compute posterior distributions over these parameters.

For numerical efficiency, we precompute Kimura distributions for 0 to 200 cell divisions and initial heteroplasmies from 0 to 1 in steps of 0.01 and use these precomputed distributions as a lookup table in the inference process. For numerical efficiency, we set effective population size to 50. A post-hoc correction can be used to interpret the results from this setup in terms of any other population size (see below).

To account for the fact that heteroplasmy measurements may have some associated uncertainty, we implement a degree of granularity within the model. For example, a granularity of 0.01 means that heteroplasmy values are rounded to the nearest 0.01. This both allows for measurement noise and improves computational speed; we will show that our results are robust to different choices of this parameter.

We write $\{\underline{D_i}\} = \{\underline{D_{i,ME}}, \underline{D_{i,CE}}, \underline{D_{i,CL}}, \underline{D_{i,CI}}\}$ for the set of observations in family i, with elements respectively corresponding to Mother Early leaf, Child Early leaf, Child Late leaf, and Child Inflorescence. We write S_{Cj} for the latent variable associated with ancestral cell heteroplasmy at developmental stage j. The likelihood associated with measurements, in the model without a segregated germline, is then

$$L(\{\underline{D_i}\} \mid \underline{n}, h_{0i}) = P(\underline{D_{i,ME}} \mid h_{0i}; n_0) \int dS_{C1}\, P(S_{C1} \mid h_{0i}; n_0+n_1+n_2+n_3)\, P(\underline{D_{i,CE}} \mid S_{C1}; n_0)$$
$$\times \int dS_{C2}\, P(S_{C2} \mid S_{C1}; n_0)\, P(\underline{D_{i,CL}} \mid S_{C2}; n_1) \int dS_{C3}\, P(S_{C3} \mid S_{C2}; n_1)\, P(\underline{D_{i,CI}} \mid S_{C3}; n_2), \quad [1]$$

So that SC1 is the precursor to EL and SC2, SC2 is the precursor to LL and SC3, and SC3 is the precursor to INF (Fig. 1A). With a segregated germline the corresponding expression is

$$L(\{\underline{D_i}\} \mid \underline{n}, h_{0i}) = P(\underline{D_{i,ME}} \mid h_{0i}; n_0) \int dS_{C1}\, P(S_{C1} \mid h_{0i}; n_2+n_3)\, P(\underline{D_{i,CE}} \mid S_{C1}; n_0)\, P(\underline{D_{i,CI}} \mid S_{C1}; n_2)$$
$$\times \int dS_{C2}\, P(S_{C2} \mid S_{C1}; n_0)\, P(\underline{D_{i,CL}} \mid S_{C2}; n_1), \quad [2]$$

So that SC1 is the precursor to EL, INF, and SC2, and SC2 is the precursor to LL. With completely separate developmental lineages we have

$$L(\{\underline{D_i}\} \mid \underline{n}, h_{0i}) = P(\underline{D_{i,ME}} \mid h_{0i}; n_0) \int dS_{C1}\, P(S_{C1} \mid h_{0i}; n_2+n_3)\, P(\underline{D_{i,CE}} \mid S_{C1}; n_0)\, P(\underline{D_{i,CL}} \mid S_{C1}; n_1)$$
$$\times P(\underline{D_{i,CI}} \mid S_{C1}; n_2), \quad [3]$$

So that SC1 is the precursor to all lineages, which develop independently.

An important difference between the models is whether samples at different stages can have different population means. In the separate lineages model, EL, LL, and INF pedigrees all come from the same precursor, so have the same population mean. In the linear model, each pedigree begins with a (latent) sample from a previously segregated population (Fig. 1B), so population means can differ (Supplementary Fig. S1). They also differ in the accumulated amount of segregation at the population level. The “linear germline” model enforces a monotonic increase in segregation (hence in V’(h)) through development – hence EL ≤ LL ≤ INF ≤ cross-generation. The “all separate” model supports a more flexible picture where INF < EL, for example. However, although these relationships hold statistically at the population level, a given set of samples may not reflect them: for example, a sample of inflorescences may not capture the full possible spread of values and may thus suggest a lower variance than the true case. The full likelihood-based inference process below accounts for these sampling issues.

Given one of the above likelihood functions for a family set of observations $\{\underline{D_i}\}$, the likelihood associated with a full set of observations is

$$L(D \mid \underline{n}, \underline{h_0}) = \Pi_{\text{families } i}\, L(\{\underline{D_i}\} \mid \underline{n}, h_{0i}) \quad [4]$$

Effective population sizes

Preuten et al. [2010] find 50 or fewer mtDNAs in stems and flowers. Wang et al. [2010] found egg cells from Arabidopsis to possess 59.0 copies of mtDNA on average. Gao et al. [2018] do not quantify mtDNA molecules in *Arabidopsis* but observe around 250 mtDNA nucleoids in mature eggs and mature zygotes, and 100-200 mtDNA nucleoids per cell during embryogenesis, with a doubling between early apical cells and mature apical cells. We choose an effective population size of 50 for consistency with those studies where mtDNA copy number is more directly observed.

In a comprehensive survey across species, Greiner et al. [2020] report an increase in plastids per cell in *Arabidopsis* development from 4-10 in the meristematic region, through 22-34 in young leaves, to 50-90+ in mature leaves. Corresponding ptDNA counts per plastid (per cell) are given as 8-21 (71-146), 48-84 (997-2476), 79-121 (2900-5500+). We choose an effective population size of 7, corresponding to the central estimate for the meristematic observations, and assuming that plastids are internally genetically homogeneous [Scarcelli et al., 2016]. This assumption may be challenged in the case of recent mutations (see Discussion).

For numerical convenience we used a population size of $N_e = 50$ in the numerical simulations. Following the usual parameterisation of the Kimura distribution for mtDNA work [Wonnapinij et al. 2010, Giannakis et al. 2023],

$$b = \exp(-n / N_e)$$

$$\approx (1 - 1/N_e)^n, \qquad [5]$$

Using this approximation (which is not perfect for low N_e) we can immediately interpret an inferred value of n for N_e as equivalent to a value n' for N'_e :

$$(1 - 1/N_e)^n = (1 - 1/N'_e)^{n'}$$

$$n \log (1 - 1/ N_e) = n' \log (1 - 1/ N'_e)$$

$$n' = n \log (1 - 1/ N_e) / \log (1 - 1/ N'_e) \qquad [6]$$

so that, for example, $n = 10$ divisions for $N_e = 50$ give roughly the same heteroplasmy distribution as $n' = 20$ divisions for $N_e = 100$. We can then scale the results for $N_e = 50$, chosen for numerical convenience in our simulation, to the required effective population size in our estimates of biological reality. Hence, any of the inferred numbers n of segregating events we report (using $N_e = 50$ for mtDNA and $N_e = 7$ for ptDNA) can readily be interpreted for another effective population size N_e' by multiplying by the factor $\log (1 - 1/ N_e) / \log (1 - 1/ N'_e)$, which for most values is close to N_e/N_e' (Supplementary Fig. S6). Finally, effective “bottleneck size” N_b (the effective population size if variance is generated by a single event) can be recovered from our inferred n with

$$N_b = 1 / (1-(1-1/N_e)^n) \quad [7]$$

Reversible jump MCMC

We use reversible jump MCMC to identify the support for different models of developmental histories [Green, 1995; Dellaportas et al., 2002; Kirk et al., 2013]. We explored several options for relating parameters in each model class, which all gave convergent results in the long-term limit of the MCMC chains, but found the best mixing between model classes to be achieved simply using $n_i^{(1)} = n_i^{(2)} = n_i^{(3)}$ for all developmental stages *i* and with model classes given by superscripts (1: linear germline; 2: separate germline; 3: all separate lineages), enforcing these (and preserving h_0 values) as deterministic proposal rules upon a proposed shift from model *i* to model *j*. These expressions immediately provide the (trivial) mapping functions $g_{ij}(\underline{n^{(i)}})$ for implementing such a step from model *i* to model *j* [Green, 1995; Dellaportas et al., 2002]. All models have the same dimensionality and the Jacobean determinants associated with each of these mapping functions are all one. We employ uniform priors on all parameters and model indices, making the acceptance rule for the RJMCMC implementation equivalent to the normal Metropolis-Hastings acceptance rule when a between-model step is proposed. We propose such steps with probability 1/3, employing the above perturbation to parameters when this option is not chosen. MCMC chains were run over 10^5 samples, discarding 10^4 as burn-in and subsequently recording every 10th sample.

Estimating and simulating variance due to gene conversion

The parameter κ in the main text is the rate constant associated with the gene conversion processes WT+MU → WT+WT and WT+MU → MU+MU [Edwards et al., 2021]. In a simple picture we could assume that half our N_e = 50 mtDNAs are WT and half are MU. Then the rate of gene conversion is $\kappa \times 25 \times 25$, which for κ = 0.007 per cell division gives ~4 events per cell division or ~4/50 = 0.08 events per mtDNA per cell division.

The derivation of this expression depends on a linear noise approximation, and the rates in the above argument will of course vary as segregation proceeds. To provide a more precise estimate, we implemented a simple stochastic simulation of binomial cell divisions, random re-amplification, and gene conversion in a model cellular population. We simulated these processes for various gene conversion rates and 300 cell divisions and asked what gene conversion rates were needed to generate a given normalised heteroplasmy variance V'(h) within ~34 cell divisions (Supplementary Fig. S4).

Data and code availability

All data and code is freely available at https://github.com/StochasticBiology/plant-segregation. The inference code is written in C; the data curation and visualisation is written in R [R Core Team, 2022], using libraries readxl [Wickham and Bryan, 2022], stringr [Wickham, 2019], ggplot2 [Wickham, 2016], and gridExtra [Auguie, 2017].

Acknowledgements

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). IGJ gratefully acknowledges support from the Peder Sather Center. DBS and AKB are supported by NIH Grant R01 GM118046. The authors are grateful to Ben Williams for valuable discussion.

References

Abdelnoor, R.V., Yule, R., Elo, A., Christensen, A.C., Meyer-Gauen, G. and Mackenzie, S.A., 2003. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proceedings of the National Academy of Sciences, 100(10), pp.5968-5973.

Albert, B., Godelle, B., Atlan, A., De Paepe, R. and Gouyon, P.H., 1996. Dynamics of plant mitochondrial genome: model of a three-level selection process. Genetics, 144(1), pp.369-382.

Allen, J.F. and Martin, W.F., 2016. Why have organelles retained genomes?. Cell systems, 2(2), pp.70-72.

Arrieta-Montiel, M., Lyznik, A., Woloszynska, M., Janska, H., Tohme, J. and Mackenzie, S., 2001. Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. Genetics, 158(2), pp.851-864.

Arrieta-Montiel, M.P., Shedge, V., Davila, J., Christensen, A.C. and Mackenzie, S.A., 2009. Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics, 183(4), pp.1261-1268.

Atlan, A. and Couvet, D., 1993. A model simulating the dynamics of plant mitochondrial genomes. Genetics, 135(1), pp.213-222.

Arimura, S.I., 2018. Fission and fusion of plant mitochondria, and genome maintenance. Plant physiology, 176(1), pp.152-161.

Auguie, B., 2017. _gridExtra: Miscellaneous Functions for "Grid" Graphics_. R package version 2.3, <https://CRAN.R-project.org/package=gridExtra>.

Bentley, K.E., Mandel, J.R. and McCauley, D.E., 2010. Paternal leakage and heteroplasmy of mitochondrial genomes in Silene vulgaris: evidence from experimental crosses. *Genetics*, *185*(3), pp.961-968.

Broz, A.K., Keene, A., Fernandes Gyorfy, M., Hodous, M., Johnston, I.G. and Sloan, D.B., 2022. Sorting of mitochondrial and plastid heteroplasmy in Arabidopsis is extremely rapid and depends on MSH1 activity. Proceedings of the National Academy of Sciences, 119(34), p.e2206973119.

Burgstaller, J.P., Johnston, I.G., Jones, N.S., Albrechtova, J., Kolbe, T., Vogl, C., Futschik, A., Mayrhofer, C., Klein, D., Sabitzer, S. and Blattner, M., 2014. MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. Cell reports, 7(6), pp.2031-2041.

Burgstaller, J.P., Kolbe, T., Havlicek, V., Hembach, S., Poulton, J., Piálek, J., Steinborn, R., Rülicke, T., Brem, G., Jones, N.S. and Johnston, I.G., 2018. Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. Nature communications, 9(1), pp.1-12.

Burian, A., De Reuille, P.B. and Kuhlemeier, C., 2016. Patterns of stem cell divisions contribute to plant longevity. Current Biology, 26(11), pp.1385-1394.

Christensen, A.C., 2014. Genes and junk in plant mitochondria—repair mechanisms and selection. Genome biology and evolution, 6(6), pp.1448-1453.

Chustecki, J.M., Gibbs, D.J., Bassel, G.W. and Johnston, I.G., 2021. Network analysis of Arabidopsis mitochondrial dynamics reveals a resolved tradeoff between physical distribution and social connectivity. *Cell systems*, *12*(5), pp.419-431.

Chustecki, J.M., Etherington, R.D., Gibbs, D.J. and Johnston, I.G., 2022. Altered collective mitochondrial dynamics in the Arabidopsis msh1 mutant compromising organelle DNA maintenance. Journal of experimental botany, 73(16), pp.5428-5439.

Clegg, M.T., Gaut, B.S., Learn Jr, G.H. and Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. Proceedings of the National Academy of Sciences, 91(15), pp.6795-6801.

Cree, L.M., Samuels, D.C., de Sousa Lopes, S.C., Rajasimha, H.K., Wonnapinij, P., Mann, J.R., Dahl, H.H.M. and Chinnery, P.F., 2008. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. Nature genetics, 40(2), pp.249-254.

Day, A. and Madesis, P., 2007. DNA replication, recombination, and repair in plastids. In Cell and molecular biology of plastids (pp. 65-119). Springer, Berlin, Heidelberg.

Dellaportas, P., Forster, J.J. and Ntzoufras, I., 2002. On Bayesian model and variable selection using MCMC. Statistics and Computing, 12(1), pp.27-36.

Edwards, D.M., Røyrvik, E.C., Chustecki, J.M., Giannakis, K., Glastad, R.C., Radzvilavicius, A.L. and Johnston, I.G., 2021. Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. PLoS biology, 19(4), p.e3001153.

Fernandes Gyorfy, M., Miller, E.R., Conover, J.L., Grover, C.E., Wendel, J.F., Sloan, D.B. and Sharbrough, J., 2021. Nuclear–cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. *The Plant Journal*, *108*(1), pp.219-230.

Furner I, J. and Pumfrey, J.E., 1992. Cell fate in the shoot apical meristem of Arabidopsis thaliana. Development, 115(3), pp.755-764.

Galtier, N., 2011. The intriguing evolutionary dynamics of plant mitochondrial DNA. BMC biology, 9(1), pp.1-3.

Gao, L., Guo, X., Liu, X.Q., Zhang, L., Huang, J., Tan, L., Lin, Z., Nagawa, S. and Wang, D.Y., 2018. Changes in mitochondrial DNA levels during early embryogenesis in Torenia fournieri and Arabidopsis thaliana. The Plant Journal, 95(5), pp.785-795.

Giannakis, K., Arrowsmith, S.J., Richards, L., Gasparini, S., Chustecki, J.M., Røyrvik, E.C. and Johnston, I.G., 2022. Evolutionary inference across eukaryotes identifies universal features shaping organelle gene retention. Cell Systems.

Giannakis, K., Broz, A.K., Sloan, D.B. and Johnston, I., 2023. Avoiding misleading estimates using mtDNA heteroplasmy statistics to study bottleneck size and selection. G3: Genes|Genomes|Genetics*,* jkad068

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika, 82(4), pp.711-732.

Greiner, S., 2012. Plastome mutants of higher plants. In Genomics of chloroplasts and mitochondria (pp. 237-266). Springer, Dordrecht.

Greiner, S., Golczyk, H., Malinova, I., Pellizzer, T., Bock, R., Börner, T. and Herrmann, R.G., 2020. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. The Plant Journal, 102(4), pp.730-746.

Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F. and Dietrich, A., 2014. The plant mitochondrial genome: dynamics and maintenance. Biochimie, 100, pp.107-120.

Hagemann, R., 2010. The foundation of extranuclear inheritance: plastid and mitochondrial genetics. *Molecular Genetics and Genomics*, *283*(3), pp.199-209.

Irish, V.F. and Sussex, I.M., 1992. A fate map of the Arabidopsis embryonic shoot apical meristem. Development, 115(3), pp.745-753.

Jajoo, R., Jung, Y., Huh, D., Viana, M.P., Rafelski, S.M., Springer, M. and Paulsson, J., 2016. Accurate concentration control of mitochondria and nucleoids. Science, 351(6269), pp.169-172.

Johnston, I.G., Gaal, B., Neves, R.P.D., Enver, T., Iborra, F.J. and Jones, N.S., 2012. Mitochondrial variability as a source of extrinsic cellular noise. PLoS computational biology, 8(3), p.e1002416.

Johnston, I.G., Burgstaller, J.P., Havlicek, V., Kolbe, T., Rülicke, T., Brem, G., Poulton, J. and Jones, N.S., 2015. Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. Elife, 4, p.e07464.

Johnston, I.G., 2019a. Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. Molecular plant, 12(6), pp.764-783.

Johnston, I.G., 2019b. Varied mechanisms and models for the varying mitochondrial bottleneck. Frontiers in Cell and Developmental Biology, 7, p.294.

Khakhlova, O. and Bock, R., 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. The Plant Journal, 46(1), pp.85-94.

Kimura, M., 1955. Solution of a process of random genetic drift with a continuous model. Proceedings of the National Academy of Sciences, 41(3), pp.144-150.

Kirk, P., Thorne, T. and Stumpf, M.P., 2013. Model selection in systems and synthetic biology. Current opinion in biotechnology, 24(4), pp.767-774.

Lanfear, R., 2018. Do plants have a segregated germline?. PLoS biology, 16(5), p.e2005439.

Logan, D.C., 2006. The mitochondrial compartment. Journal of experimental botany, 57(6), pp.1225-1243.

Lonsdale, D.M., Brears, T., Hodge, T.P., Melville, S.E. and Rottmann, W.H., 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 319(1193), pp.149-163.

Mandel, J.R., Ramsey, A.J., Holley, J.M., Scott, V.A., Mody, D. and Abbot, P., 2020. Disentangling complex inheritance patterns of plant organellar genomes: an example from carrot. *Journal of Heredity*, *111*(6), pp.531-538.

Maréchal, A. and Brisson, N., 2010. Recombination and the maintenance of plant organelle genome stability. New Phytologist, 186(2), pp.299-317.

McCauley, D.E., 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. New Phytologist, 200(4), pp.966-977.

Miller-Messmer, M., Kühn, K., Bichara, M., Le Ret, M., Imbault, P. and Gualberto, J.M., 2012. RecA-dependent DNA repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. *Plant Physiology*, *159*(1), pp.211-226.

Mohanta, T.K., Mishra, A.K., Khan, A., Hashem, A., Abd_Allah, E.F. and Al-Harrasi, A., 2020. Gene loss and evolution of the plastome. Genes, 11(10), p.1133.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L. and 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(13), pp.6960-6966.

Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K. and Börner, T., 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. The Plant Journal, 64(6), pp.948-959.

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rowan, B.A., Oldenburg, D.J. and Bendich, A.J., 2010. RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. Journal of experimental botany, 61(10), pp.2575-2588.

Scarcelli, N., Mariac, C., Couvreur, T.L.P., Faye, A., Richard, D., Sabot, F., Berthouly-Salazar, C. and Vigouroux, Y., 2016. Intra-individual polymorphism in chloroplasts from NGS data: Where does it come from and how to handle it?. Molecular ecology resources, 16(2), pp.434-445.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D. and Lohmann, J.U., 2005. A gene expression map of Arabidopsis thaliana development. Nature genetics, 37(5), pp.501-506.

Seguí-Simarro, J.M. and Staehelin, L.A., 2009. Mitochondrial reticulation in shoot apical meristem cells of Arabidopsis provides a mechanism for homogenization of mtDNA prior to gamete formation. *Plant signaling & behavior*, *4*(3), pp.168-171.

Shedge, V., Arrieta-Montiel, M., Christensen, A.C. and Mackenzie, S.A., 2007. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *The Plant Cell*, *19*(4), pp.1251-1264.

Stadler, T., Pybus, O.G. and Stumpf, M.P., 2021. Phylodynamics for cell biologists. *Science*, *371*(6526), p.eaah6266.

Stewart, J.B. and Chinnery, P.F., 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nature Reviews Genetics, 16(9), pp.530-542.

Virdi, K.S., Laurie, J.D., Xu, Y.Z., Yu, J., Shao, M.R., Sanchez, R., Kundariya, H., Wang, D., Riethoven, J.J.M., Wamboldt, Y. and Arrieta-Montiel, M.P., 2015. Arabidopsis MSH1 mutation alters the epigenome and produces heritable changes in plant growth. Nature communications, 6(1), pp.1-9.

Wai, T., Teoli, D. and Shoubridge, E.A., 2008. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. Nature genetics, 40(12), pp.1484-1488.

Wallace, D.C. and Chalkia, D., 2013. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. Cold Spring Harbor perspectives in biology, 5(11), p.a021220.

Wang, D.Y., Zhang, Q., Liu, Y., Lin, Z.F., Zhang, S.X. and Sun, M.X., 2010. The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. The Plant Cell, 22(7), pp.2402-2416.

Watson, J.M., Platzer, A., Kazda, A., Akimcheva, S., Valuchova, S., Nizhynska, V., Nordborg, M. and Riha, K., 2016. Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. Proceedings of the National Academy of Sciences, 113(43), pp.12226-12231.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,

Wickham, H., Bryan, J., 2022. _readxl: Read Excel Files_. R package version 1.4.0, <https://CRAN.R-project.org/package=readxl>.

Wickham, H., 2019. _stringr: Simple, Consistent Wrappers for Common String Operations_. R package version 1.4.0, <https://CRAN.R-project.org/package=stringr>.

Wilton, P.R., Zaidi, A., Makova, K. and Nielsen, R., 2018. A population phylogenetic view of mitochondrial heteroplasmy. Genetics, 208(3), pp.1261-1274.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J., 2007. An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. PloS one, 2(8), p.e718.

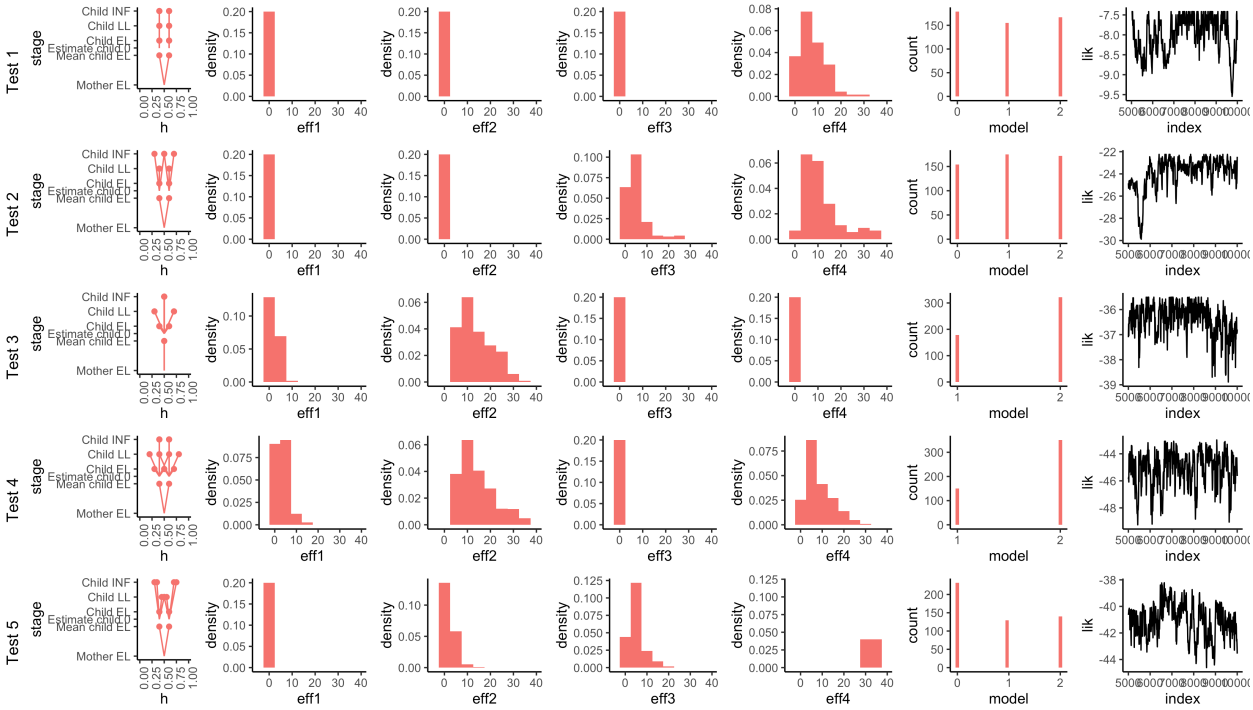
Woloszynska, M., 2010. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. Journal of experimental botany, 61(3), pp.657-671.

Wonnapinij, P., Chinnery, P.F. and Samuels, D.C., 2008. The distribution of mitochondrial DNA heteroplasmy due to random genetic drift. The American Journal of Human Genetics, 83(5), pp.582-593.

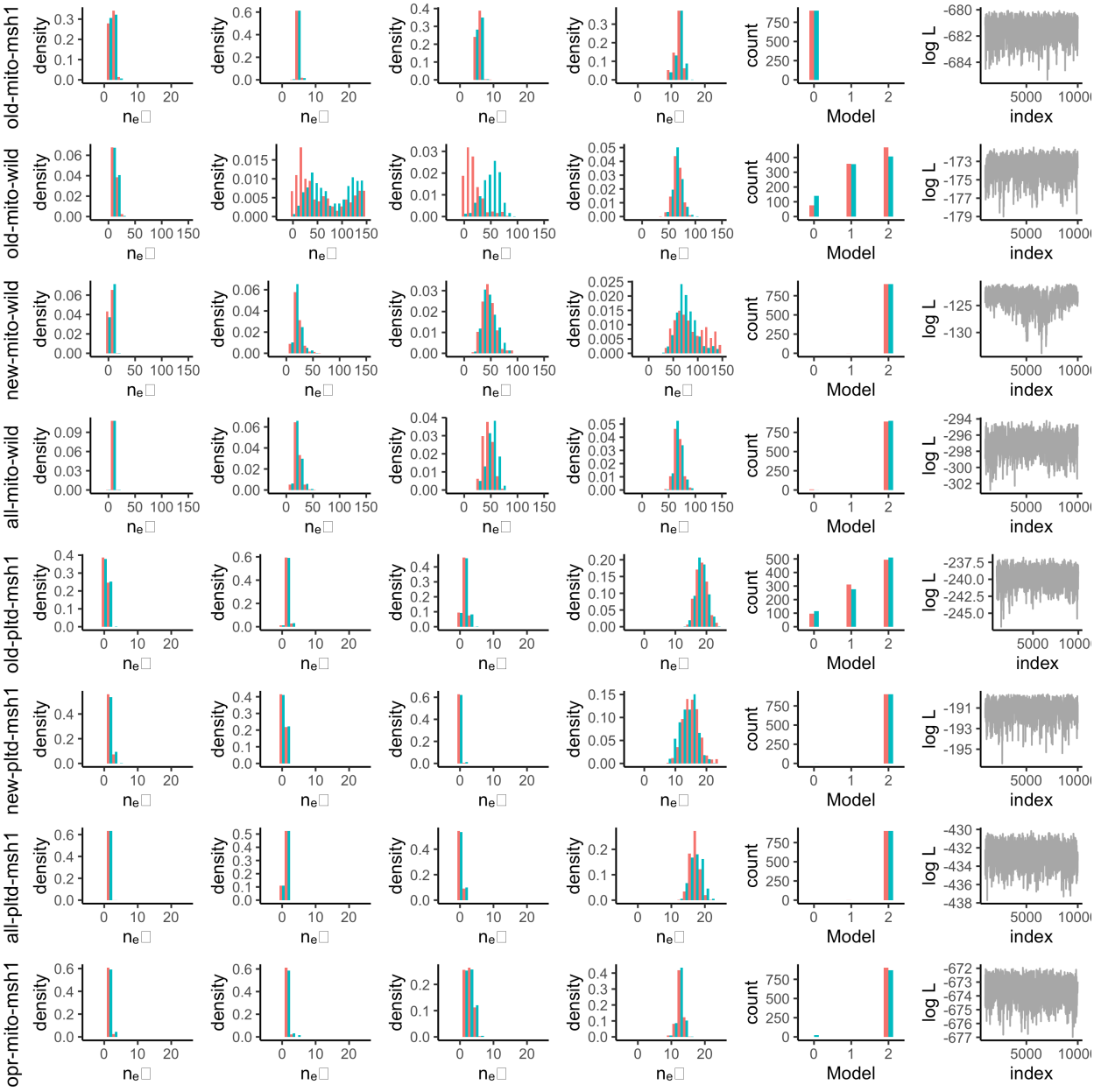
Wu, Z., Waneka, G., Broz, A.K., King, C.R. and Sloan, D.B., 2020. MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proceedings of the National Academy of Sciences, 117(28), pp.16448-16455.

Zhang, H., Burr, S.P. and Chinnery, P.F., 2018. The mitochondrial DNA genetic bottleneck: inheritance and beyond. Essays in Biochemistry, 62(3), pp.225-234.

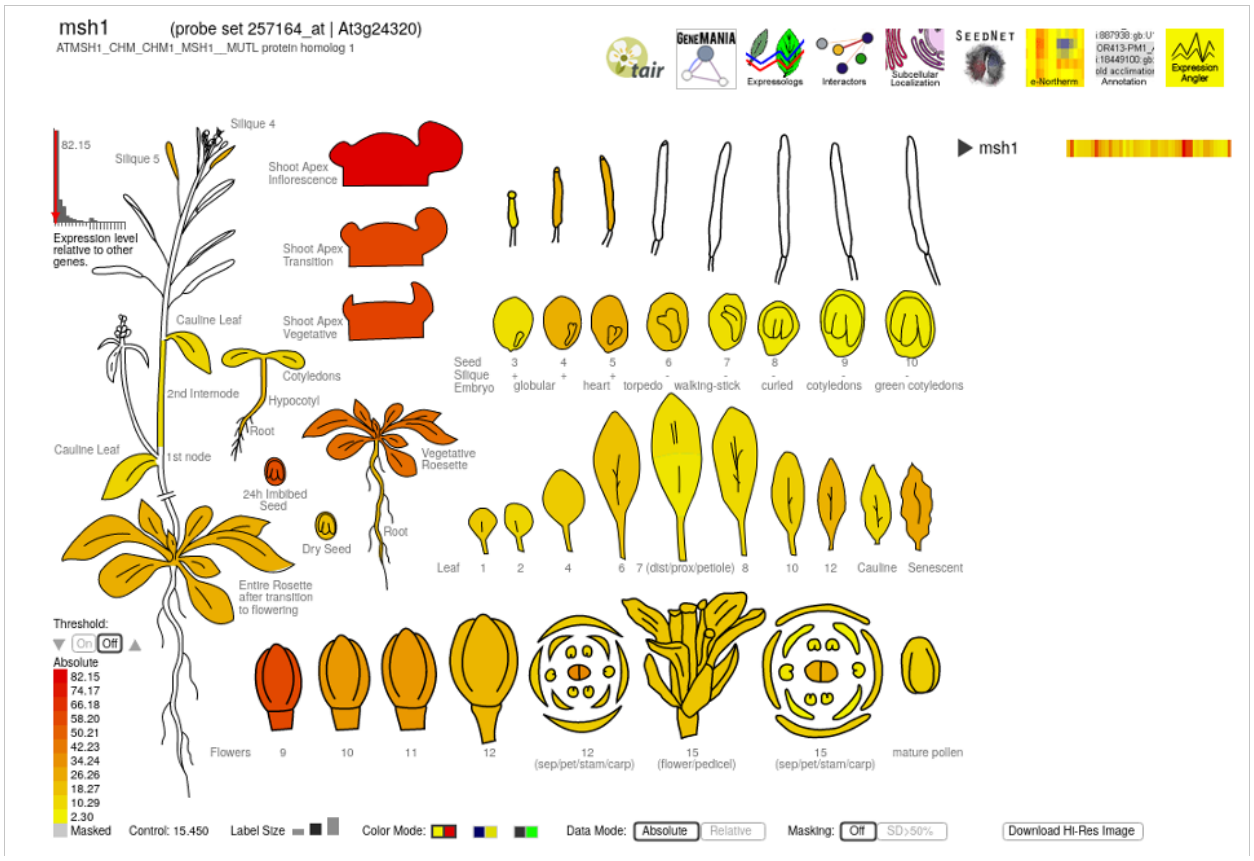
Supplementary Information



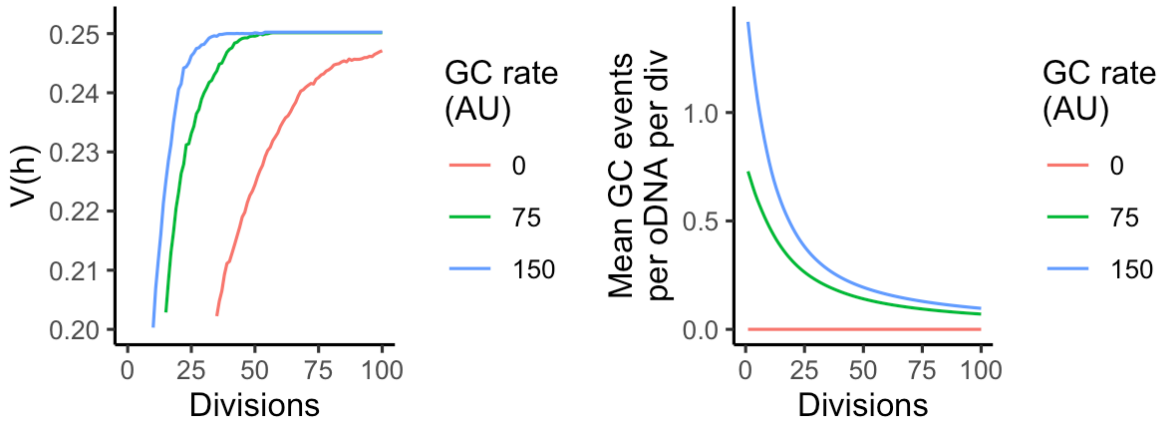
Supplementary Figure S1. **Validating model and inference approach.** Each row corresponds to a synthetic dataset generated to match a different type of segregation dynamics. The synthetic observations are shown in the first column, followed by the inferred effective segregating events to EL, LL, INF, and next-generation stages (eff1-4); the inferred model index (0, linear; 1, separate germline; 2, all separate); and finally a trace of likelihood over the MCMC chain as a readout of chain dynamics. Individual experiments reflect (1) segregation between generations, generating diversity between siblings but not within plants; (2) segregation in inflorescence development (and possibly between generations) but not in somatic tissue; (3) segregation only in somatic tissue, with a separate germline; (4) segregation between generations and in somatic tissue, but with germline protected; (5) segregation throughout linear germline, with precursor cells causing shifts in mean (see Methods). In case (1), segregation between generations but nowhere else is inferred, with uniform posteriors over model index in the absence of further information. In case (2), segregation at inflorescence development but not in somatic tissue is inferred, with a linear model favoured. In case (3), zero segregation in the germline and nonzero in somatic tissue is inferred, with models 1 and 2 (separate germline) inferred. Case (4) mirrors case (3) but with between-generation segregation also inferred. Case (5) supports the linear germline model as others cannot account for the shifts in mean heteroplasmy between stages.



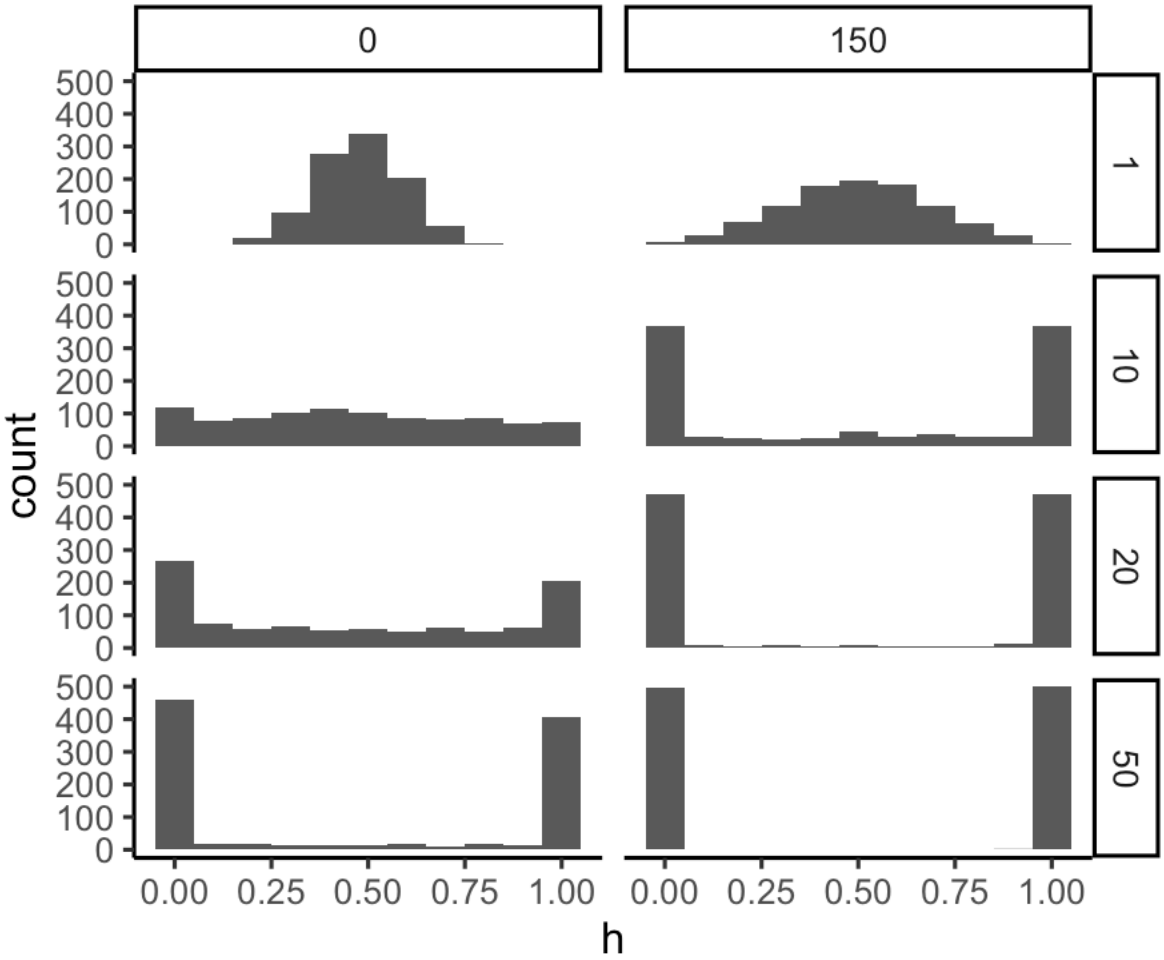
Supplementary Figure S2. **Inferred behaviour for different datasets.** Each row is the result of inference on the given dataset (*old*, existing data, *new*, new data, *all*, combined data, *opr*, outlier pruned as in main text; *mito*, MT, *pltd*, PT). Effective numbers of segregating events to EL, LL, inflorescence, and between-generation stages (eff1-4); the inferred model index (0, linear; 1, separate germline; 2, all separate); and finally a trace of likelihood over the MCMC chain as a readout of chain dynamics. Results for two independent MCMC chains (red and blue) are shown for all except the likelihood traces. Divergence in the “old-mito-wild” case reflects the unidentifiability of within-plant segregation parameters from this between-generational data.



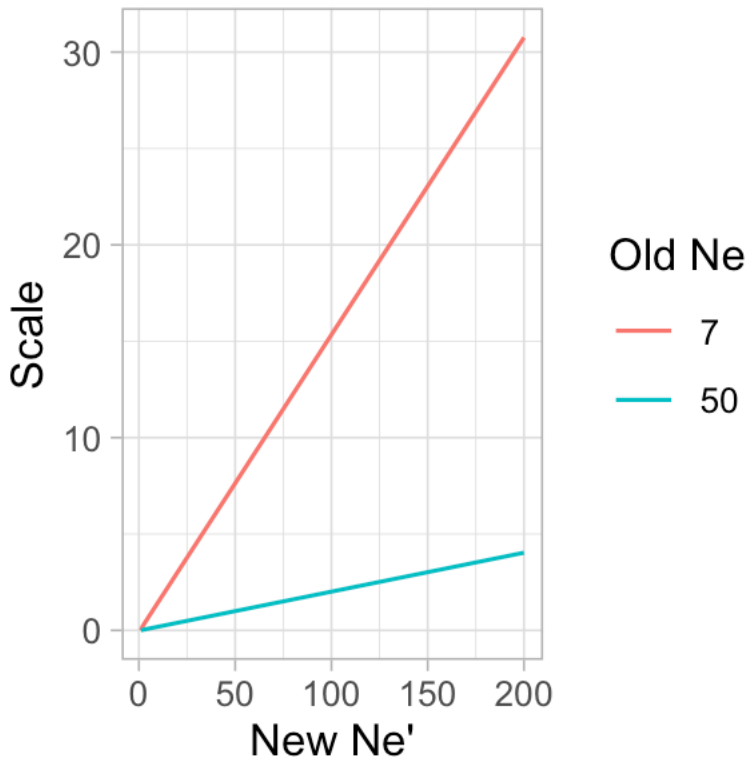
Supplementary Figure S3. **Msh1 expression patterns during development.** Data from Schmid et al. [2005], visualised by the "eFP browser" from the Bio-Analytic Resource for Plant Biology [Winter et al., 2007].



Supplementary Figure S4. **Simulated segregation with and without gene conversion.** (left) V(h) with number of divisions for different rates of gene conversion attempts (GC rate). (right) Actual gene conversion events per mtDNA per division, with number of divisions for different R. Within 34 divisions, the R = 75 and R = 150 cases readily generate the V(h) ~ 0.25 (corresponding to V'(h) ~ 1 for these simulations where h = 0.5) values observed for 75 divisions of the R=0 case, corresponding to a mean around 0.25 gene conversion events per mtDNA per cell cycle.



Supplementary Figure S5. **Predicted heteroplasmy distributions over cell divisions.** Example model predictions for heteroplasmy distributions in mtDNA populations of size N_e = 50, with a given number of cell divisions (rows). (left) No gene conversion, modelling the *msh1* case; (right) gene conversion at the rate suggested by our analysis in the wildtype plants.



Supplementary Figure S6. **Scaling factors for converting effective population sizes.** To interpret a number of inferred segregating events *n* from a population with N_e = 7 or 50 with a new population size N_e', read off the scale factor corresponding to the new population size on the horizontal axis and scale *n* by this factor. For most cases this scale factor is very close to N_e/N_e'.