

Stochastic organelle genome segregation through *Arabidopsis* development and reproduction

Amanda K. Broz¹ , Daniel B. Sloan¹  and Iain G. Johnston^{2,3} 

¹Department of Biology, Colorado State University, Fort Collins, CO 80523, USA; ²Department of Mathematics, University of Bergen, Bergen, 5007, Norway; ³Computational Biology Unit, University of Bergen, Bergen, 5007, Norway

Author for correspondence:
Iain G. Johnston
Email: iain.johnston@uib.no

Received: 29 April 2023
Accepted: 7 September 2023

New Phytologist (2024) **241**: 896–910
doi: 10.1111/nph.19288

Key words: *Arabidopsis*, bottleneck, chloroplasts, development, inheritance, mitochondria, organelle DNA, segregation.

Summary

- Organelle DNA (oDNA) in mitochondria and plastids is vital for plant (and eukaryotic) life. Selection against damaged oDNA is mediated in part by segregation – sorting different oDNA types into different cells in the germline. Plants segregate oDNA very rapidly, with oDNA recombination protein MSH1 a key driver of this segregation, but we have limited knowledge of the dynamics of this segregation within plants and between generations. Here, we reveal how oDNA evolves through *Arabidopsis thaliana* development and reproduction.
- We combine stochastic modelling, Bayesian inference, and model selection with new and existing tissue-specific oDNA measurements from heteroplasmic *Arabidopsis* plant lines through development and between generations.
- Segregation proceeds gradually but continually during plant development, with a more rapid increase between inflorescence formation and the next generation. When MSH1 is compromised, the majority of observed segregation can be achieved through partitioning at cell divisions. When MSH1 is functional, mtDNA segregation is far more rapid; we show that increased oDNA gene conversion is a plausible mechanism quantitatively explaining this acceleration.
- These findings reveal the quantitative, time-dependent details of oDNA segregation in *Arabidopsis*. 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; Clegg *et al.*, 1994; Palmer *et al.*, 2000; Allen & Martin, 2016; Mohanta *et al.*, 2020; Giannakis *et al.*, 2022). Plant cells typically contain populations that range from dozens to thousands of mtDNA and ptDNA molecules (Preuten *et al.*, 2010; Wang *et al.*, 2010; Greiner *et al.*, 2020; Fernandes Gyorffy *et al.*, 2021), contained within their respective organelles (Barr *et al.*, 2005; Woloszynska, 2010; McCauley, 2013; 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 (Burgstaller *et al.*, 2018; Stewart & Chinnery, 2015; Zhang

et al., 2018; Johnston, 2019b). 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.

How plants limit the inheritance of these damaging mutations is less well understood (Barr *et al.*, 2005; Woloszynska, 2010; Galtier, 2011; McCauley, 2013). 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. Previous work characterising inheritance and sorting of heteroplasmy in carrot (Mandel *et al.*, 2020) described rather little evidence for mitochondrial segregation during plant development, with most observations involving a loss of heteroplasmy between generations. Such intergenerational sorting has also been observed in *Silene*, where only 17% of offspring retained heteroplasmy that was present in their mother (Bentley *et al.*, 2010). The heteroplasmy levels involved in the carrot study were typically extreme (*c.* 1% frequency of the minor allele), meaning that within-plant segregation would be very hard to detect. However, one notable

instance was recorded of a moderately heteroplasmic offspring (31% minor allele frequency) arising from a < 1% heteroplasmic mother and father, suggesting that a mechanism for substantial amplification of minor alleles may nonetheless be present. Barnard-Kubow *et al.* (2017) reported substantial vegetative sorting of ptDNA in *Campanulastrum americanum*, acting to resolve heteroplasmy arising from the plant's biparental plastid inheritance. More general, qualitative observations of variegated phenotypes also suggest that vegetative sorting (within-plant segregation during development) must be possible (Greiner *et al.*, 2015).

Recent experimental evidence has shown that generational sorting out of plant mtDNA and ptDNA is extremely rapid compared with animals (Broz *et al.*, 2022). This work showed that this sorting depends on *MSH1*, a gene responsible for controlling recombination activity in oDNA (Abdelnoor *et al.*, 2003). Although the precise nature and mechanism of this control is yet to be determined (Arrieta-Montiel *et al.*, 2009; Christensen, 2014; Virdi *et al.*, 2015), *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 (Day & Madesis, 2007; Shedje *et al.*, 2007; Maréchal & Brisson, 2010; Rowan *et al.*, 2010; 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 (Preuten *et al.*, 2010; Wang *et al.*, 2010; Greiner *et al.*, 2020; Edwards *et al.*, 2021), 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 (Burian *et al.*, 2016; Lanfear, 2018), 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 (Kirk *et al.*, 2013; Lanfear, 2018).

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 reveal 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.

Materials and Methods

Plant material and growth

The initial generation and selection of heteroplasmic *Arabidopsis thaliana* L. 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 wild-type 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 (to be described later); both have been reported to be full allelic knock-outs (Broz *et al.*, 2022). All progeny were confirmed to be heterozygous for *MSH1*. Seeds of desired lines were vernalised in water at 4°C for 3 d, sown in 3-inch pots containing Pro-Mix BX media, and grown under short day conditions (10 h : 14 h, light : dark) on light racks with fluorescent bulbs ($c. 150 \mu\text{E m}^{-2} \text{s}^{-1}$) at ambient temperature ($c. 25^\circ\text{C}$). An initial fully expanded rosette leaf sample was taken at 4 wk of growth to identify heteroplasmic individuals. Three additional leaves were sampled at 5 wk of growth. These 4–5-wk-old leaf samples are considered 'early leaf' (EL) for subsequent analyses. At 8 wk, four additional leaf samples were taken. Two were harvested from the base of the rosette. These leaves were already fully expanded at 5 wk 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 wk from the top of the rosette, emerging from the SAM at a later timepoint than ELs, and are considered as late leaf 'LL' in the analysis. Inflorescence tissue (INF) was harvested after plants began to bolt. The number of families sampled, and the range of samples in each family were in the original study (Broz *et al.*, 2022): mtDNA *msh1* (five families, 23–57 samples per family); ptDNA *msh1* (seven families, 11–47 samples per family); mtDNA WT (six families, 12–38 samples per family); in this study, ptDNA *msh1* (one family, 64 samples in family); mtDNA WT (three families, 16–37 samples per family). Samples were of two main types: mother-offspring, where EL measurements were taken in a mother and across its offspring, and within offspring, where EL, LL, and INF samples were taken within multiple offspring in a family.

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.*, 2020) and ddPCR assays were designed to track these SNVs within plants and between generations. Allele-specific primers and probes were designed for each SNV and droplet generation, and reading was performed using Bio-Rad QX200 system. This study used the specific loci plastid 26 553, mitochondria 91 017, and mitochondria 334 038, 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 numts have enough sequence divergence that nuclear and plastid copies can be unambiguously distinguished.

Developmental history models

We 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 the “linear” model of Fig. 1(a) (Wilton *et al.*, 2018; Stadler *et al.*, 2021). An example of a possible relationship between heteroplasmy distributions in this case is illustrated in Fig. 1(b).

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 the “separate germline” model of Fig. 1(a), 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 the “all separate” model of Fig. 1(a), 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 time points – 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

Our statistical approach requires a ‘likelihood’ function, giving the probability of making our experimental observations given a

particular model and parameterisation (mathematical terms here and throughout are described in Table 1). To assign a likelihood to our tissue observations given a developmental model, we need to (1) estimate the ancestral cell heteroplasmies and (2) estimate the probability of observing a tissue heteroplasmy given the ancestral value and some parameterised description of segregation (Burgstaller *et al.*, 2014, 2018). For (1), we treat ancestral cell heteroplasmies as latent (unobserved) variables and integrate the likelihood over all possible values for each. For (2), we use the Kimura distribution (Kimura, 1955; Wonnapijit *et al.*, 2008) to describe the probability of observing a given heteroplasmy in individual tissue samples, creating a stochastic model with a full likelihood function (Broz *et al.*, 2022; Giannakis *et al.*, 2022). We change variables from the ‘drift parameter’ b to a number of ‘effective segregation events’ $n = \log b/(1-1/N_e)$ – described in detail 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; described 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 0–200 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 (to be described later).

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 $\{D_i\} = \{D_{i,MEL}, D_{i,EL}, D_{i,LL}, D_{i,INF}\}$ for the set of observations in family i , with elements respectively corresponding to Mother Early leaf (MEL), Child Early leaf (EL), Child Late leaf (LL), and Child Inflorescence (INF). The underlines mean that we are in general talking about vectors of observations. We write S_{C_j} for the unobserved (latent) variable associated with ancestral cell heteroplasmy at developmental stage j . The likelihood associated with a set of measurements is built up iteratively, considering all possible values of unobserved variables, and how subsequent behaviour might vary thereafter. The basic “building block” of these expressions is $P(b|h_0; n)$, which is the probability of seeing heteroplasmy b given an initial heteroplasmy of h_0 and n effective segregation events. These probabilities are described by the Kimura distribution (Kimura, 1955; Wonnapijit *et al.*, 2008; Table 1). For example, imagine we had initial heteroplasmy h_0 , then n_0 effective segregation events before an unobserved developmental state S_{C_1} , after which n_1 effective segregation events generate the population from which a sample D is taken. We could write $\int dS_{C_1} P(S_{C_1}|h_0; n_0) P(D|S_{C_1}; n_1)$, which considers, for all possible values of S_{C_1} ($\int dS_{C_1}$ means we

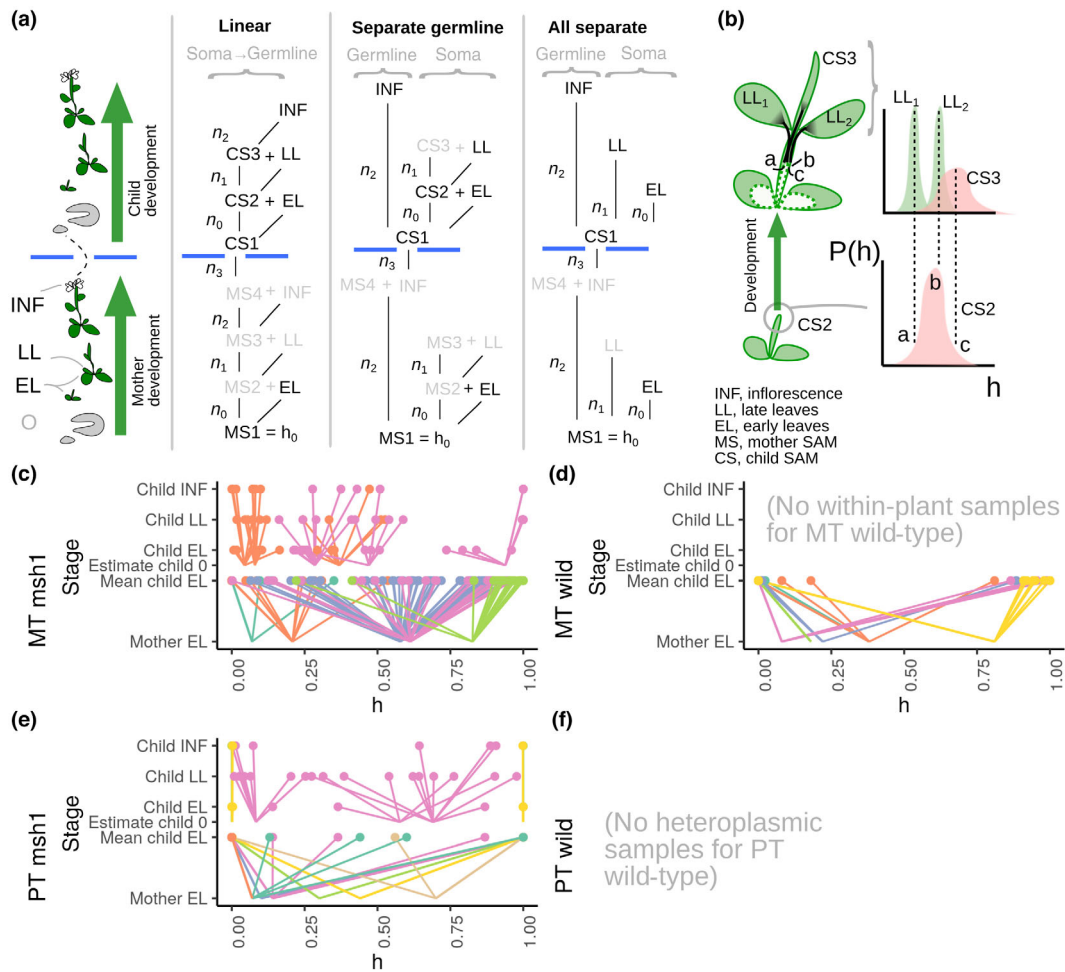


Fig. 1 Models and data for heteroplasmy segregation in plant development. (a) Developmental models for observations of heteroplasmy (h , the proportion of mutant organelle DNA type in a sample) in *Arabidopsis thaliana*. MS_i and CS_i are the unobserved (latent) ancestral cells at different developmental stages (O, original precursor state; EL, early leaf; LL, late leaf; INF, inflorescence) in Mother and Child shoot apical meristem (SAM). The blue horizontal bars denote the generation of sex cells and the establishment of a new generation. Greyed-out elements are unidentifiable given our observations and play no role in our model. n_i corresponds to the number of effective segregation events (model cell divisions) at each developmental stage. (b) Example of model for heteroplasmy h 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 (LL₁ and LL₂) 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 LL₁ and LL₂ are recorded) or unseen (the CS₃ distribution plays a latent role in our model). (c–f) Observed heteroplasmy data through development in different heteroplasmic plant families from Broz *et al.* (2022): (c) mtDNA in mutant *msh1* background; (d) mtDNA in wild-type background (no within-plant data was taken here in Broz *et al.*, 2022); (e) ptDNA in mutant *msh1* background; (f) no heteroplasmic samples for wild-type ptDNA available in Broz *et al.* (2022). Samples are taken between generations (lower stages; comparing mother EL to a range of offspring EL) and within offspring (upper stages; measuring EL–LL–INF within offspring); within-offspring measurements were not taken for wild-type mtDNA in Broz *et al.* (2022). Different colours correspond to different families (with different founder mothers). The ‘fanning out’ of individual sample heteroplasmy over time corresponds to increasing sample-to-sample variance.

integrate over all these values), the probability of an S_{C1} value given the initial heteroplasmy and n_0 events ($P(S_{C1}|h_0; n_0)$) and the probability of our observation D given that S_{C1} value and n_1 further events ($P(D|S_{C1}; n_1)$). The actual expressions are longer

(and depend on the other n parameters of our model) but consist of building blocks like these, combined according to the developmental models in Fig. 1. For the model without a segregated germline, the likelihood is then

$$L(\{\underline{D}_i\}|\underline{z}, h_0) = P(\underline{D}_{i,MEL}|h_0; n_0) \int dS_{C1} P(S_{C1}|h_0; n_0 + n_1 + n_2 + n_3) P(\underline{D}_{i,EL}|S_{C1}; n_0) \\ \int dS_{C2} P(S_{C2}|S_{C1}; n_0) P(\underline{D}_{i,LL}|S_{C2}; n_1) \int dS_{C3} P(S_{C3}|S_{C2}; n_1) P(\underline{D}_{i,INF}|S_{C3}; n_2)$$

Eqn 1

So that S_{C1} is the precursor to EL and S_{C2} , S_{C2} is the precursor to LL and S_{C3} , and S_{C3} is the precursor to INF (Fig. 1a). With a segregated germline, the corresponding expression is

$$L(\{\underline{D}_i\}|\underline{z}, h_{0i}) = P(\underline{D}_{i,MEL}|h_{0i}; n_0) \int dS_{C1} P(S_{C1}|h_{0i}; n_2 + n_3) P(\underline{D}_{i,EL}|S_{C1}; n_0) \\ P(\underline{D}_{i,INF}|S_{C1}; n_2) \int dS_{C2} P(S_{C2}|S_{C1}; n_0) P(\underline{D}_{i,LL}|S_{C2}; n_1) \quad \text{Eqn 2}$$

So that S_{C1} is the precursor to EL, INF, and S_{C2} , and S_{C2} is the precursor to LL. With completely separate developmental lineages, we have

$$L(\{\underline{D}_i\}|\underline{z}, h_{0i}) = P(\underline{D}_{i,MEL}|h_{0i}; n_0) \int dS_{C1} P(S_{C1}|h_{0i}; n_2 + n_3) P(\underline{D}_{i,EL}|S_{C1}; n_0) P(\underline{D}_{i,LL}|S_{C1}; n_1) P(\underline{D}_{i,INF}|S_{C1}; n_2) \quad \text{Eqn 3}$$

So that S_{C1} is the precursor to all lineages, which develop independently. These likelihood functions are the mathematical analogues to the graphical illustration of developmental models in Fig. 1.

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 (Supporting Information 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 $V(b)$) through development – hence $EL \leq LL \leq INF \leq$ 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 D is the product of likelihoods across the different independent families in our dataset:

$$L(D|\underline{z}; h_0) = \prod_{\text{families } i} L(\{\underline{D}_i\}|\underline{z}; h_{0i}) \quad \text{Eqn 4}$$

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 *c.* 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 to 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), and 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 the Discussion section).

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 (Wonnapijit *et al.*, 2008; Giannakis *et al.*, 2023), $b = \exp(-n/N_e) \approx (1 - 1/N_e)^n$. 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 : as $(1 - 1/N_e)^n = (1 - 1/N'_e)^{n'}$, $n' = n \log(1 - 1/N_e) / \log(1 - 1/N'_e)$, 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 (Fig. S2). 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 $N_b = 1 / (1 - (1 - 1/N_e)^n)$.

Reversible jump MCMC

We use reversible jump MCMC (RJMCMC) to identify the statistical support for different models of developmental histories

given observations (Green, 1995; Dellaportas *et al.*, 2002; Kirk *et al.*, 2013). Briefly, RJMCMC simultaneously explores different parameterisations of models and different model structures, assigning support to model-parameter combinations that are most compatible with the data. To search the joint model-parameter space, a way of relating parameters between models must be specified. 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 $g_{ij}(z^{(i)})$ 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, corresponding to no prior favouring of one model structure over another, and no prior favouring of particular parameter values. This makes 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 one-third, employing the above perturbation to parameters when this option is not chosen. MCMC chains were run over 10^7 samples, discarding 10^4 as burn-in and subsequently recording every 10^{th} 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 $\text{WT} + \text{MU} \rightarrow \text{WT} + \text{WT}$ and $\text{WT} + \text{MU} \rightarrow \text{MU} + \text{MU}$ (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 $\kappa = 0.007$ per cell division gives *c.* 4 events per cell division or *c.* $4/50 = 0.08$ 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 reamplification, 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 *c.* 34 cell divisions (Fig. S3).

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

Table 1 Mathematical terms used in the manuscript.

Mathematical/statistical model	A quantitative description of how observations may be generated by a biological system. Statistical models describe a distribution from which observations can be 'drawn'; the shape of these distributions will be influenced by the <i>parameters</i> of the model.
Parameters	A set of values in a mathematical model that describes its behaviour. These often represent the rates of biological processes, or the magnitudes of their effects on the system.
Heteroplasmy	Written throughout the text as h . The proportion of a reference oDNA type – usually a mutant type – in a population, which may be within a cell or in a bulk sample.
Kimura distribution	A statistical model, giving a theoretical distribution used to describe how the cell-to-cell distribution of heteroplasmy values spreads out as segregation occurs. Takes parameters describing the mean heteroplasmy and the amount of segregation.
Reversible jump MCMC	A computational approach that explores different model structures and their possible parameters. It returns probability distributions (called 'posterior' distributions) describing which model structures, and which parameter values, are most supported by a given set of observations.
Bottleneck size/ (normalised) heteroplasmy variance	$V(h)$ is cell-to-cell variance of heteroplasmy values h . This is often 'normalised' by dividing by a function of mean heteroplasmy: $V'(h) = V(h)/(h(1-h))$. This accounts for the fact that the scale of $V(h)$ can depend on mean heteroplasmy. Bottleneck size is the number of oDNAs that, when binomially sampled, would generate a given amount of variance: $N_b = 1/V'(h)$.
Effective segregation event	Our study's representation of a developmental event generating cell-to-cell heteroplasmy variance. We picture a population of N_e oDNA molecules in a cell. Each is randomly assigned to one of two daughter cells. Each daughter then randomly reamplifies their oDNA population back up to N_e . The combination of partitioning and reamplification constitutes an effective segregation event (and is a model for cell dynamics during development).

quantitative model predicting the distribution of heteroplasmy at the different developmental and generational time points we observe (Johnston *et al.*, 2015; Burian *et al.*, 2016; Burgstaller *et al.*, 2018; Wilton *et al.*, 2018). Throughout this work, heteroplasmy is defined as the proportion of a mutant oDNA type in a sample: so a sample of 60% mutant and 40% wild-type oDNA has heteroplasmy $h = 0.6$; this and other quantitative terms are listed as a glossary in Table 1. We analysed 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 (Furner & Pumfrey, 1992; Irish & Sussex, 1992; Burian *et al.*, 2016). 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. This is compatible with evidence in this system, which found only weak bias for some alleles (Broz *et al.*, 2022) and evidence from other systems (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 (Fig. 1a,b). 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 an ‘ontogenetic phylogeny’ picture, which tracks 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 apply it to a description of a continuous heteroplasmy level as it varies through plant development. Our mathematical model describes and links the distributions of heteroplasmy in the estimated stem cell populations through and between generations (Fig. 1a,b; see the **Materials and Methods** section). We consider three different models of development, 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 (rather than fitting using coarse-grained, uncertain summary statistics like heteroplasmy variance), increasing the statistical power of our approach (Giannakis *et al.*, 2023).

The amount of segregation occurring between each developmental period is quantified in our model as ‘effective segregation events’. This draws upon a picture of binomial cell divisions, where a cell has an effective population size of N_e oDNA molecules. At a cell division, each oDNA is randomly assigned to one of the two daughter cells. The daughter cells’ oDNA populations are then expanded back to N_e through ‘relaxed’ replication, where oDNA molecules randomly replicate (Chinnery & Samuels, 1999). Because of the random partitioning and replication, each division generates some variance in heteroplasmy between daughter cells: the number n of such cell divisions that would generate the observed variance in heteroplasmy is our number of ‘effective segregation events’. We use this variable rather than a ‘bottleneck size’ or ‘drift parameter’ (Wonnapijit *et al.*, 2008; Johnston, 2019b) because (1) it corresponds to a biological ‘null model’ where variance is generated by cell divisions alone (to be described later); and (2) 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 the **Materials and Methods** section), that cellular oDNA population size $N_e = 50$ for mtDNA (Preuten *et al.*, 2010; Wang *et al.*, 2010) and 7 for ptDNA (the latter corresponding to seven genetically homogeneous organelles (Scarcelli *et al.*, 2016; Greiner *et al.*, 2020)). We adopt binomial cell divisions and reamplification as a convenient null model with some empirical support (Johnston *et al.*, 2012, 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 basal rosette leaves between 4 and 8 wk of growth), late-emerging leaves (LL, upper rosette leaves that were fully expanded after 8 wk of growth), and inflorescences (INF; Fig. 1a; see the **Materials and Methods** section), reflecting tissues generated progressively later in development from the SAM. These data include observations of mtDNA heteroplasmy in wild-type and *msb1* mutant backgrounds, and ptDNA heteroplasmy in the *msb1* mutant. All wild-type lines measured were homoplasmic in ptDNA, likely due to the high rate of plastid segregation in wild-type plants (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), a computational method which simultaneously estimates which of a set of different mathematical models is most supported by data, and the parameters of those models (Green, 1995; Dellaportas *et al.*, 2002). RJMCMC gives ‘posterior’ probability distributions for each parameter and model index, describing the probability of different mechanisms given the data and any prior information (see the **Materials and Methods** section; 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 (Fig. S1), and confirmed that inference results were numerically stable (Fig. S4). 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).

Fig. 2 shows the inferred number of effective segregation events at different stages of plant development and between generations, across the different model structures in Fig. 1(a). 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 \rightarrow O) is substantially greater than that occurring within a single plant up to the inflorescence stage (O \rightarrow INF). In the *msb1* mutant, a total of between 9 and 15 events are inferred to occur for mtDNA and between 15 and 25 for ptDNA between generations. In the wild-type, between 50 and 100 events – on average about a sevenfold increase in segregation – are inferred to occur between generations for mtDNA.

To compare with other studies, we can consider the ‘normalised heteroplasmy variance’ $V(h)$, which is the cell-to-cell

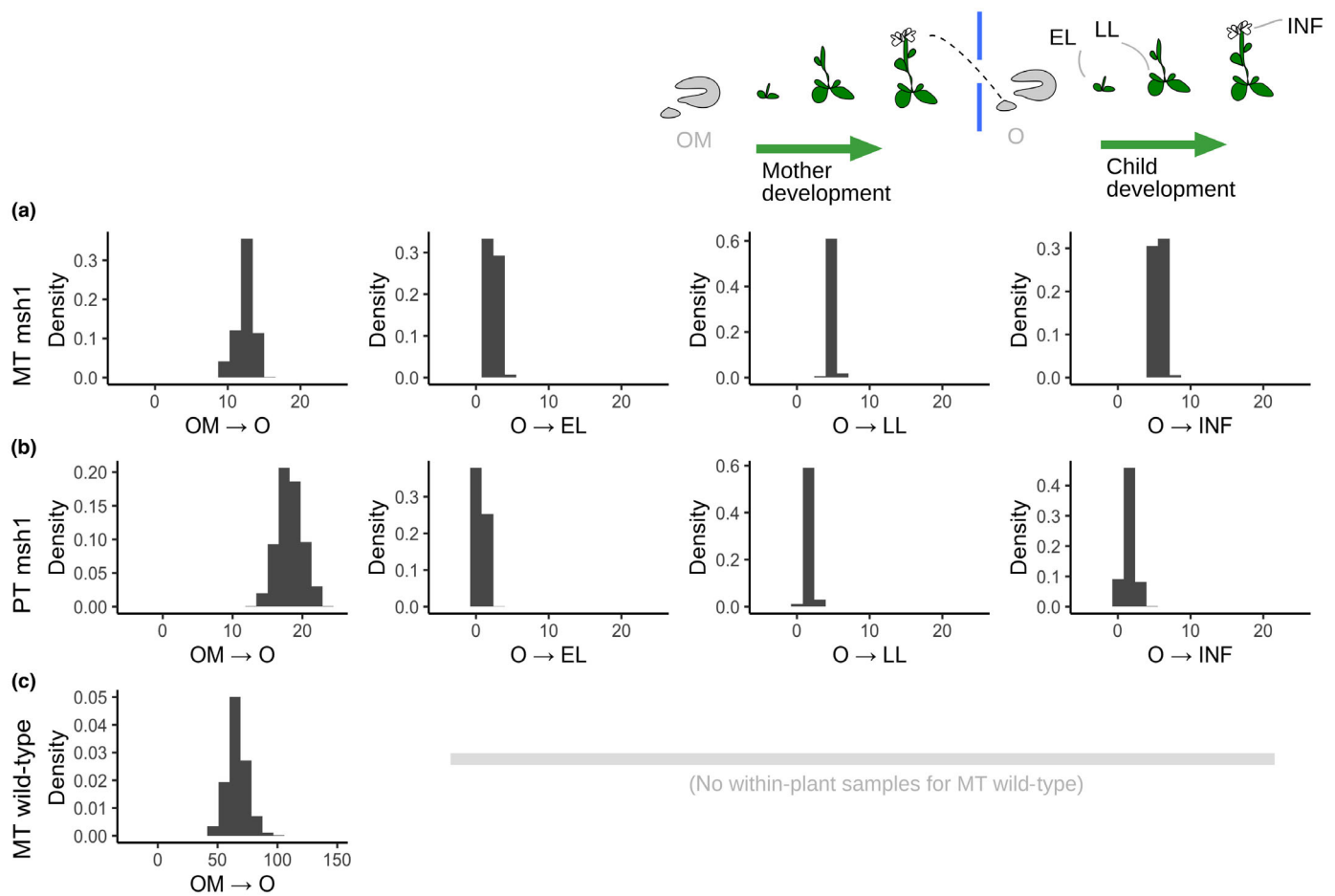


Fig. 2 Posteriors from inference process. Posterior distributions, inferred across models, for the effective segregation events from a precursor state (O for child, OM for mother) to different tissue precursors (EL, early leaf; LL, late leaf; INF, inflorescence), and between generations (OM → O) in *Arabidopsis thaliana*. (a) *msh1* mtDNA ($N_e = 50$); (b) *msh1* ptDNA ($N_e = 7$); (c) wild-type mtDNA ($N_e = 50$, different scale); no within-plant data was taken here in Broz *et al.* (2022), so the details of vegetative segregation cannot be inferred.

variance in heteroplasmy normalised by $h(1 - h)$, where h is the cell-to-cell mean heteroplasmy. The ‘bottleneck size’ N_b , the effective population size if all heteroplasmy variance was generated by a single binomial sample, is $1/V(h)$. Our inferred numbers of effective segregation events 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 wild-type mtDNA; hence, ‘bottleneck sizes’ of $c. 4$ for *msh1* mtDNA, $c. 1$ for *msh1* ptDNA, and $c. 1$ for wild-type mtDNA. 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.

Segregation differences in samples within a generation were less pronounced, with comparatively few effective segregation events inferred to occur up to the generation of early leaves (fully expanded at 4–5 wk of growth), and few more inferred to occur up to late leaf generation (fully expanded at 8 wk of growth). The means of each estimated parameter 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 wild-type mtDNA, and *msh1* ptDNA sampling was also somewhat limited. Based on the seven-fold scaling of mtDNA segregation from the *msh1* mutant to the wild-type, 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

To further illuminate the developmental dynamics of *Arabidopsis* heteroplasmy, we measured mitochondrial heteroplasmy across developmental profiles in lines where MSH1 functionality was recovered by backcrossing to a wild-type male, while preserving the heteroplasmy that was present in the female. The heteroplasmy dynamics in these lines are expected to reflect those in the

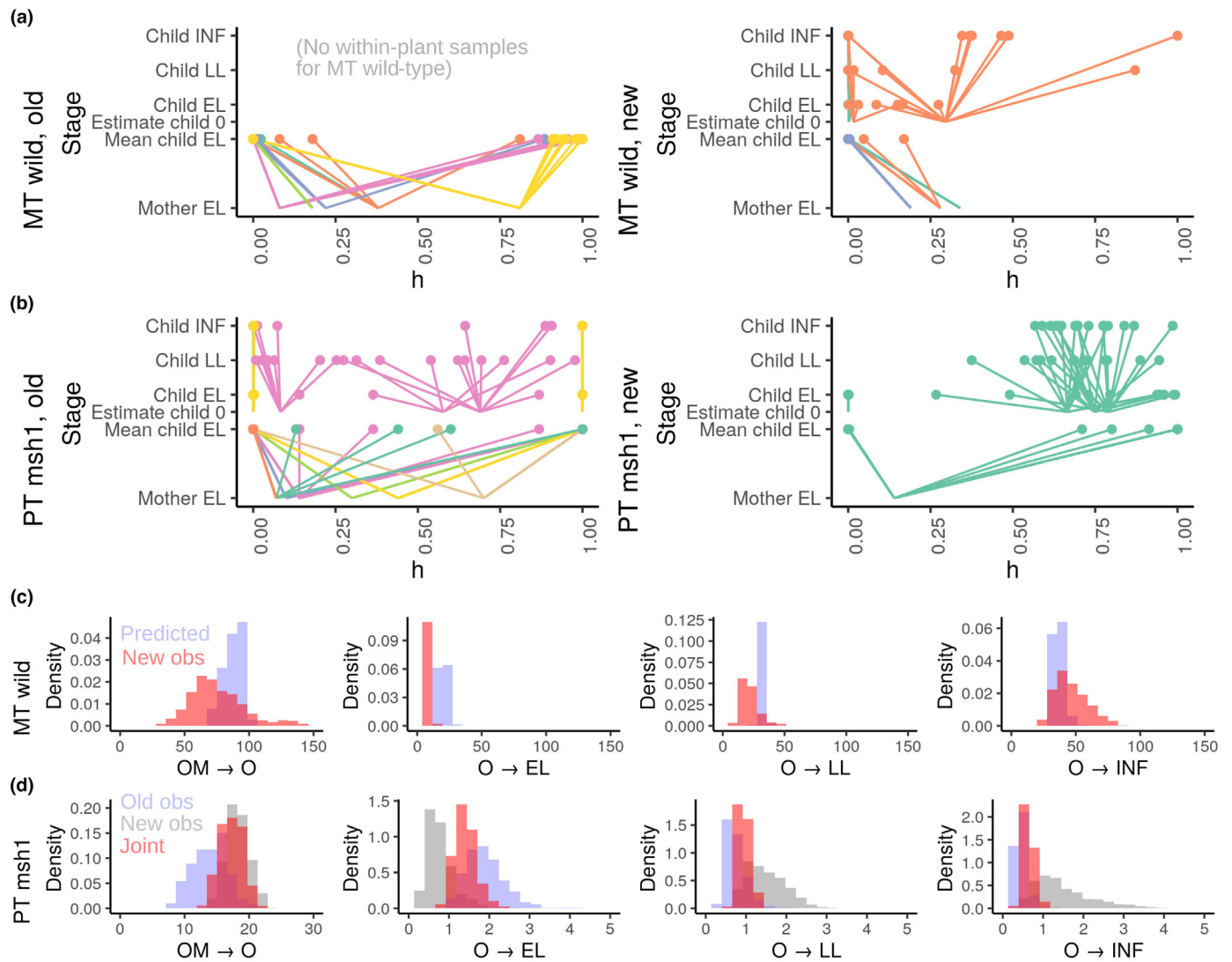


Fig. 3 New data and predicted segregation behaviour. (a, b) Previous ('old', left) and new (right) oDNA observations for (a) wild-type mtDNA and (b) *msh1* ptDNA in *Arabidopsis thaliana*. Data are displayed as heteroplasmy levels h measured at different developmental stages. Different colours correspond to different families (with different founder mothers). (c) Within-plant segregation dynamics for wild-type mtDNA, plotted as the probability of a given number of effective segregation events between different developmental stages. Predictions (blue) from scaling the *msh1* observations sevenfold to match between-generation observations; (red) inferred effective segregation events from new data. (d) Segregation dynamics of *msh1* ptDNA; previous observations (blue); new observations (grey); and refined estimates inferred from the joint dataset (red). Developmental stages: O, original precursor state; EL, early leaf; LL, late leaf; INF, inflorescence.

wild-type (where heteroplasmy rarely arises because of low mutation rates and the rapid sorting). The new observations are shown in the right panels of Fig. 3(a,b).

Matching our predictions, we found dramatically accelerated mtDNA segregation in the wild-type at the late leaf and inflorescence stages, with the rates inferred from new observations compatible with the seven-fold scaling predicted from the between-generations data (Fig. 3c). However, the extent of wild-type mtDNA segregation before early-leaf development was lower than this hypothesis predicted (with only 0.2% posterior probability shared between the two distributions) – and more similar to the lower levels in the *msh1* mutant. This difference suggests a refinement to our predicted picture – that the increased segregation activity of MSH1 is mainly manifest in later

development, which in turn is in qualitative agreement with observed patterns of MSH1 expression (Fig. S5).

Our new ptDNA observations also matched the predictions inferred from previous data, with the increased volume of observations substantially refining the estimates of effective segregation events at different developmental stages (Fig. 3d). The new observations were always compatible with the (more uncertain) estimated parameters from the original measurements and combined to provide a tightly defined estimate of segregation dynamics through development. Assuming, as before, an effective population size $N_e = 7$, the number of effective segregation events is quite limited from early leaf to late leaf to inflorescence, with an over 10-fold further increase in segregation following between generations. It seems likely that this dramatic segregation

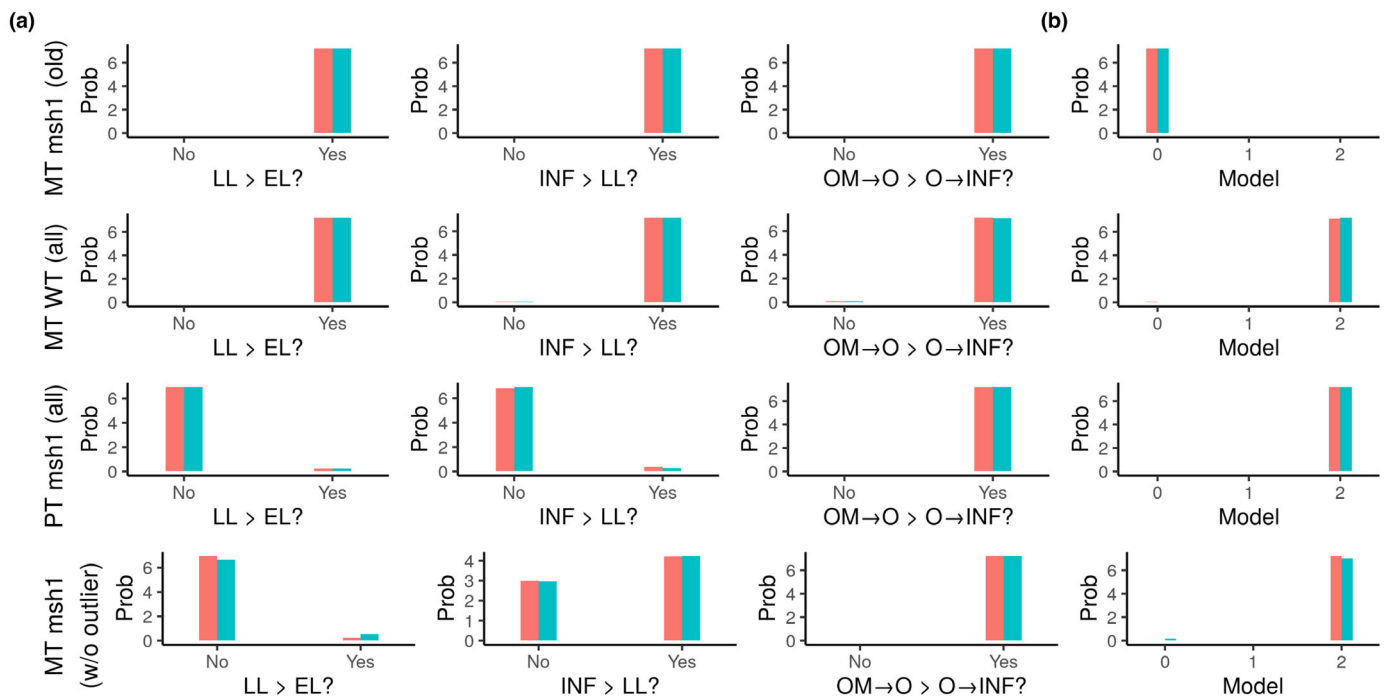


Fig. 4 Patterns and models of segregation through development inferred from combined heteroplasmy profiles. (a) Evidence for progressive vegetative segregation through development in *Arabidopsis thaliana*. Each plot asks whether the extent of segregation over one period is greater than that over another. LL > EL corresponds to late-leaf segregation exceeding early-leaf segregation; INF > LL corresponds to inflorescence segregation exceeding late-leaf segregation; OM → O > O → INF corresponds to whether between-generation segregation (early mother to early offspring) exceeds vegetative segregation (early offspring to inflorescence). The probability for yes/no answers to these questions is given, with two independent computational estimates plotted to demonstrate numerical convergence. (b) Probabilities of different model structures 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 Supporting Information Fig. S4). Developmental stages: OM, original precursor state of mother plant; O, original precursor state; EL, early leaf; LL, late leaf; INF, inflorescence.

between generations is due to a severe physical bottleneck on ptDNA, perhaps involving the inheritance of approximately only one homoplasmic organelle (see the Discussion section).

To ask whether within-generation segregation was a genuinely continuous process, we next explored the probability that the magnitude of segregation increased sequentially through developmental stages (e.g. whether the amount of segregation experienced by late leaves exceeded that experienced by early leaves). Here, we found evidence for continuous mitochondrial segregation through development when functional MSH1 is present, but more limited support when MSH1 was compromised (Fig. 4a). When MSH1 is compromised, segregation patterns can be explained by all segregation occurring in early development before early-leaf sampling; with functional MSH1, mitochondrial segregation proceeds continuously through development. Given our limited dataset, it remains open to what extent MSH1 influences within-generation segregation in plastids.

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

Arabidopsis has been estimated to undergo *c.* 34 germline cell divisions between generations (Watson *et al.*, 2016). In the *msh1*

mutant, the number of inferred effective segregation events (averages *c.* 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 $N_c = 50$ mtDNAs and $N_c = 7$ ptDNAs, meaning that the observed heteroplasmy variance could then be readily accounted for through random cell divisions and reamplification alone.

In the wild-type mtDNA, much more segregation is observed than can be accounted for by 34 cell divisions – the average number of inferred events is *c.* 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 wild-type 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 section, we consider alternative mechanisms. That reference characterised the contribution of gene conversion to normalised heteroplasmy variance $V'(h)$ 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 *msb1* and wild-type mtDNA is roughly 0.5, this expression suggests that a rate of $\kappa = 0.007$ per cell division (corresponding to $c. 0.1$ gene conversion events per mtDNA per cell division; see the [Materials and Methods](#) section) would be sufficient to generate the observed segregation patterns over $c. 34$ cell divisions.

The expression for how much variance is generated by gene conversion employed a particular mathematical assumption (a linear noise approximation) that may be challenged by the substantial segregation magnitudes involved in this system. To check these results, we constructed a simulation-based stochastic model for oDNA during development, including the binomial cell divisions and relaxed replication used previously, and a variable rate of gene conversion in a population of $N_e = 50$ oDNA molecules (see the [Materials and Methods](#) section). We asked what rates of gene conversion were required to generate the observed $V'(h)$ within $c. 34$ cell divisions, finding support for figure $c. 0.25$ events per mtDNA per cell cycle (Fig. S3). This simulation model provides predictions for heteroplasmy distributions at any given stage of plant development (Fig. S6). 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 (Schmid *et al.*, 2005; Winter *et al.*, 2007; Fig. S5) 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 parameter estimates we have presented are integrated over all the model structures in Fig. 1(a), so that they reflect ‘universal’ behaviour regardless of the support for the individual models. However, the RJMCMC process also estimates the statistical support for our different models of the plant germline. Interestingly, we initially observed some diversity in the estimated probabilities over these different model structures (Fig. 4b). The mtDNA *msb1* data have strong support for the ‘linear germline’ model, while the mtDNA wild-type and ptDNA *msb1* data provide strong support for the ‘all separate lineages’ model (Fig. S4).

To interpret these findings, it helps to consider the behaviour of heteroplasmy statistics under the different models. Under ‘all separate lineages’, samples from different developmental stage (EL, LL, and INF) reflect the mean heteroplasmy of an early stem cell (CS1) and have independent heteroplasmy variances. Under a ‘linear germline’, progressive sampling events form the precursor state for each developmental stage. Differences in heteroplasmy mean can therefore arise due to this sampling, and the heteroplasmy 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 *msb1* data have one developmental lineage in particular that suggests a strong shift in mean heteroplasmy (top right of Fig. 1c), where an initial heteroplasmy $c. 0.85$ 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). However, the substantial potential for a single family of observations to alter the weighting of these results means we cannot make definitive claims here, and further characterisation of somatic heteroplasmy in wild-type 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 mathematical models; the models also make further predictions about heteroplasmy distributions at any stage of plant development (Fig. S6). 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 10-fold increase in segregation strength for mtDNA, leading to rapid shifts towards homoplasmy, which cannot be explained by cell divisions alone.

We do not have measurements of heteroplasmic ptDNA on the wild-type background – all lines measured so far have been homoplasmic. The predictions of this theory for wild-type 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),

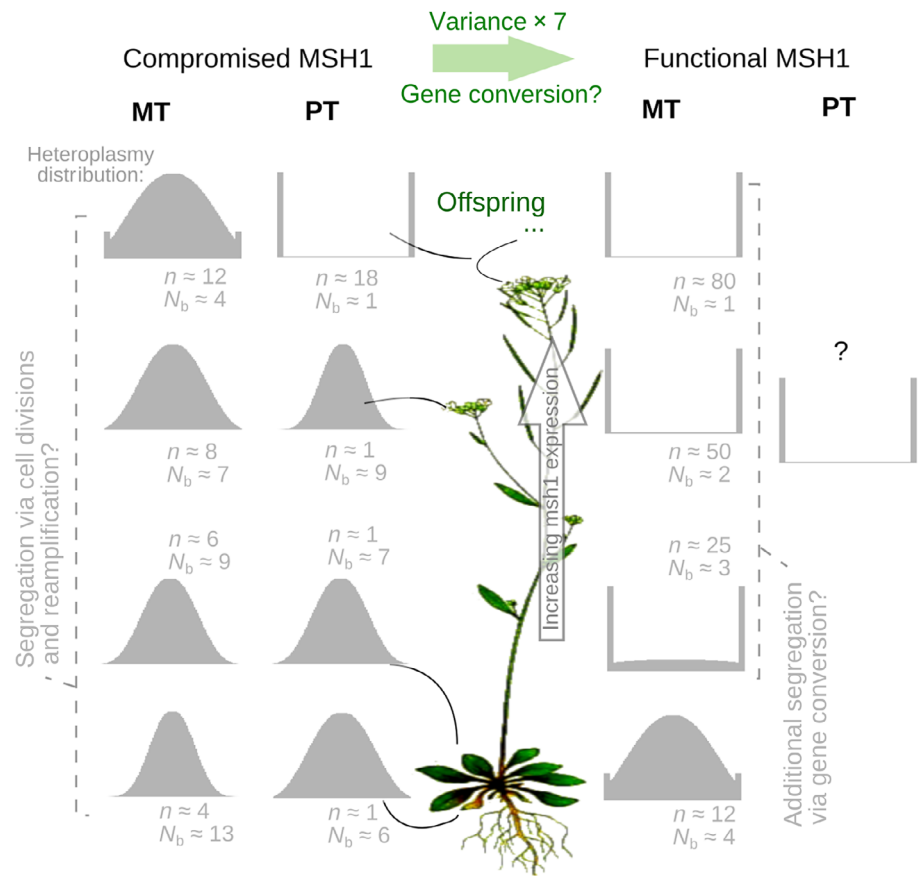


Fig. 5 Summary of inferred segregation dynamics within plants and between generations. Illustrative distributions of heteroplasmy in *Arabidopsis thaliana*, corresponding to the inferred mean segregation magnitude (n segregating events, for $N_e = 50$ mtDNAs or $N_e = 7$ ptDNAs; and N_b , effective bottleneck size). Distributions at each developmental stage, and an initial heteroplasmy of 0.5, are shown for mtDNA (MT) and ptDNA (PT) in wild-type and *msh1* mutants (all wild-type PT observations are homoplasmic, so no inference is possible; see the Discussion section 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.

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 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 the Materials and Methods section), but are not expected to be universally true or perfectly precise single values. oDNA population sizes change through development (see the Materials and Methods section and references therein) and oDNA partitioning at cell divisions may be more or less tightly controlled than a binomial distribution (Johnston *et al.*, 2015; Jajoo *et al.*, 2016). 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 wild-type 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 the Materials

and Methods section and Fig. S2) 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 mtDNA 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. We do not find evidence for vegetative ptDNA sorting in the *Arabidopsis msh1* mutant between early-leaf and inflorescence stages, in contrast to results in *Campanulastrium americanum* (Barnard-Kubow *et al.*, 2017). Our results suggest that ptDNA in *msh1 Arabidopsis* is already substantially sorted by the early-leaf stage (Figs 4, 5). It may be that the *msh1* mutation slows vegetative sorting in plastids as it does in mitochondria and that further plastid vegetative sorting during later stages of development would be observed in a wild-type background, or that ptDNA sorting after this stage is indeed more limited in *Arabidopsis*, which does not exhibit biparental plastid inheritance and therefore may have experienced less pressure to evolve vegetative ptDNA sorting.

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 wild-type, 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 (Arrieta-Montiel *et al.*, 2001; Abdelnoor *et al.*, 2003; 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 strand 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 vs wild-type 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 (Fig. S5; 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 co-localisation 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.*, 2021, 2022). Together, these observations suggest a linked physical and genetic axis of control acting to shape oDNA through plant generations.

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 R35 GM148134. The authors are grateful to Ben Williams and David Logan for valuable discussion.

Competing interests

None declared.

Author contributions

AKB, DBS and IGJ conceived the research, analysed data, and wrote the manuscript. AKB performed the experiments. IGJ performed modelling and simulation and drafted the manuscript.

ORCID

Amanda K. Broz  <https://orcid.org/0000-0001-5663-3569>

Iain G. Johnston  <https://orcid.org/0000-0001-8559-3519>

Daniel B. Sloan  <https://orcid.org/0000-0002-3618-0897>

Data availability

All data and code are freely available at <https://github.com/StochasticBiology/plant-segregation>. The inference code is written in C; the data curation and visualisation are written in R (R Core Team, 2022), using libraries READXL (Wickham & Bryan, 2022), STRINGR (Wickham, 2019), GGLOT2 (Wickham, 2016), and GRIDEXTRA (Auguie, 2017).

References

- Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA. 2003. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proceedings of the National Academy of Sciences, USA* 100: 5968–5973.
- Albert B, Godelle B, Atlan A, De Paepe R, Gouyon PH. 1996. Dynamics of plant mitochondrial genome: model of a three-level selection process. *Genetics* 144: 369–382.
- Allen JF, Martin WF. 2016. Why have organelles retained genomes? *Cell Systems* 2: 70–72.
- Arimura SI. 2018. Fission and fusion of plant mitochondria, and genome maintenance. *Plant Physiology* 176: 152–161.
- Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S. 2001. Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 158: 851–864.
- Arrieta-Montiel MP, Shedje V, Davila J, Christensen AC, Mackenzie SA. 2009. Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics* 183: 1261–1268.
- Atlan A, Couvet D. 1993. A model simulating the dynamics of plant mitochondrial genomes. *Genetics* 135: 213–222.
- Augue B. 2017. *gridExtra: miscellaneous functions for "Grid" graphics*. R package v.2.3. [WWW document] URL <https://CRAN.R-project.org/package=gridExtra> [accessed 09 September 2017].
- Barnard-Kubow KB, McCoy MA, Galloway LF. 2017. Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. *New Phytologist* 213: 1466–1476.
- Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* 168: 39–50.
- Bentley KE, Mandel JR, McCauley DE. 2010. Paternal leakage and heteroplasmy of mitochondrial genomes in *Silene vulgaris*: evidence from experimental crosses. *Genetics* 185: 961–968.
- Broz AK, Keene A, Fernandes Gyorfy M, Hodous M, Johnston IG, Sloan DB. 2022. Sorting of mitochondrial and plastid heteroplasmy in Arabidopsis is extremely rapid and depends on MSH1 activity. *Proceedings of the National Academy of Sciences, USA* 119: e2206973119.
- Burgstaller JP, Johnston IG, Jones NS, Albrechtova J, Kolbe T, Vogl C, Futschik A, Mayrhofer C, Klein D, Sabitzer S *et al.* 2014. MtDNA segregation in heteroplasmic tissues is common *in vivo* and modulated by haplotype differences and developmental stage. *Cell Reports* 7: 2031–2041.
- Burgstaller JP, Kolbe T, Havlicek V, Hembach S, Poulton J, Piálék J, Steinborn R, Rüllicke T, Brem G, Jones NS *et al.* 2018. Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. *Nature Communications* 9: 2488.
- Burian A, De Reuille PB, Kuhlemeier C. 2016. Patterns of stem cell divisions contribute to plant longevity. *Current Biology* 26: 1385–1394.
- Chinnery PF, Samuels DC. 1999. Relaxed replication of mtDNA: a model with implications for the expression of disease. *The American Journal of Human Genetics* 64: 1158–1165.
- Christensen AC. 2014. Genes and junk in plant mitochondria – repair mechanisms and selection. *Genome Biology and Evolution* 6: 1448–1453.
- Chustecki JM, Etherington RD, Gibbs DJ, Johnston IG. 2022. Altered collective mitochondrial dynamics in the Arabidopsis *msh1* mutant compromising organelle DNA maintenance. *Journal of Experimental Botany* 73: 5428–5439.
- Chustecki JM, Gibbs DJ, Bassel GW, Johnston IG. 2021. Network analysis of Arabidopsis mitochondrial dynamics reveals a resolved tradeoff between physical distribution and social connectivity. *Cell Systems* 12: 419–431.
- Clegg MT, Gaut BS, Learn GH Jr, Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795–6801.
- Cree LM, Samuels DC, de Sousa Lopes SC, Rajasimha HK, Wonnapijit P, Mann JR, Dahl HHM, Chinnery PF. 2008. A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nature Genetics* 40: 249–254.
- Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: *Cell and molecular biology of plastids*. Berlin, Heidelberg, Germany: Springer, 65–119.
- Dellaportas P, Forster JJ, Ntzoufras I. 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing* 12: 27–36.
- Edwards DM, Røyrvik EC, Chustecki JM, Giannakis K, Glastad RC, Radzvilavicius AL, Johnston IG. 2021. Avoiding organelle mutational meltdown across eukaryotes with or without a germline bottleneck. *PLoS Biology* 19: e3001153.
- Fernandes Gyorfy M, Miller ER, Conover JL, Grover CE, Wendel JF, Sloan DB, Sharbrough J. 2021. Nuclear–cytoplasmic balance: whole genome duplications induce elevated organellar genome copy number. *The Plant Journal* 108: 219–230.
- Furner IJ, Pumfrey JE. 1992. Cell fate in the shoot apical meristem of *Arabidopsis thaliana*. *Development* 115: 755–764.
- Galtier N. 2011. The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology* 9: 1–3.
- Gao L, Guo X, Liu XQ, Zhang L, Huang J, Tan L, Lin Z, Nagawa S, Wang DY. 2018. Changes in mitochondrial DNA levels during early embryogenesis in *Torenia fournieri* and *Arabidopsis thaliana*. *The Plant Journal* 95: 785–795.
- Giannakis K, Arrowsmith SJ, Richards L, Gasparini S, Chustecki JM, Røyrvik EC, Johnston IG. 2022. Evolutionary inference across eukaryotes identifies universal features shaping organelle gene retention. *Cell Systems* 13: 874–884.
- Giannakis K, Broz AK, Sloan DB, Johnston IG. 2023. Avoiding misleading estimates using mtDNA heteroplasmy statistics to study bottleneck size and selection. *G3: Genes, Genomes, Genetics* 13: jkad068.
- Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.
- Greiner S. 2012. Plastome mutants of higher plants. In: *Genomics of chloroplasts and mitochondria*. Dordrecht, the Netherlands: Springer, 237–266.
- Greiner S, Golczyk H, Malinova I, Pellizzer T, Bock R, Börner T, Herrmann RG. 2020. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *The Plant Journal* 102: 730–746.
- Greiner S, Sobanski J, Bock R. 2015. Why are most organelle genomes transmitted maternally? *BioEssays* 37: 80–94.
- Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. 2014. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100: 107–120.
- Hagemann R. 2010. The foundation of extranuclear inheritance: plastid and mitochondrial genetics. *Molecular Genetics and Genomics* 283: 199–209.
- Irish VF, Sussex IM. 1992. A fate map of the Arabidopsis embryonic shoot apical meristem. *Development* 115: 745–753.
- Jajoo R, Jung Y, Huh D, Viana MP, Rafelski SM, Springer M, Paulsson J. 2016. Accurate concentration control of mitochondria and nucleoids. *Science* 351: 169–172.
- Johnston IG. 2019a. Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. *Molecular Plant* 12: 764–783.
- Johnston IG. 2019b. Varied mechanisms and models for the varying mitochondrial bottleneck. *Frontiers in Cell and Developmental Biology* 7: 294.
- Johnston IG, Burgstaller JP, Havlicek V, Kolbe T, Rüllicke T, Brem G, Poulton J, Jones NS. 2015. Stochastic modelling, Bayesian inference, and new *in vivo* measurements elucidate the debated mtDNA bottleneck mechanism. *eLife* 4: e07464.
- Johnston IG, Gaal B, Neves RPD, Enver T, Iborra FJ, Jones NS. 2012. Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Computational Biology* 8: e1002416.
- Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal* 46: 85–94.
- Kimura M. 1955. Solution of a process of random genetic drift with a continuous model. *Proceedings of the National Academy of Sciences, USA* 41: 144–150.
- Kirk P, Thorne T, Stumpf MP. 2013. Model selection in systems and synthetic biology. *Current Opinion in Biotechnology* 24: 767–774.
- Lanfar R. 2018. Do plants have a segregated germline? *PLoS Biology* 16: e2005439.

- Logan DC. 2006. The mitochondrial compartment. *Journal of Experimental Botany* 57: 1225–1243.
- Lonsdale DM, Brears T, Hodge TP, Melville SE, Rottmann WH. 1988. The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 319: 149–163.
- Mandel JR, Ramsey AJ, Holley JM, Scott VA, Mody D, Abbot P. 2020. Disentangling complex inheritance patterns of plant organellar genomes: an example from carrot. *Journal of Heredity* 111: 531–538.
- Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* 186: 299–317.
- McCauley DE. 2013. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytologist* 200: 966–977.
- Miller-Messmer M, Kühn K, Bichara M, Le Ret M, Imbault P, Gualberto JM. 2012. RecA-dependent DNA repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. *Plant Physiology* 159: 211–226.
- Mohanta TK, Mishra AK, Khan A, Hashem A, Abd-Allah EF, Al-Harrasi A. 2020. Gene loss and evolution of the plastome. *Genes* 11: 1133.
- Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, 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, USA* 97: 6960–6966.
- Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T. 2010. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal* 64: 948–959.
- R Core Team. 2022. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL <https://www.R-project.org/> [accessed 26 July 2023].
- Rowan BA, Oldenburg DJ, Bendich AJ. 2010. RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. *Journal of Experimental Botany* 61: 2575–2588.
- Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, Berthouly-Salazar C, 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: 434–445.
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics* 37: 501–506.
- Seguí-Simarro JM, Staehelin LA. 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: 168–171.
- Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* 19: 1251–1264.
- Stadler T, Pybus OG, Stumpf MP. 2021. Phylodynamics for cell biologists. *Science* 371: eaah6266.
- Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics* 16: 530–542.
- Virdi KS, Laurie JD, Xu YZ, Yu J, Shao MR, Sanchez R, Kundariya H, Wang D, Riethoven JJM, Wamboldt Y *et al.* 2015. Arabidopsis MSH1 mutation alters the epigenome and produces heritable changes in plant growth. *Nature Communications* 6: 1–9.
- Wai T, Teoli D, Shoubridge EA. 2008. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nature Genetics* 40: 1484–1488.
- Wallace DC, Chalkia D. 2013. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harbor Perspectives in Biology* 5: a021220.
- Wang DY, Zhang Q, Liu Y, Lin ZF, Zhang SX, Sun MX. 2010. The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. *Plant Cell* 22: 2402–2416.
- Watson JM, Platzer A, Kazda A, Akimcheva S, Valuchova S, Nizhynska V, Nordborg M, Riha K. 2016. Germline replications and somatic mutation accumulation are independent of vegetative life span in Arabidopsis. *Proceedings of the National Academy of Sciences, USA* 113: 12226–12231.
- Wickham H. 2016. *GGPLOT2: elegant graphics for data analysis*. New York, NY, USA: Springer-Verlag.
- Wickham H. 2019. *stringr: simple, consistent wrappers for common string operations*. R package v.1.4.0. [WWW document] URL <https://CRAN.R-project.org/package=stringr> [accessed 26 July 2023].
- Wickham H, Bryan J. 2022. *readxl: read excel files*. R package v.1.4.0. [WWW document] URL <https://CRAN.R-project.org/package=readxl> [accessed 26 July 2023].
- Wilton PR, Zaidi A, Makova K, Nielsen R. 2018. A population phylogenetic view of mitochondrial heteroplasmy. *Genetics* 208: 1261–1274.
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An “Electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS ONE* 2: 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: 657–671.
- Wonnapijit P, Chinnery PF, Samuels DC. 2008. The distribution of mitochondrial DNA heteroplasmy due to random genetic drift. *The American Journal of Human Genetics* 83: 582–593.
- Wu Z, Waneka G, Broz AK, King CR, Sloan DB. 2020. MSH1 is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. *Proceedings of the National Academy of Sciences, USA* 117: 16448–16455.
- Zhang H, Burr SP, Chinnery PF. 2018. The mitochondrial DNA genetic bottleneck: inheritance and beyond. *Essays in Biochemistry* 62: 225–234.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Validating model and inference approach.

Fig. S2 Scaling factors for converting effective population sizes.

Fig. S3 Simulated segregation with and without gene conversion.

Fig. S4 Inferred behaviour for different datasets.

Fig. S5 msh1 expression patterns during development.

Fig. S6 Predicted heteroplasmy distributions over cell divisions.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.