Somatic evolution of stem cell mutations in long-lived plants

Frank Johannes[1]

[1] Technical University of Munich, Plant Epigenomics, Emil-Ramann-Str. 4, 85354, Freising, Germany; correspondence: frank@johanneslab.org

Long-lived perennial plants accumulate numerous somatic mutations with age. Mutations originating in stem cells at the shoot apex often become fixed in large sectors of the plant body due to cell lineage drift during repeated branching. Understanding the somatic evolution of such mutations requires knowledge of the effective stem cell population size, the cellular bottleneck strength during branch initiation, and the mutation rate. Here we show that these parameters can be estimated directly from cell-layer-specific DNA sequencing data, thus filling a gap where no other *in vivo* method exists.

Keywords: Somatic mutations, somatic mutation rate, shoot apical meristem, stem cells, cell lineage tracing, trees, site frequency spectrum, somatic drift, DNA sequencing, single cell, clonal reproduction

Many perennial plants exhibit extraordinarily long lifespans, ranging from centuries to millennia. Trees are a good example of this. Even at an advanced age they grow vigorously, iteratively forming new branches that produce leaves and fruits. All new aerial plant structures emerge from the shoot apical meristem (SAM), a population of cells situated at the tip of each growing shoot (Lyndon (1998)). The SAM consists of stem cells and other meristematic cells, which continually divide, either to self-renew or to differentiate. Somatic mutations that arise during stem cell divisions are particularly consequential as they can be propagated through the plant architecture over time, creating substantial intra-organism genetic heterogeneity. There is considerable interest in understanding the adaptive significance of this heterogeneity, as well as the mechanisms that contribute to the rate and patterns of mutation accumulation within the plant (Schoen and Schultz (2019)).

In higher plants, new lateral branches are initiated from a small number of precursor cells at the periphery of the SAM (Burian et al. (2016), Shi et al. (2016)). Their cell lineages trace back to a group of "apical stem cells" (ASCs) at the SAM center. The ASCs are distinct from surrounding cells due to their spatial positioning and slow mitotic division rate, and thus constitute an important unit of

analysis (Burian (2021)). The limited number of precursors involved in branching creates cellular bottlenecks, allowing mutations in ASCs to become fixed in large sectors of the plant structure through somatic drift (Reusch et al. (2021)).

Such fixed somatic mutations sometimes confer advantageous phenotypic effects on specific plant parts or modules. This phenomenon has been extensively used in horticulture and fruit crop breeding, where mutant branches with desirable traits are propagated through techniques such as grafting, cuttings, or tissue culture (Foster and Aranzana (2018)). Advantageous somatic mutations have also been observed in natural settings (e.g. Padovan et al. (2013)), though they are likely rare. By contrast, deleterious ASC mutations are much more common (Eyre-Walker and Keightley (2007)), and contribute to a gradual build-up of mutational load over time (Scofield and Schultz (2006)). The harmful effects of these mutations are often hidden, however, because they typically to arise as recessive alleles in a heterozygous state (Schoen and Schultz (2019)). As a result, the somatic evolution of ASC mutations is thought to be largely shaped by neutral somatic drift during branching (Klekowski and Kazarinova-Fukshansky (1984b), Chen et al. (2024), Yu et al. (2020), Tomimoto and Satake (2023)), rather than by selection acting on differences in fitness among cell lineages or plant modules (Folse and Roughgarden (2012), Antolin and Strobeck (1985), Klekowski and Kazarinova-Fukshansky (1984a), Pineda-Krch and Fagerstrom (1999), Otto and Orive (1995)).

Nonetheless, deleterious somatic mutations can be subject to strong selection once they enter the haploid gametic phase of the plant life cycle, where their effects are fully exposed. Mutations that escape purging during this stage can be transmitted to offspring. In this way they can contribute to inbreeding depression at the population level (Lesaffre (2021)). Long-lived plant species appear to mitigate this problem by maintaining high outcrossing rates and large population sizes (Scofield and Schultz (2006)), both of which reduce inbreeding coefficients. Interestingly, an additional strategy utilized by long-lived plants to limit the accumulation of somatic mutations is to reduce the mutation rate itself (Lanfear et al. (2013), Johannes (2024), Shahryary et al. (2020)). This rate reduction seems to be particularly pronounced in ASC-derived cell lineages that eventually form the gametes (Goel et al. (2024)).

Recent DNA sequencing studies, mainly in tree species, have shown that the yearly rate of fixed somatic mutations is at least an order of magnitude lower than that observed in annual plants (Johannes (2024)). This trend is also evident in the lower evolutionary substitution rates detected in long-lived angiosperms and gymnosperms compared to short-lived species (De La Torre et al. (2017)). It has been proposed that these reduced rates are achieved by limiting the number of mitotic divisions in ASCs per unit time (e.g. Petit and Hampe (2006), Lanfear et al. (2013), Burian (2021), Johannes (2024)), thereby decreasing the likelihood of errors during DNA replication. An alternative, yet unexplored, hypothesis is that long-lived species have larger effective stem cell populations or recruit a greater number of precursor cells during branch initiation, both of which would dilute

the effects of somatic drift and reduce the rate of fixation. Theoretical models predict that larger stem cell populations could also buffer the impact of deleterious mutations, should they display some level of dominance (Klekowski and Kazarinova-Fukshansky (1984a)).

To begin to test these hypotheses and to study the somatic evolution of ASC mutations from a quantitative perspective, detailed knowledge of the effective stem cell population size, the cellular bottleneck strength during branch initiation, and the mutation rate are needed for diverse plant systems. Despite recent technological advances in live imaging (e.g. Bradamante (2024)) and Crisper-based cell lineage tracing (e.g. Lu et al. (2024)), direct measurements of these parameters remain experimentally challenging. A major obstacle is to obtain measurements *in vivo*, and in a manner that is applicable across a wide range of plant species, including those that cannot be easily transformed genetically. Moreover, neither technology provides information about mutation rates.

One promising alternative strategy is to develop theoretical models of shoot branching that describe how these parameters affect changes in the frequency of ASC mutations in the SAM cell population over time. As these mutation frequencies can be measured directly using bulk sequencing of selected somatic tissues, fitting the model to such data could yield statistical estimates of the unknown parameters. Although theoretical models of somatic evolution have been developed, none are currently designed to be applied to actual somatic sequencing data for parameter estimation. This is particularly true of early theoretical models that focused on mutant cells as units of analysis (e.g. Klekowski and Kazarinova-Fukshansky (1984b), Klekowski and Kazarinova-Fukshansky (1984a), Otto and Orive (1995), Pineda-Krch and Fagerstrom (1999)). More recent models do have a DNA focus, but their application appears to require single-cell data of individual SAMs (Iwasa et al. (2023), Chen et al. (2024)), which is technically infeasible at the moment, or they make strong numerical assumptions about the very parameters that require estimation (Tomimoto and Satake (2023), Yu et al. (2024)).

Here we overcome many of these challenges and demonstrate that a modeling approach in combination with bulk sequencing of somatic cells is feasible. We begin by laying out the biological framework underlying the theoretical branching model. We then develop the model step-by-step and show how it can be used to derive and compute the expected frequency spectrum of ASC mutations in the SAM cell population. Finally, we provide a proof-of-principle demonstration of how parameter estimates can be obtained by fitting the model to recent layer-enriched DNA sequencing data of an apricot tree.

Biological framework

The shoot apical meristem (SAM) is a zone of actively dividing cells at the shoot tip, driving the formation of new stems, leaves, and flowers. In seed plants, the SAM is organized into a tunica-corpus structure, with the tunica comprising the

outer cell layers and the corpus forming the inner region. Most species possess a tunica that consists of one or two layers (L1, L2) where cells divide anticlinally (perpendicular to the surface). The underlying corpus (L3) has cells dividing in multiple directions (Lyndon (1998)) (Fig. S1). Stem cells in the SAM have the capacity for self-renewal and for producing progenitor cells that differentiate (Heidstra and Sabatini (2014)). Live imaging and cell lineage tracing in Arabidopsis and tomato (Burian et al. (2016)) have identified 3-4 resident stem cells at the very tip of the SAM surface that distinguish themselves from surounding cells by their extremely slow division rates (Lyndon (1998), Kwiatkowska (2008))(Fig.1A). These so-called apical stem cells (ASCs) are the origin of all SAM cell lineages, including those forming lateral organs and the germline (Burian (2021)). A similarly arranged, but independent, population of ASCs is present in the L2 and the inner L3 layer, just below the SAM surface. We will only consider the L1 case further, but similar arguments holds for the other layers.

The cellular basis of branching

In higher plants, lateral branches arise from axillary buds, which form in the axils of developing leaves along the stem (Fig. S2). These buds originate from clusters of cells called axillary meristems (AMs) (Fig.1A, Fig. S2). AMs develop from a small group of precursor cells that are sequestered from the flanks of the SAM (Burian et al. (2016), Shi et al. (2016)), a process known as the "detached meristem model" of shoot branching (Wang and Jiao (2018)). These precursor cells retain meristematic properties within the leaf axil, gradually proliferating to form active AMs that later produce new lateral branches (Fig. S2). Once established, these AMs function as the SAM of the emerging shoot, possessing the same structural organization, self-maintenance capabilities, and ability to generate new organs as the original SAM (Nicolas and Laufs (2022)). In this way, shoot branching reflects a developmental shift, where the SAM of the primary shoot gives rise to new SAMs on lateral branches (Fig.1B, Fig. S2).

Somatic drift and cell lineage fixation

In the time interval between two branching events, the ASC population self-renews through asymmetrical cell divisions (Fig. S3). Each apical stem cell produces two descendant cells, where one retains its stem cell function at the central position of the SAM, while the other is displaced toward the SAM periphery, where it undergoes a series of symmetrical cell divisions (Fig. S3). In the L1 layer, this latter process generates spatially defined clonal sectors in the SAM, with each sector having its origin in one of the apical stem cells at the center (Burian et al. (2016)) (Fig.1A,B). The sampling of L1-derived precursor cells at the SAM periphery is therefore an indirect sampling of ASCs, albeit via their clonal descendants. Depending on the number of clonal sectors and the degree of spatial constraints in AM establishment, precursor sampling can be a source of substantial somatic drift (Reusch et al. (2021), Chen et al. (2024)), leading to ASC lineages becoming either fixed or lost in the SAM over time (Fig.1C). In the limiting case, AMs are

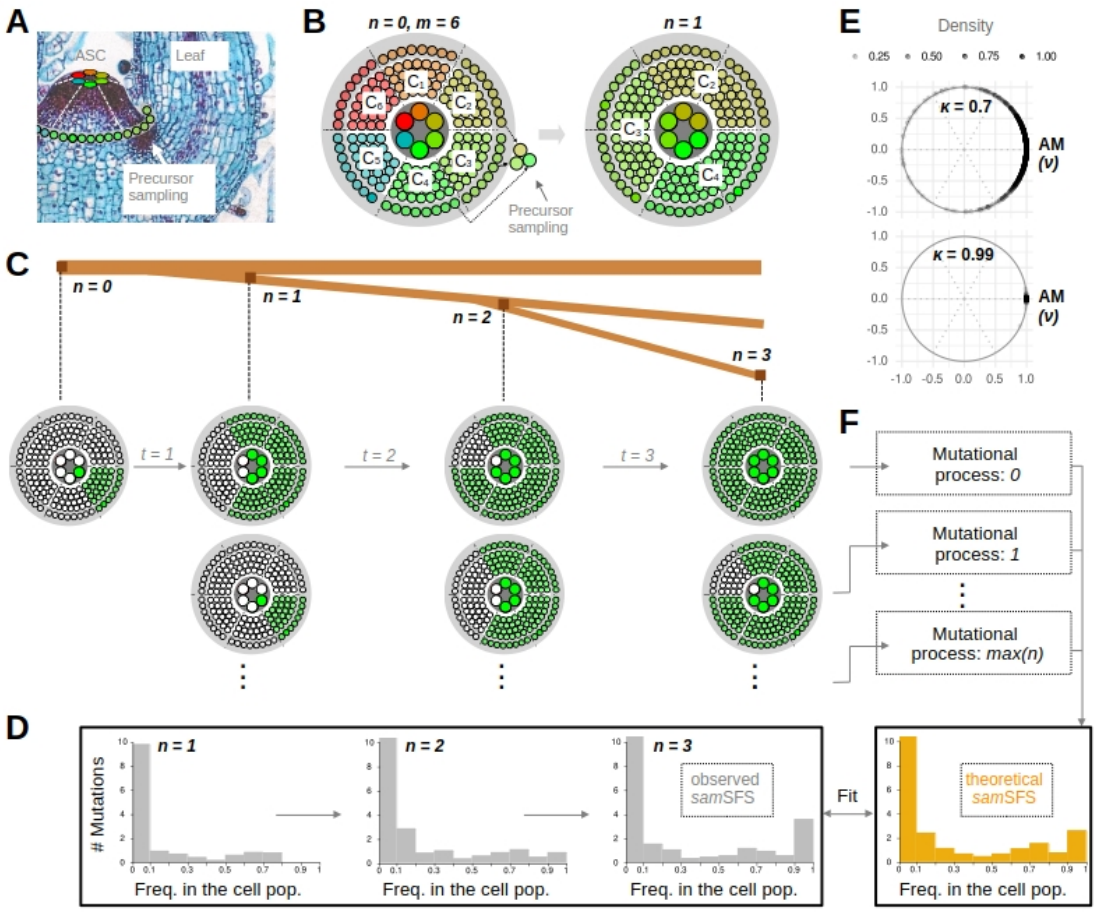

Figure 1: Biological basis and conceptual framework. A. Side-view of the shoot apical meristem (SAM). The colored circles at the tip represent six ($m = 6$) apical stem cells (ASC). Clonal descendants of these ASCs are shown in corresponding colors at the SAM periphery. The arrow points to the site of axillary meristem (AM) establishment. **B.** A schematic top-view of the SAM. Six clonal sectors ($C_1, \cdots, C_6$) are shown at $n = 0$, with each sector corresponding to one of the ASC at the tip. At first branching ($n = 1$), precursors are "sampled" from C_2, C_3 and C_4 during AM establishment, and give rise to a new SAM. The cellular composition of the new SAM reflects the proportional sampling of these sectors. **C.** Cell lineage composition of the SAM across three nested branching events. Here, lineages from clonal sector C_4 become fixed in the SAM over time. In time intervals t_1, t_2 and t_3 the ASC self-renews. Mutations in ASCs that occur in these intervals follow the fate of the cell lineages in which they arise. There is a new mutational processes, starting with each new branching iteration n, giving rise to mutations that are nested within those that appeared earlier. **D.** The frequencies of these mutations within the SAM cell population defines the site frequency spectrum (*sam*SFS) at n. **E.** Modeling of precursor sampling using a wrapped density distribution. The distribution governs the location of AM establishment (ν) and the density spread of selected precursor around the site of AM establishment (κ). Sampling can be highly localized (bottom) or more spread around the site of AM establishment (top). **F.** Modeling of precursor sampling along with the mutational processes during nested branching yields the theoretical *sam*SFS, which can be fitted to the observed *sam*SFS for parameter estimation.

initiated from a single cell from each of the SAM layers. In this scenario, there is immediate fixation of one clonal lineage in the new SAM. However, if precursor sampling is more "relaxed", involving cellular contributions from multiple clonal

sectors, fixation may require several nested branching iterations to reach fixation (Fig.1C).

Mutation accumulation in the SAM

The developmental fate of *de novo* mutations within the SAM is closely related to the fate of the cell lineages in which they arise. We focus exclusively on mutations within the ASC population. We assume that these mutations occur with some rate μ (per bp per year) during the self-renweal phase between branching events, and that their frequencies in the SAM evolve neutrally through somatic drift. To see this, consider a mutation that arises in an apical stem cell in the first internode time interval t_1 (Fig.1C). This mutation essentially "barcodes" all cell lineages that derive from this cell, and its fixation dynamics is therefore equivalent to the fixation of the cell lineage that carries it.

However, the interval t_1 is not the only interval contributing mutations to the SAM. New and independent mutations also arise in ASCs in any subsequent branching intervals $t_2, t_3, \ldots t_n$, generating mutational "barcodes" that are nested in those acquired in earlier intervals (Fig.1C). The continuous flow of *de novo* mutations into the SAM, coupled with somatic drift, therefore generates a distribution of mutational frequencies in the SAM during branching, with developmentally older mutations being already fixed at high frequency in the SAM population, and more recent mutations appearing at intermediate to low frequencies. We define the complete mutation frequency distribution as the SAM site frequency spectrum (*sam*SFS) (Fig.1D). The shape of the *sam*SFS at any point in time holds statistical information about the underlying branching process, including the mutation rate (μ), the size of the ASC population (m) as well as the spatial constraints during precursor sampling (κ).

In what follows we develop a theoretical model that incorporates these parameters (Fig.1E,F). Using the theory we derive the theoretical *sam*SFS, and show that it can be fitted to layer-specific DNA sequencing data for parameter estimation.

Theory

Unit circle representation of the SAM

Let us represent the L1 surface layer of the SAM as a unit circle in polar coordinates (Fig.1E). The unit circle is defined as:

$$S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\},$$

which describes all points (x, y) on the boundary of a circle with a radius of 1 in the two-dimensional plane. We assume that the unit circle is divided into m cell partitions, labeled $C_1, C_2, \ldots, C_m$. Each partition corresponds to a subpopulation of clonal cells descended from a single ASC at the center of the SAM (Fig.1B).

The angular width of each partition i, denoted by $\Delta\theta_i$, is defined as the difference between its starting angle a_i and its ending angle b_i:

$$\Delta\theta_i = b_i - a_i.$$

We suppose that at $n = 0$, when the SAM is first established during development, all m partitions are of equal size (Fig.1B). Hence, the initial angular width for each sector is:

$$\Delta\theta_i^{(0)} = \frac{2\pi}{m},$$

with the starting angle $a_i^{(0)}$ and the ending angle $b_i^{(0)}$ given by:

$$a_i^{(0)} = (i-1)\cdot\Delta\theta_i^{(0)}, \quad b_i^{(0)} = i\cdot\Delta\theta_i^{(0)},$$

where $i = 1, \ldots, m$. The uniform partitioning of clonal sectors at $n = 0$, implies that the probability of each sector $C_i^{(0)}$ is simply:

$$P(C_i^{(0)}) = \frac{1}{m}.$$

Branching: precursor sampling and re-normalization

Branching is initiated by the formation of AMs at the SAM periphery (Fig.1A,B). It starts with the selection (i.e. sampling) of a small number of peripheral cells to act as precursors. For simplicity, several previous models assumed that this sampling is uniform (Klekowski and Kazarinova-Fukshansky (1984b), Klekowski and Kazarinova-Fukshansky (1984a), Klekowski et al. (1985), Otto and Orive (1995), Chen et al. (2024), Yu et al. (2024), Pineda-Krch and Fagerstrom (1999), Antolin and Strobeck (1985)), so that all peripheral cells have an equal chance of being selected. In reality, the sampling processes is spatially constrained, often involving only cells that are in close proximity of each other. The degree of this constraint can depend on factors such as the precise SAM topology and the actual number of precursors, all of which can vary across species (Schnablová et al. (2017)). To account for these spatial considerations Tomimoto and Satake (2023) proposed to model precursor sampling using a wrapped distribution on a unit circle (Fig.1E). A wrapped distribution is derived by taking a standard linear distribution and "wrapping" it around the circle to ensure that the random radian variable θ remains confined to the interval $[0, 2\pi)$. The general form of the probability density function (PDF) of a wrapped distribution is:

$$f(\theta \mid \nu, \kappa) = \sum_{q=-\infty}^{\infty} g(\theta + 2\pi q \mid \nu, \kappa), \tag{1}$$

where q is an integer index that represents how many times the real line has been "wrapped" around the circle (Kato and Jones (2013)). As can be seen, the distribution depends on two parameters, ν and κ. Both have important biological

interpretations. The parameter ν is the location parameter and gives the "mean direction"; that is, the angle around which the distribution is centered. Biologically, this sets the location of AM initiation and lateral bud formation along the SAM periphery (Fig.1E). The parameter κ is the "concentration" parameter. It controls how tightly the probability mass is concentrated around ν. In biological terms, κ determines the degree of spatial constraint in precursor selection. For small κ, the distribution is spread out more uniformly around the circle, while for large κ, the distribution becomes tightly concentrated around ν (Fig.1E). The use the parameter κ as a way to quantify the bottleneck strength during precursor sampling.

Now, consider the first branching event at $n = 1$. We are interested in the probability that cells from clonal partitions $C_1^0, C_2^0, \ldots, C_m^0$ are selected as precursors for AM initiation at some location $\nu^{(1)}$ along the SAM periphery. This probability can be derived for each C_i by integrating Equation 1 over the angular range of each partition:

$$\int_{a_i^{(0)}}^{b_i^{(0)}} f(\theta \mid \nu^{(1)}, \kappa)\, d\theta = F(b_i^{(0)} \mid \nu^{(1)}, \kappa) - F(a_i^{(0)} \mid \nu^{(1)}, \kappa).$$

where $F(\cdot \mid \nu, \kappa)$ is the cummulative distribution function (CDF), $a_i^{(0)}$ and $b_i^{(0)}$ are the angular start and end points of the i-th partition just prior to branching, respectively, and $\nu^{(1)}$ is the location of AM initiation. As the AM matures into a bud and forms a new SAM on the emerging branch, the clonal partitions have to be re-normalized to account for the new cellular proportions:

$$P(C_i^{(1)}) = \frac{F(b_i^{(0)} \mid \nu^{(1)}, \kappa) - F(a_i^{(0)} \mid \nu^{(1)}, \kappa)}{\sum_{j=1}^{m} \left[F(b_j^{(0)} \mid \nu^{(1)}, \kappa) - F(a_j^{(0)} \mid \nu^{(1)}, \kappa)\right]}.$$

where the summation in the denominator ensures that the total probability across all $C_1^1, C_2^1, \ldots, C_m^1$ remains 1. These updated probabilities thus present the proportion of SAM cells following branching at $n = 1$, whose lineages trace back to apical stem cell i at $n = 0$.

Nested branching and cell lineage evolution

The above arguments suggest the following recurrence relation for iterative nested branching events:

$$P(C_i^{(n+1)}) = \frac{F(b_i^{(n)} \mid \nu^{(n+1)}, \kappa) - F(a_i^{(n)} \mid \nu^{(n+1)}, \kappa)}{\sum_{j=1}^{m} \left[F(b_j^{(n)} \mid \nu^{(n+1)}, \kappa) - F(a_j^{(n)} \mid \nu^{(n+1)}, \kappa)\right]}. \tag{2}$$

with updated boundaries:

$$b_i^{(n+1)} = a_i^{(n)} + 2\pi \cdot P(C_i^{(n+1)}), \quad a_{i+1}^{(n+1)} = b_i^{(n+1)}.$$

Notice that we allow the mean direction $\nu^{(n)}$ of the wrapped distribution to change with each branching event n. This should account for the fact that the location

of AM initation along the SAM periphery is not fixed over time. We model the dynamic behavior of $\nu^{(n)}$ in two ways. First, we assume that $\nu^{(n)}$ rotates around the unit circle according to the following deterministic function:

$$\nu^{(n)} = \nu_0 + \omega n, \tag{3}$$

where ω is a fixed angular velocity. This model can capture different forms of phyllotaxy; that is, the periodic placement of leaf primordia and their accompanying AMs along the axis of shoot growth (Kuhlemeier (2017)). For example, we can have $\omega = \pi$ so that AM initiation is shifted by a half a circle with each branching iteration, thus completing a full rotation around the SAM with every second branching event. A second model assumes that $\nu^{(n)}$ is randomly chosen, at each branching iteration, from a uniform distribution on the interval $[0, 2\pi)$, i.e.,

$$\nu^{(n)} \sim \mathcal{U}(0, 2\pi).$$

This model is realistic if one considers that not all AMs actually develop into lateral branches. It could therefore reflect some level of stochasticity in AM activation or dormancy. In addition, the spatial coordinates of the clonal sectors in the newly established SAM are not fixed, but re-established randomly with respect to the site of AM initiation.

The evolution and convergence of the recurrence Equation 2 can be evaluated numerically using:

$$\left| P\left(C_i^{(n+1)}\right) - P\left(C_i^{(n)}\right) \right| < \epsilon,$$

where ϵ is a small positive number. Note, that this also implies that the SAM partition boundaries stabilize:

$$\left| a_i^{(n+1)} - a_i^{(n)} \right| < \delta \quad \text{and} \quad \left| b_i^{(n+1)} - b_i^{(n)} \right| < \delta,$$

where δ is a small positive number.

Fixation and fixation times

We seek to explore the conditions, at convergence, under which the probability $P(C_i)$ for a given partition reaches a state of effective "fixation", meaning it approaches or effectively reaches a probability close to 1. For a given partition $C_i^{(n)}$, we define fixation as:

$$P(C_i^{(n)}) \geq 1 - \epsilon,$$

where ϵ is a small positive number, indicating that $P(C_i^{(n)})$ is "sufficiently close" to 1 to be considered fixed. Now, let N_i denote the fixation time for partition C_i. This is defined as the smallest branching iteration n for which the fixation criterion is first met:

$$N_i = \min\{n \,|\, P(C_i^{(n)}) \geq 1 - \epsilon\}. \tag{4}$$

While the above derivation focuses on the fixation of cell lineages that originate in ASC i, there is a direct connection with the fixation of *de novo* mutations. As already argued above, a mutation that first arises in ASC i in time interval t_1 "barcodes" all cell lineages that derive from this ASC, and its fixation dynamics is therefore equivalent to that of the cell lineages themselves. Since the lineages of only one of the m ASCs converges to fixation, the probability of a mutation in a random ASC going to fixation is therefore simply m^{-1} (Klekowski and Kazarinova-Fukshansky (1984b)).

The site frequency spectrum of apical stem cell mutations

We can build on the above insights to derive the theoretical *sam*SFS; that is, the site frequency spectrum (SFS) of *de novo* mutations that originate in ASCs. The *sam*SFS quantifies how many mutations are found at different frequencies within the SAM cell population at any point n. To set ideas, consider again those mutations that arose in ASC i during time interval t_1. Their expected number in the genome is:

$$\lambda_1 = 2G \cdot \mu \cdot t_1,$$

where $2G$ denotes two times the haploid genome size, t_1 is the time measured in units of ASC divisions or in years, and μ is the mutation rate per bp, per haploid genome, per unit time. The proportion of SAM cells that share these mutations are given by the sector probabilities $P(C_1^{(n)} \ldots, P(C_m^{(n)})$, derived above. However, as pointed out above, additional mutations arise in ASCs in subsequent branching intervals $t_2, t_3, \ldots t_n$, thus generating mutational "barcodes" that are nested in those acquired in earlier intervals (Fig.1C).

To formalize this idea, we must consider a generalization of the recurrence equations introduced above (see Equation 2), and allow for multiple mutation-drift processes to evolve simultaneously, starting at different branching iterations, but sharing a common wrapped probability distribution at each step. For each process k, which begins with branching iteration $n = k$, the unit circle is initially divided into m equal partitions. The partition for process k at iteration n could be denoted as:

$$C_{i,k}^{(n)} = [a_{i,k}^{(n)}, b_{i,k}^{(n)}),$$

where i is the partition number, n is the current branching iteration, and k denotes the specific process starting at $n = k$. The initial probabilities for each partition are equal: $P(C_{i,k}^{(n)}) = \frac{1}{m}$.

At each branching event a shared wrapped distribution $f(\theta; \nu^{(n)}, \kappa)$, with mean direction $\nu^{(n)}$ and concentration parameter κ, governs the assignment of probability mass to partitions. The probability updates of these partitions is given by:

$$P(C_{i,k}^{(n+1)}) = \frac{F(b_{i,k}^{(n)} \mid \nu^{(n+1)}, \kappa) - F(a_{i,k}^{(n)} \mid \nu^{(n+1)}, \kappa)}{\sum_{j=1}^{m} \left[F(b_{j,k}^{(n)} \mid \nu^{(n+1)}, \kappa) - F(a_{j,k}^{(n)} \mid \nu^{(n+1)}, \kappa)\right]}.$$

Assuming an infinite sites models, *de novo* mutations arising in any of the k processes can therefore be uniquely assigned to one of $m \times k$ categories, where m is the number of SAM partitions. The expected number of *de novo* mutations that originate in time interval t_k from processes k is:

$$\lambda_k = 2G \cdot \mu \cdot t_k.$$

To construct the *sam*SFS at any given iteration n, we need to examine the probabilities $P(C_{i,k}^{(n)})$ and classify each into a set of predefined, non-overlapping frequency ranges over the interval $(0, 1]$, excluding zero. Let each range be denoted as:

$$R_j = (r_{j-1}, r_j),$$

where r_{j-1} and r_j represent the lower and upper bounds of each range, respectively, and $j = 1, 2, \ldots, P$, with P representing the total number of ranges over $(0, 1]$.

We introduce an indicator function I_{R_j} that "activates" when $P(C_{i,k}^{(n)})$ falls within the range R_j:

$$I_{R_j}(P(C_{i,k}^{(n)})) = \begin{cases} 1 & \text{if } P(C_{i,k}^{(n)}) \in R_j, \\ 0 & \text{otherwise.} \end{cases}$$

To count how many of the m partitions $C_{i,k}^{(n)}$ fall within each range R_j for a given iteration n and process k, we sum the indicator function over all $i = 1, 2, \ldots, m$:

$$\text{Count}(R_j, n, k) = \sum_{i=1}^{m} I_{R_j}(P(C_{i,k}^{(n)})).$$

This provides the total count of $C_{i,k}^{(n)}$ values whose probabilities fall within range R_j at branching iteration n for process k. Hence, the total number of *de novo* mutations falling into value range j across all k processes is given by:

$$L_{\Psi,j}^{(n)} = \sum_{k} \lambda_k \cdot \text{Count}(R_j, n, k).$$

The distribution of $L_{\Psi}^{(n)}$ over all P frequency ranges defines the theoretical *sam*SFS at branching iteration n, where the parameter vector $\Psi = (\nu, \kappa, \mu, m)$ contains all the model parameters, except of $t_1, \ldots, t_n$, G and n, which are assumed to be known.

Data application and parameter estimation

Goodness-of-fit statistic: Given a set of mutations identified from a layer-specific sequencing analysis of the SAM, or its derived lateral organs, we can construct the observed *sam*SFS and denote it by $L_{obs}^{(n)}$. The Goodness-of-fit statistic comparing $L_{obs}^{(n)}$ with $L_{\Psi}^{(n)}$ is:

$$\chi_{\Psi}^2 \mid n, t_1, \ldots, t_n, G = \sum_{j=1}^{P} \frac{\left(L_{\text{obs},j}^{(n)} - L_{\Psi,j}^{(n)}\right)^2}{L_{\Psi,j}^{(n)}},$$

where P is the total number of frequency ranges over which the *sam*SFS is defined. Under the null hypothesis, the statistic is distributed as a Chi-square distribution with $P-1$ degrees of freedom. To evaluate the Goodness-of-fit statistic we require estimates of the model parameters, Ψ. There are a number of strategies to obtain these estimates.

Estimating Ψ: One strategy is to treat the Chi-square statistic as an objective function and minimize it with respect to Ψ:

$$\hat{\Psi} = \arg\min_{\Psi} \chi^2_{\Psi}. \tag{5}$$

This strategy should work even if mutation data from only a single sample is available, provided the mutation counts are large enough. With multiple samples, say $s = 1, \ldots, S$, obtained from independent branches, the mutation counts could be simply aggregated in $L^{(n)}_{obs}$. By "independent branches", we mean branches that share as little mutational history as possible. Because this is not entirely possible, given the shared stem segment at the bottom of the tree, any shared *de novo* mutations could be identified beforehand and only counted once. With multiple samples, the formulation of $L^{(n)}_{\Psi}$ requires the following adjustment:

$$\lambda_k = (2G \cdot S) \cdot \mu \cdot t_k.$$

Assuming that S is sufficiently large, an alternative estimation strategy is to assume that there are S Poisson-distributed counts within each of the P frequency categories. Denote the Poisson mean parameter for category j as $L^{(n)}_{j,\Psi}$. Let $x_{j,s}$ be the s-th observed count in category j, where $s = 1, \ldots, S$ and $j = 1, \ldots, P$. The probability of observing $x_{j,s}$ given the mean $L^{(n)}_{j,\Psi}$ is:

$$P(X_{j,s} = x_{j,s} \mid L^{(n)}_{j,\Psi}) = \frac{(L^{(n)}_{j,\Psi})^{x_{j,s}} e^{-L^{(n)}_{j,\Psi}}}{x_{j,s}!}.$$

Because each category contributes equally to the total probability, we weigh each category's contribution by P^{-1}. The overall likelihood function for the observed data across all categories and realizations is the product of the likelihoods from each category:

$$L(\Psi) = \prod_{j=1}^{P} \left(\prod_{s=1}^{S} \frac{(L^{(n)}_{j,\Psi})^{x_{j,s}} e^{-L^{(n)}_{j,\Psi}}}{x_{j,s}!} \right)^{\frac{1}{P}}.$$

The log-likelihood is:

$$\log L(\Psi) = \frac{1}{P} \sum_{j=1}^{P} \sum_{s=1}^{S} \left(x_{j,s} \log L^{(n)}_{j,\Psi} - L^{(n)}_{j,\Psi} - \log(x_{j,s}!) \right).$$

To find the parameter estimates $\hat{\Psi}$ that maximize $\log L(\Psi)$, we need to solve:

$$\hat{\Psi} = \arg\max_{\Psi} \log L(\Psi).$$

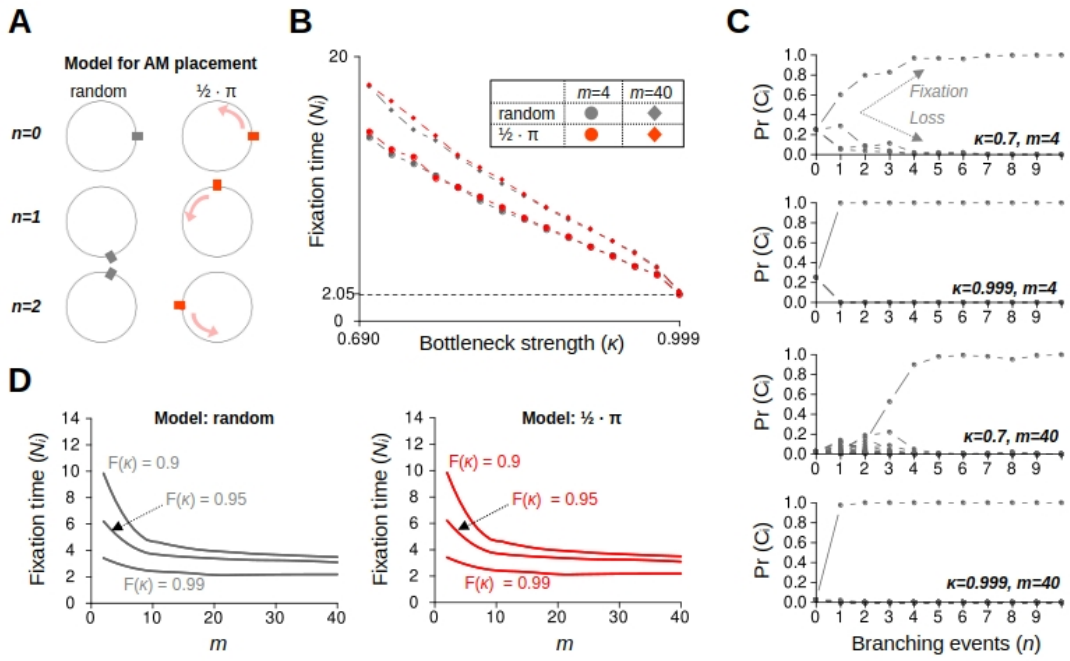

Figure 2: Fixation times of apical stem cell mutations. A. Two models of axillary meristem (AM) placement. In the first model, AMs are established at random locations along the SAM periphery with each branching iteration. In the second model, AM locations are sampled according to a phylotaxic process in steps of $0.5 \cdot \pi$. **B.** Computation of fixation times as a function of the concentration parameter κ and the size of the ASC population m for the two AM placement models. Fixation times denote the number of nested branching events that are required for a cell lineage that originates in one of the ASC cells at $n = 0$ to reach a relative frequency of 1 within the SAM. These fixation times also hold for *de novo* mutations that have arisen in the ASC self-renewal phase t_1. **C.** Shown are several numerical realizations of the stochastic processes that leads to the fixation (and loss) of cell lineages. The probability of clonal sector C_i (y-axis) as a function of the number of nested branching events (x-axis). The probability measures the proportion of all SAM cells that are descendants of apical stem cell i at $n = 0$. Fixation is relatively slow in situations where κ is relaxed ($\kappa = 0.7$) and m is large ($m = 40$), leading to intermediate frequency in early branching iterations. **D.** Fixation times become nearly independent of m as $\kappa \to 1$. The reason is that precursor sampling becomes progressively restricted to a single clonal sector, creating strong somatic drift. In the limiting case, only a single precursor cell is chosen for AM establishment.

Depending on the specific form of the wrapped distribution, estimates for Ψ will have to rely on numerical methods, as many lack closed-form CDFs. However, it may be possible to derive analytical expressions for some special cases. Future work could explore this possibility in more detail.

Results and discussion

Fixation times of apical stem cell mutations

We computed the expected fixation time (N) of a *de novo* mutation occuring in an apical stem cell, given that the clonal lineage of that cell had gone to fixation.

Note that "fixation time" is defined here as in Equation 4 to mean the number of branching events, rather than absolute time. Our goal was to assess to which extent N varies as a function of the ASC population size (m) and the concentration parameter (κ). To perform these computations, we assumed a wrapped Gauchy distribution underlying Equation 2 (see Methods). We explored a parameter regime, where κ varied over a grid from 0.7 to 0.999 for two different values of m (4 and 40). The lower bound grid value of κ (0.7) was chosen such that at least 90% of density of the wrapped Gauchy was contained within the interval $[0, \pi]$ of the unit circle (see Methods). In other words, precursor sampling was restricted to one half of the SAM to reflect the fact that AM formation is not dispersed around the SAM periphery. Finally, we considered two models for AM initiation (Fig.2A). The first model sampled the location of AM initiation along the SAM periphery from a uniform distribution at each branching iteration, while the second model sampled according to a phylotaxic process in steps of $0.5 \cdot \pi$ (see Theory section).

We found that fixation times are mainly influenced by the concentration parameter κ. For $\kappa = 0.7$, fixation was reached after approximately 12 to 18 nested branching events, but after only ≈ 2 for $\kappa = 0.999$ (Fig.2B,C). Without providing formal proof, it is intuitive that when $\kappa \to 1$, the sampling convergences to a single-cell bottleneck, so that $N \to 1$, as only one clonal sector will be sampled. Interestingly, the impact of m on the fixation time was relatively modest. Its largest effect occurred for small values κ and became gradually less prominent with increasing κ. This observation points at an interesting relationship between m and κ: the impact of ASC population size m becomes negligible as soon as the concentration parameter is large enough to ensure that precursor sampling is restricted to only one of the m SAM clonal sectors. To illustrate this point, we assessed different values of m ranging from 2 to 40, and calculated the corresponding κ such that 90%, 95% and 99% of the total Gauchy density was contained within the angular width, $\Delta\theta = 2\pi \cdot m^{-1}$, for each m (see Methods). As expected, we found that fixation time N becomes nearly independent of m as $\kappa \to 1$ (Fig.2D). Finally, a comparison between the random and the phyllotaxy models did not reveal notable differences for any of the parameter combinations, although we detected slight fixation delays when phyllotaxy was accounted for (Fig.2B,D, Table S1).

*sam*SFS dynamics during repeated nested branching

Recent work in *A. thaliana* suggested that branching is initiated from only one precursor cell of each of the L1, L2 and L3 layers (He et al. (2024)). However, it is currently unknown if this holds generally across species. Our analysis shows that if spatial sampling is more relaxed, the somatic fixation of ASC mutations would require several branching iterations (e.g. Fig.2C). As new mutations arise continuously during ASC self-renewal there will always be some mutations that originate in the internode interval just prior to the latest branching event, or in the one just before that and so on. The result is a mixture of mutation frequencies within the SAM, ranging from older mutations that are already fixed (frequency of 1) to more recent mutations that are present in only a few cells (frequency close

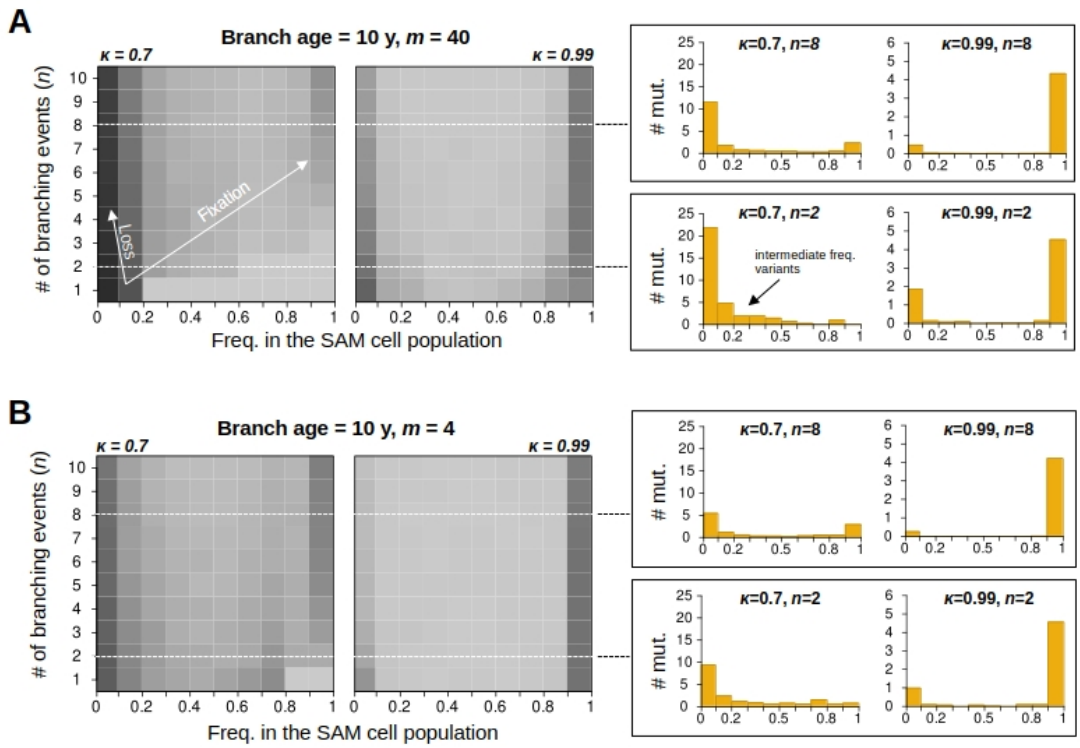

Figure 3: *sam*SFS dynamics during nested branching. Computation of the *sam*SFS assuming a 10 year old branch, a haploid mutation rate of 10^{-9} per bp per year, and a haploid genome size of 240 Mb. **A.** Given a relatively large ASC population size ($m = 40$), shown are the changes in the shape of the *sam*SFS during successive nested branching events under a relaxed and strong drift regime ($\kappa = 0.7$ vs. $\kappa = 0.99$, respectively). **B.** As in A.), but assuming a relatively small ASC population size ($m = 4$). Right panels: For each scenario, the *sam*SFS is highlighted after two ($n = 2$) and after eight branching events ($n = 8$). While drift rapidly generates a U-shaped *sam*SFS in the course of repeated nested branching, mutations do appear at intermediate frequencies in the SAM cell population. This phenomenon is most pronounced in early branching iterations (e.g. $n = 2$), when drift is weak ($\kappa = 0.7$) and the ASC population is large ($m = 40$).

to zero) (Chen et al. (2024)). As described in the Theory section, we refer to the complete frequency distribution as the *sam*SFS.

It is intuitive that the *sam*SFS is shaped by rate of *de novo* mutations (μ) as well as somatic drift (via m and κ). As the number of nested branching iterations (n) becomes large ($n \to \infty$), the *sam*SFS should converge to a mutation-drift equilibrium. However, this limiting case never occurs in reality, because plants initiate relatively few nested branching events during their life-time. In tree species, for instance, the life-time number of nested branching events is usually no more than five or six. Moreover, recent DNA sequencing studies searching for somatic mutations in trees, particularity in fruit trees, have sampled terminal branches that had experienced as few as two nested branching events (e.g. Goel et al. (2024)). It is therefore important to understand the temporal dynamics of the *sam*SFS in finite time; that is, over time-scales that are empirically relevant.

To begin to motivate how the *sam*SFS changes during iterative nested branching, we considered a hypothetical branch that had undergone 1, 2, 3, . . . , 10 equally spaced nested branching events. We assumed that the total age of the branch is 10 years, so that the internode age is e.g. five years when only two branching events occurred, and one year when 10 branching events. This ensured that the total number of *de novo* mutations that arise over time is independent of the number of branching events. Similar to the previous section, we considered two distinct values for the concentration parameter κ (0.7 and 0.99) to reflect relaxed and strong spatial constraint in precursor sampling, respectively, as well two very different values for the size of the ASC population m (4 and 40).

We found that, in the case where the spatial constraint is relatively relaxed ($\kappa = 0.7$) and the ASC population large ($m = 40$), a considerable proportion of somatic variants appear at low to intermediate frequencies in the SAM population, particularly in early branching iterations (Fig.3A). However, the shape of the *sam*SFS changes progressively with increased branching iterations, with low to intermediate frequency variants either becoming fixed or lost due to drift (Fig.3A). A similar, but less pronounced pattern, can be seen when the ASC population is small ($m = 4$). As expected, the situation is different for large κ (e.g. $\kappa = 0.99$). In this case, somatic drift efficiently clears low to intermediate variants, even after only a few nested branching events, and generates a characteristic U-shaped SFS (Fig.3B). That said, low frequency variants do persist in this case, especially when the ASC population is large $m = 40$ and the number of branching events is small.

The above insights reflect recent observations made by genomic studies in trees, which found a wide-range of low and high frequency variants (e.g. Schmitt et al. (2024)), even after layer-specific sequencing was taken into account (e.g. Goel et al. (2024), Amundson et al. (2023)). The utility of the *sam*SFS is that it can be applied to such sequencing data to obtain parameter estimates of m, κ and μ, given that n and the ages of the internode segments is known. In the following section we provide a proof-of-principle demonstration of such an approach.

Estimating the *sam*SFS in an apricot tree

Goel et al. (2024)) recently analyzed somatic mutations in multiple branches of an apricot tree (*Prunus armeniaca cv. Rojo Pasión*) (Fig.4A). In contrast to previous tree studies (Duan et al. (2022), Hanlon et al. (2019), Hofmeister et al. (2020), Wang et al. (2019), Orr et al. (2020), Perez-Roman et al. (2022), Schmid-Siegert et al. (2017), Xie et al. (2016), Satake et al. (2023)), the authors performed DNA sequencing of two distinct tissues that are known to be derived from different SAM cell layers (fruit peel: L1-derived; mesocarp: L2-derived). Their analysis revealed that *de novo* meristematic mutations appear in a layer-specific fashion and that L1 layer mutations are much more frequent than L2-derived mutations, though no mutation rates were reported.

Building on these data, we focused on variant calls of L1-derived mutations from six branches (Fig.4A, Methods). The observed SFS of each of these branches is

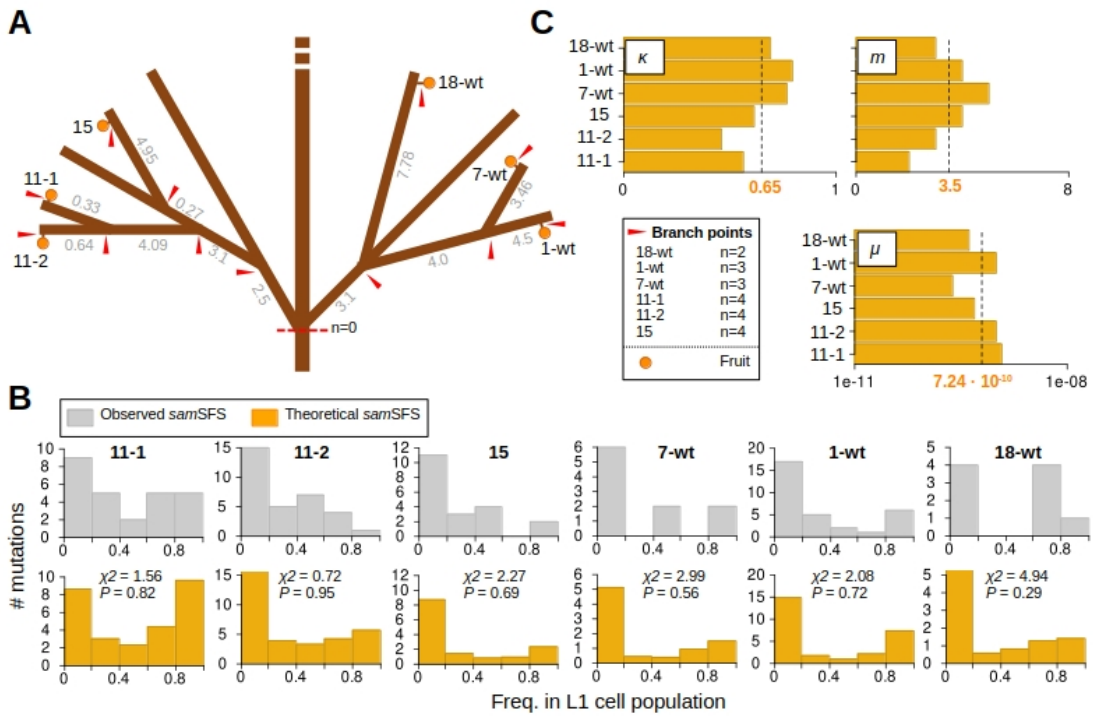

Figure 4: Data application: A. Topology of the apricot tree (*Prunus armeniaca cv. Rojo Pasión*) analyzed by Goel et al. (2024). Layer-specific DNA sequencing was performed for six different branches (11-2, 11-1, 15, 1-wt, 7-wt, 18-wt). Branching bottlenecks are indicated by red arrows and were counted from the tree base ($n = 0$) to the location of each sample on the terminal branches. Gray numbers denote the approximate age (in years) for all internode segments. **B.** Top panel: Observed *sam*SFS; Bottom panel: estimated theoretical *sam*SFS. The random AM placement model provided the best overall fit to the data (Table S3). The goodness-of-fit statistics and their corresponding P values are shown for this model. **C.** Estimates of the concentration parameter (κ), the size of the apical stem cell population (m) and the SAM L1-layer haploid mutation rate per bp per year (μ) for each sample.

displayed in Fig.4B (Method). Our goal was to apply the theoretical *sam*SFS to these data to 1. assess if the model fits the observations, and to 2. obtain statistical estimates of the ASC population size (m), the concentration parameter (κ) and the SAM L1-layer mutation rate (μ). We fitted the *sam*SFS model to each branch separately (see Methods). This allowed us to assess whether the estimates replicate across semi-independent samples. Overall, the best-fitting model assumed a random placement of AMs with each branching iteration (Table S3, see Theory section). According to the Goodness-of-fit statistics, this model did not diverge significantly from the observed SFS on any of the branches (Fig.4B, Table S3). This means that our model agrees well with the data.

Inspection of the parameter estimates of this model revealed a consistently low ASC population size in all samples ($\overline{\hat{m}} = 3.5, SE = 0.17$), a number that is in line with live imaging observations in Arabidopsis and tomato (3-4 ASC cells) Burian et al. (2016)). Moreover, the model estimated a haploid mutation rate of $\overline{\hat{\mu}} \approx 7.24 \cdot 10^{-10}$ ($SE = 8.41 \cdot 10^{-11}$) per bp per year , which corresponds to a

diploid rate of $1.48 \cdot 10^{-9}$. This value is in the range of previous estimates in the Prunus family (e.g. *Prunus persica*: $\mu = 0.82 \cdot 10^{-9}$; *Prunus mume*: $\mu = 2.2 \cdot 10^{-9}$) (Wang et al. (2019)). Finally, the model estimates of the concentration parameter were $\overline{\hat{\kappa}} \approx 0.65$ ($SE = 0.028$). This value is somewhat low, and suggests precursor sampling is not entirely restricted to a single SAM clonal sector, on average. However, a technical caveat with the estimates of κ is that both high- and low-frequency variants are likely undersampled in the post-processed sequencing data. Under-sampling of high-frequency variants could stem from layer contamination during tissue dissection. Under-sampling of low frequency mutations, on the other hand, may be due to stringent filtering of rare variant calls that resemble sequencing noise. The net effect is that the concentration parameter is biased downward, so that it mimics a relaxed drift scenario. It should be possible to correct these issues in future empirical studies, for example by using layer-specific marker genes in combination with cell sorting prior to sequencing, and by fine-tuning pre-processing thresholds for variant calling.

Despite these caveats, our model agrees well with the observed data, and parameter estimates for μ and m are consistent with those reported in other species. Hence, a statistical analysis of the *sam*SFS in combination with DNA sequencing data appears to be feasible methodological approach, which can yield estimates of meristematic parameters that are difficult to obtain with other methods.

Conclusion and outlook

Here we developed a theoretical model that describes how the frequency of *de novo* mutations in apical stem cells (ASCs) evolve in the SAM cell population due to neutral somatic drift during repeated branching. The process is a function of the ASC population size (m), the strength of the cellular bottleneck during branch initiation (κ), and the stem cell mutation rate (μ). Using the model, we derived a method for computing the mutation site frequency spectrum of the SAM (*sam*SFS), and showed that it can be applied to layer-specific DNA sequencing data for parameter estimation. This is a useful result because there are currently no alternative methods for measuring these parameters *in vivo*, and certainly not simultaneously.

Perhaps the most important result is that our model agrees well with the observed *sam*SFS, as seen in our proof-of-prinicple data application. This fact implies that the neutral mutation-drift assumption is sufficient to explain the data, and that additional processes such as layer-exchange and/or cell lineage selection may be less relevant, although they have featured heavily in several theoretical models (e.g. Iwasa et al. (2023), Klekowski et al. (1985), Pineda-Krch and Lehtila (2002)). That said, statistical inference may have been somewhat underpowered in the apricot data due to the small number of somatic mutations (i.e. observations). Subtle divergences from our model assumptions may have therefore gone undetected. One strategy to increase statistical power in future applications is to aggregate somatic mutations from multiple branches. A second, perhaps complimentary,

strategy is to use spontaneous somatic epimutations, instead of mutations, as input to the model. Spontaneous epimutations refer to mitotically heritable CG methylation changes that arise stochastically during cell divisions (Johannes and Schmitz (2019)). Because their rate of accumulation is several orders of magnitude higher than the somatic genetic mutation rate (van der Graaf et al. (2015), Hazarika et al. (2022), Denkena et al. (2021), Schmitz et al. (2011), Becker et al. (2011)), they would yield more "observations" (per unit time) for more robust model inference (Yao et al. (2021), Yao et al. (2023)).

Although our data application was restricted to apricot, the method has clear applications beyond trees. It can be applied, generally, to any plant species that uses branching as a developmental or reproductive strategy. This also includes species that propagate clonally via ramets, suckers, tillers, etc. Examples include clonal populations of the seagrass (*Zostera marina*) (Yu et al. (2019)), or the large Aspen (*Populus tremuloides*) clones of the Pando forest (Pineau et al. (2024)). Since our model assumes a neutral mutation-drift processes, it could serve as a starting point to detect evidence for cell lineage, module or ramet selection in natural settings. Possible model extensions could include the direct incorporation of a selection parameter, and the possibility of L1/L2/L3 stem cell layer-exchange, as these processes may be more important over long time-scales of clonal evolution.

We expect that further tests of our model on a broader range of plant species and data modalities will lead to further refinements. Such efforts promise to accelerate our basic understanding of the somatic evolution of stem cell mutations in long-lived plants.

Materials and Methods

Implementation of the wrapped Cauchy distribution

To implement the theoretical models described in the main text, we needed to assume a specific form for the underlying wrapped distribution, $f(\theta; \nu, \kappa)$. A flexible choice for modeling precursor sampling is the wrapped Gauchy distribution. Its probability density function (PDF) is given by:

$$f(\theta \mid \nu, \kappa) = \frac{1 - \kappa^2}{2\pi \left(1 + \kappa^2 - 2\kappa \cos(\theta - \nu)\right)},$$

where $\kappa \in [0, 1)$. The bounds on κ facilitate an easy interpretation of the strength of the bottleneck during branching.

Exploratory analysis of fixation times

To explore fixation times, we restricted the analysis to a realistic parameter range of values of κ such that 90% of the cumulative probability density of the wrapped distribution is contained within one-half of the unit circle. This constraint reflects microscopy studies showing that AM initiation and precursor sampling are

localized phenomena, occurring in "narrow" regions along the SAM periphery. To determine the minimum κ value that meets this condition, we computed the cumulative probability over the interval $[-\pi/2, \pi/2]$ as a function of κ. Specifically, the cumulative probability for a given κ was calculated by integrating the probability density function (PDF) of the wrapped Cauchy distribution around the mode ν:

$$F(\kappa) = \int_{\nu-\frac{\pi}{2}}^{\nu+\frac{\pi}{2}} \frac{1-\kappa^2}{2\pi\left(1+\kappa^2-2\kappa\cos(\theta-\nu)\right)}\, d\theta.$$

This computation was repeated over a grid of increasing κ values until the condition $F(\kappa) \geq 0.90$ was met. Based on the wrapped Cauchy distribution, the minimum κ value that satisfies this requirement is approximately 0.7. We used a similar approach to generate Fig.2B. There, the goal was to find the minimum value of κ corresponding to different m, so that the above probability criterion was met in each case. Hence, for a given m, integration was over the interval $[\nu - \pi/m, \nu + \pi/m]$.

*sam*SFS analysis in apricot

We reanalyzed data from the study of Goel et al. (2024), who sequenced fruits from six different branches (11-1, 11-2, 15, 7-wt, 1-wt, 18-wt) of an apricot tree. A unique aspect of this study is that, prior to sequencing, the authors dissected apricot fruits into two distinct tissues that are known to be derived from different SAM cell layers (fruit peel: L1-derived; mesocarp: L2-derived). For each branch, we obtained a dataset of *de novo* mutations (Table S2). To guard against variant over-filtering, we worked with a rawer version of the final dataset the authors used in their original publication. To obtain this data, the following filtering steps were applied: Reads were trimmed, aligned, and deduplicated. Positions with at least three reads supporting the variant allele and displaying mapping quality ≥10 and base quality ≥30 were retained. Further positions were removed if they: were heterozygous between the two haplotypes of the assembly, had too low/high read depth in either L1 or L2 sample, or did not have sufficient variant allele frequency (VAF) variation between the layers, represented positions that were mutated in all of the leaf samples. Since apricot is diploid the VAF values of each mutation were multiplied by two to obtain a proxy of the proportion of cells carrying the mutation in a given bulk sample. Adjusted VAF values >1 were set to 1.

To construct the observed *sam*SFS, we divided the vector of VAFs into $u = 5$ equally-sized intervals over the range $[0, 1]$ and counted the values in each interval. The interval boundaries $[d_1, d_2, \ldots, d_{u+1}]$ were defined, with d_1 set at $\frac{1}{100}$. For each interval $[d_q, d_{q+1})$, we counted the VAFs, v_r, that satisfied $d_q \leq v_r < d_{q+1}$. These counts formed a vector, representing the observed SFS (Fig.4B). To be able to fit the theoretical *sam*SFS to each of the samples, we needed information about the parameter n and as well as the internode time intervals $t_1, \ldots, t_n$. Starting from the observed tree topology, we determined n for each sample, by counting the number of bottlenecks along a branching path from the tree base to the location

of that sample. The tree used in this study was 11 years old. However, the ages of the internode segments was not available, but only provided in terms of "meters of growth". We therefore converted the internode segments to approximate time intervals $t_1, \ldots, t_n$ (in years). To do this, we assumed a constant growth rate and expressed the meter-length of each segment as a proportion of the total meter-length of the branching path from the tree base to the sampled fruits. These proportions were then multiplied by the total tree age to obtain the approximate ages of each segment in terms of years.

We fitted the theoretical *sam*SFS to the observed SFS by minimizing the Goodness-of-fit statistic shown in Equation 5. Minimization was carried out using a grid search over parameters μ (grid range: 10^{-11} to 10^{-8}), κ (grid range: 0 to 1) and m (grid range: 1 to 14). We considered three different *sam*SFS models. The first model assumed that the location AM initiation at the SAM periphery is randomly chosen, at each branching iteration n, from a uniform distribution on the unit circle (Model: random-each-iter in Table S3). The the other two models accounted for phylotaxy, and used Equation 3 to rotate AMs initiation in deterministic steps of π (Model: fixed-start-rotate-pi in Table S3) or steps of $0.5 \cdot \pi$ (Model: fixed-start-rotate-half-pi in Table S3) at each branching iteration. The model fits as well as the parameter estimates are provided in Table S. The estimation procedure was implemented in *R* (see gitcode).

References

Amundson, K. R., Marimuthu, M. P. A., Nguyen, O., Sarika, K., DeMarco, I. J., Phan, A., Henry, I. M., and Comai, L. (2023). Differential mutation rates in plant meristematic layers. *bioRxiv*, page 2023.09.25.559363.

Antolin, M. F. and Strobeck, C. (1985). The population genetics of somatic mutation in plants. *Am. Nat.*, 126(1):52–62.

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, 480(7376):245–249.

Bradamante, G. (2024). Live imaging of the shoot apical meristem of intact, soil-grown, flowering arabidopsis plants. *Bio Protoc.*, 14(12):e5015.

Burian, A. (2021). Does shoot apical meristem function as the germline in safeguarding against excess of mutations? *Front. Plant Sci.*, 12:707740.

Burian, A., Barbier de Reuille, P., and Kuhlemeier, C. (2016). Patterns of Stem Cell Divisions Contribute to Plant Longevity. *Curr. Biol.*, 26(11):1385–1394.

Chen, Y., Burian, A., and Johannes, F. (2024). Somatic epigenetic drift during shoot branching: a cell lineage-based model. *Genetics*, page iyae091.

De La Torre, A. R., Li, Z., Van de Peer, Y., and Ingvarsson, P. K. (2017). Contrast-

ing rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Biol. Evol.*, 34(6):1363–1377.

Denkena, J., Johannes, F., and Colomé-Tatché, M. (2021). Region-level epimutation rates in arabidopsis thaliana. *Heredity*, pages 1–13.

Duan, Y., Yan, J., Zhu, Y., Zhang, C., Tao, X., Ji, H., Zhang, M., Wang, X., and Wang, L. (2022). Limited accumulation of high-frequency somatic mutations in a 1700-year-old osmanthus fragrans tree. *Tree Physiol.*, 42(10):2040–2049.

Eyre-Walker, A. and Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nat. Rev. Genet.*, 8(8):610–618.

Folse, 3rd, H. J. and Roughgarden, J. (2012). Direct benefits of genetic mosaicism and intraorganismal selection: modeling coevolution between a long-lived tree and a short-lived herbivore: Direct benefits of genetic mosaicism. *Evolution*, 66(4):1091–1113.

Foster, T. M. and Aranzana, M. J. (2018). Attention sports fans! the far-reaching contributions of bud sport mutants to horticulture and plant biology. *Hortic Res*, 5:44.

Goel, M., Campoy, J. A., Krause, K., Baus, L. C., Sahu, A., Sun, H., Walkemeier, B., Marek, M., Beaudry, R., Ruiz, D., Huettel, B., and Schneeberger, K. (2024). The vast majority of somatic mutations in plants are layer-specific. *Genome Biol.*, 25(1):194.

Hanlon, V. C. T., Otto, S. P., and Aitken, S. N. (2019). Somatic mutations substantially increase the per-generation mutation rate in the conifer picea sitchensis. *Evol Lett*, 3(4):348–358.

Hazarika, R. R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R. J., and Johannes, F. (2022). Molecular properties of epimutation hotspots. *Nat Plants*, 8(2):146–156.

He, X., Xia, F.-N., Liu, K., Wang, J., Liu, Z., and Li, A. (2024). Mapping zygote-to-adult developmental cell phylogeny in arabidopsis thaliana reveals a 3-cell rule of branching. *Research Square*.

Heidstra, R. and Sabatini, S. (2014). Plant and animal stem cells: similar yet different. *Nat. Rev. Mol. Cell Biol.*, 15(5):301–312.

Hofmeister, B. T., Denkena, J., Colomé-Tatché, M., Shahryary, Y., Hazarika, R., Grimwood, J., Mamidi, S., Jenkins, J., Grabowski, P. P., Sreedasyam, A., Shu, S., Barry, K., Lail, K., Adam, C., Lipzen, A., Sorek, R., Kudrna, D., Talag, J., Wing, R., Hall, D. W., Jacobsen, D., Tuskan, G. A., Schmutz, J., Johannes, F., and Schmitz, R. J. (2020). A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial Populus trichocarpa. *Genome Biol.*, 21(1):259.

Iwasa, Y., Tomimoto, S., and Satake, A. (2023). The genetic structure within

a single tree is determined by the behavior of the stem cells in the meristem. *Genetics*, 223(4).

Johannes, F. (2024). Allometric scaling of somatic mutation and epimutation rates in trees. *Evolution*, 79(1):1–5.

Johannes, F. and Schmitz, R. J. (2019). Spontaneous epimutations in plants. *New Phytol.*, 221(3):1253–1259.

Kato, S. and Jones, M. C. (2013). An extended family of circular distributions related to wrapped cauchy distributions via brownian motion. *bj*, 19(1):154–171.

Klekowski, join(' ', E. J. and Kazarinova-Fukshansky, N. (1984a). Shoot apical meristems and mutation: Selective loss of disadvantageous cell genotypes. *Am. J. Bot.*, 71(1):28.

Klekowski, Jr, E. J. and Kazarinova-Fukshansky, N. (1984b). Shoot apical meristems and mutation: Fixation of selectively neutral cell genotypes. *Am. J. Bot.*, 71(1):22–27.

Klekowski, Jr, E. J., Kazarinova-Fukshansky, N., and Mohr, H. (1985). Shoot apical meristems and mutation: Stratified meristems and angiosperm evolution. *Am. J. Bot.*, 72(11):1788–1800.

Kuhlemeier, C. (2017). Phyllotaxis. *Curr. Biol.*, 27(17):R882–R887.

Kwiatkowska, D. (2008). Flowering and apical meristem growth dynamics. *J. Exp. Bot.*, 59(2):187–201.

Lanfear, R., Ho, S. Y. W., Jonathan Davies, T., Moles, A. T., Aarssen, L., Swenson, N. G., Warman, L., Zanne, A. E., and Allen, A. P. (2013). Taller plants have lower rates of molecular evolution. *Nat. Commun.*, 4(1):1879.

Lesaffre, T. (2021). Population-level consequences of inheritable somatic mutations and the evolution of mutation rates in plants. *Proc. Biol. Sci.*, 288(1958):20211127.

Lu, X., Zhang, Q., Wang, Z., Cheng, X., Yan, H., Cai, S., Zhang, H., and Liu, Q. (2024). Development of an inducible DNA barcoding system to understand lineage changes in arabidopsis regeneration. *Dev. Cell*, 0(0).

Lyndon, R. F. (1998). *Developmental and cell biology series: The shoot apical meristem: Its growth and development series number 34*. Developmental and cell biology series. Cambridge University Press, Cambridge, England.

Nicolas, A. and Laufs, P. (2022). Meristem initiation and de novo stem cell formation. *Front. Plant Sci.*, 13:891228.

Orr, A. J., Padovan, A., Kainer, D., Külheim, C., Bromham, L., Bustos-Segura, C., Foley, W., Haff, T., Hsieh, J.-F., Morales-Suarez, A., Cartwright, R. A., and Lanfear, R. (2020). A phylogenomic approach reveals a low somatic mutation

rate in a long-lived plant. *Proceedings of the Royal Society B: Biological Sciences*, 287(1922).

Otto, S. P. and Orive, M. E. (1995). Evolutionary consequences of mutation and selection within an individual. *Genetics*, 141(3):1173–1187.

Padovan, A., Keszei, A., Foley, W. J., and Külheim, C. (2013). Differences in gene expression within a striking phenotypic mosaic Eucalyptus tree that varies in susceptibility to herbivory. *BMC Plant Biol.*, 13:29.

Perez-Roman, E., Borredá, C., López-García Usach, A., and Talon, M. (2022). Single-nucleotide mosaicism in citrus: Estimations of somatic mutation rates and total number of variants. *Plant Genome*, 15(1):e20162.

Petit, R. and Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.*, 37:187–214.

Pineau, R. M., Mock, K. E., Morris, J., Kraklow, V., Brunelle, A., Pageot, A., Ratcliff, W. C., and Gompert, Z. (2024). Mosaic of somatic mutations in earth's oldest living organism, pando. *bioRxivorg*, page 2024.10.19.619233.

Pineda-Krch, M. and Fagerstrom, T. (1999). On the potential for evolutionary change in meristematic cell lineages through intraorganismal selection. *J. Evol. Biol.*, 12(4):681–688.

Pineda-Krch, M. and Lehtila, K. (2002). Cell lineage dynamics in stratified shoot apical meristems. *J. Theor. Biol.*, 219(4):495–505.

Reusch, T. B. H., Baums, I. B., and Werner, B. (2021). Evolution via somatic genetic variation in modular species. *Trends Ecol. Evol.*, 36(12):1083–1092.

Satake, A., Imai, R., Fujino, T., Tomimoto, S., Ohta, K., Na'iem, M., Indrioko, S., Widiyatno, Purnomo, S., Mollá-Morales, A., Nizhynska, V., Tani, N., Suyama, Y., Sasaki, E., and Kasahara, M. (2023). Somatic mutation rates scale with time not growth rate in long-lived tropical trees. *eLife*.

Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., Cattaneo, P., Schütz, F., Farinelli, L., Pagni, M., Schneider, M., Voumard, J., Jaboyedoff, M., Fankhauser, C., Hardtke, C. S., Keller, L., Pannell, J. R., Reymond, A., Robinson-Rechavi, M., Xenarios, I., and Reymond, P. (2017). Low number of fixed somatic mutations in a long-lived oak tree. *Nature Plants*, 3(12):926.

Schmitt, S., Heuret, P., Troispoux, V., Beraud, M., Cazal, J., Chancerel, , Cravero, C., Guichoux, E., Lepais, O., Loureiro, J., Marande, W., Martin-Ducup, O., Vincent, G., Chave, J., Plomion, C., Leroy, T., Heuertz, M., and Tysklind, N. (2024). Low-frequency somatic mutations are heritable in tropical trees dicorynia guianensis and sextonia rubra. *Proc. Natl. Acad. Sci. U. S. A.*, 121(10):e2313312121.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A.,

Libiger, O., Schork, N. J., and Ecker, J. R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, 334(6054):369–373.

Schnablová, R., Herben, T., and Klimešová, J. (2017). Shoot apical meristem and plant body organization: a cross-species comparative study. *Ann. Bot.*, 120(5):833–843.

Schoen, D. J. and Schultz, S. T. (2019). Somatic mutation and evolution in plants. *Annu. Rev. Ecol. Evol. Syst.*, 50(Volume 50, 2019):49–73.

Scofield, D. G. and Schultz, S. T. (2006). Mitosis, stature and evolution of plant mating systems: low-phi and high-phi plants. *Proc. Biol. Sci.*, 273(1584):275–282.

Shahryary, Y., Symeonidi, A., Hazarika, R. R., Denkena, J., Mubeen, T., Hofmeister, B., van Gurp, T., Colomé-Tatché, M., Verhoeven, K. J. F., Tuskan, G., Schmitz, R. J., and Johannes, F. (2020). AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants. *Genome Biol.*, 21(1):260.

Shi, B., Zhang, C., Tian, C., Wang, J., Wang, Q., Xu, T., Xu, Y., Ohno, C., Sablowski, R., Heisler, M. G., Theres, K., Wang, Y., and Jiao, Y. (2016). Two-step regulation of a meristematic cell population acting in shoot branching in arabidopsis. *PLoS Genet.*, 12(7):e1006168.

Tomimoto, S. and Satake, A. (2023). Modelling somatic mutation accumulation and expansion in a long-lived tree with hierarchical modular architecture. *J. Theor. Biol.*, 565:111465.

van der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., Schmitz, R. J., Colomé-Tatché, M., and Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, 112(21):6676–6681.

Wang, L., Ji, Y., Hu, Y., Hu, H., Jia, X., Jiang, M., Zhang, X., Zhao, L., Zhang, Y., Jia, Y., Qin, C., Yu, L., Huang, J., Yang, S., Hurst, L. D., and Tian, D. (2019). The architecture of intra-organism mutation rate variation in plants. *PLoS Biol.*, 17(4):e3000191.

Wang, Y. and Jiao, Y. (2018). Axillary meristem initiation—a way to branch out. *Curr. Opin. Plant Biol.*, 41:61–66.

Xie, Z., Wang, L., Wang, L., Wang, Z., Lu, Z., Tian, D., Yang, S., and Hurst, L. D. (2016). Mutation rate analysis via parent-progeny sequencing of the perennial peach. I. a low rate in woody perennials and a higher mutagenicity in hybrids. *Proc. Biol. Sci.*, 283(1841).

Yao, N., Schmitz, R. J., and Johannes, F. (2021). Epimutations define a fast-ticking molecular clock in plants. *Trends Genet.*

Yao, N., Zhang, Z., Yu, L., Hazarika, R., Yu, C., Jang, H., Smith, L. M., Ton,

J., Liu, L., Stachowicz, J. J., Reusch, T. B. H., Schmitz, R. J., and Johannes, F. (2023). An evolutionary epigenetic clock in plants. *Science*, 381(6665):1440–1445.

Yu, L., Boström, C., Franzenburg, S., Bayer, T., Dagan, T., and Reusch, T. B. H. (2019). Somatic genetic drift and multi-level selection in modular species. *bioRxiv*, page 833335.

Yu, L., Boström, C., Franzenburg, S., Bayer, T., Dagan, T., and Reusch, T. B. H. (2020). Somatic genetic drift and multilevel selection in a clonal seagrass. *Nat Ecol Evol*, 4(7):952–962.

Yu, L., Renton, J., Burian, A., Khachaturyan, M., Bayer, T., Kotta, J., Stachowicz, J. J., DuBois, K., Baums, I. B., Werner, B., and Reusch, T. B. H. (2024). A somatic genetic clock for clonal species. *Nat. Ecol. Evol.*, 8(7):1327–1336.

Acknowledgments

I thank Korbinian Schneeberger and Manish Goel for sharing and explaining their semi-filtered datasets, and Agata Burian for her quick feedback on a draft version of this manuscript.