

# Somatic Mutation Analysis in *Salix suchowensis* Reveals Early-Segregated Cell Lineages

Yifan Ren,<sup>†,1</sup> Zhen He,<sup>†,1</sup> Pingyu Liu,<sup>1</sup> Brian Traw,<sup>1</sup> Shucun Sun,<sup>2</sup> Dacheng Tian,<sup>1</sup> Sihai Yang,<sup>1</sup> Yanxiao Jia,<sup>\*,3</sup> and Long Wang<sup>ID\*,1</sup>

<sup>1</sup>State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China

<sup>2</sup>Department of Ecology, School of Life Science, Nanjing University, Nanjing, China

<sup>3</sup>State Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China

<sup>†</sup>These authors contributed equally to this work.

\*Corresponding authors: E-mails: wanglong@nju.edu.cn; jiayanxiao@163.com.

Associate editor: Gregory Wray

## Abstract

Long-lived plants face the challenge of ever-increasing mutational burden across their long lifespan. Early sequestration of meristematic stem cells is supposed to efficiently slow down this process, but direct measurement of somatic mutations that accompanies segregated cell lineages in plants is still rare. Here, we tracked somatic mutations in 33 leaves and 22 adventitious roots from 22 stem-cuttings across eight major branches of a shrub willow (*Salix suchowensis*). We found that most mutations propagated separately in leaves and roots, providing clear evidence for early segregation of underlying cell lineages. By combining lineage tracking with allele frequency analysis, our results revealed a set of mutations shared by distinct branches, but were exclusively present in leaves and not in roots. These mutations were likely propagated by rapidly dividing somatic cell lineages which survive several iterations of branching, distinct from the slowly dividing axillary stem cell lineages. Leaf is thus contributed by both slowly and rapidly dividing cell lineages, leading to varied fixation chances of propagated mutations. By contrast, each root likely arises from a single founder cell within the adventitious stem cell lineages. Our findings give straightforward evidence that early segregation of meristems slows down mutation accumulation in axillary meristems, implying a plant “germline” paralog to the germline of animals through convergent evolution.

**Key words:** plant evolution, clonal variation, somatic mutation, *Salix* mutation, plant cell lineage.

## Introduction

Somatic mutations may arise every time a cell divides, and be passed to the next generation if the cell harboring the mutation becomes a germ cell. In animals, somatic mutations usually have degenerative effects and are associated with disease and aging (Zhang and Vijg 2018). Owing to the early segregation of germlines in most animals, the later developed somatic mutations have no chance to enter the germline (the germ-plasm theory [Weismann 1892]). In plants where germline differentiates late, somatic mutations are supposed to act as an important source of innovation for plant evolution (Lanfear 2018; Plomion et al. 2018). They are frequently used as source of genetic material in artificial breeding programs to create bud sport mutants among clonal descendants (Benedict 1923; Shamel and Pomeroy 1936; Roest et al. 1981; Foster and Aranzana 2018). Assuming ever-lasting accumulation and fixation of somatic mutations in long-lived plants, the intraorganismal hypothesis supposes that they are able to maximize the within-plant heterogeneity, allowing intraorganismal selection, and providing opportunities for plants to outmaneuver enemies or adapt to changing environments (Whitham and Slobodchikoff 1981; Michel et al.

2004; Simberloff and Leppanen 2019). This hypothesis received considerable attention in the past (Whitham and Slobodchikoff 1981; Whitham 1983; Antolin and Strobeck 1985; Suomela and Ayres 1994; Simberloff and Leppanen 2019).

By contrast, recent studies have focused more on the precise detection of the plant somatic mutations as well as their potential inheritance (Watson et al. 2016; Schmid-Siegert et al. 2017; Plomion et al. 2018; Hanlon et al. 2019; Hofmeister et al. 2020; Wang et al. 2019; Orr et al. 2020; Yu et al. 2020). Through measuring fixed somatic mutations in terminal leaves from trees with lifespans of several centuries (Schmid-Siegert et al. 2017; Plomion et al. 2018; Hanlon et al. 2019; Hofmeister et al. 2020; Wang et al. 2019; Orr et al. 2020), these studies found that old trees only accumulate very few somatic mutations, which is much lower than what has been conjectured before (Sutherland and Watkinson 1986; Klekowski Jr and Godfrey 1989; O’Connell and Ritland 2004; Yong 2012; Diwan et al. 2014). The low number of fixed mutations in old trees challenges the assumption of the intraorganismal hypothesis, while implying an efficient strategy in plants to keep fixable mutations in check. Using time-lapse

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

imaging and computational modeling, Burian et al. (2016) showed that this was likely achieved by setting aside the axillary meristems early and thus reducing the number of stem cell divisions in shoot apical meristem (SAM) during plant development (Groot and Laux 2016), a system analogous to the formation of germline in animals.

These novel findings emphasized the importance of timing of segregation for certain stem cell lineages, such as the plant germline. There is a long-lasting debate on whether plants have a segregated germline in determining the heritable somatic mutations (see Lanfear's recent review [Lanfear 2018] on the historical background of related arguments). A late-segregated germline predicts that most somatic mutations are heritable, whereas an early-segregated germline will predict the contrary (Lanfear 2018). The differences in timing of germline (or more broadly, stem cell lineages) segregation therefore create strong differences in the ability of these mutations to fuel genetic variation within populations and therefore contribute to evolution (Lanfear 2018; Plomion et al. 2018). However, whether the early segregation of the stem cell lineage suggested by Burian et al. (2016) applies to woody plants has yet to be confirmed. More importantly, despite the achievements in measuring plant mutations, direct quantification of mutations for specified cell lineages remains elusive in perennial woody plants.

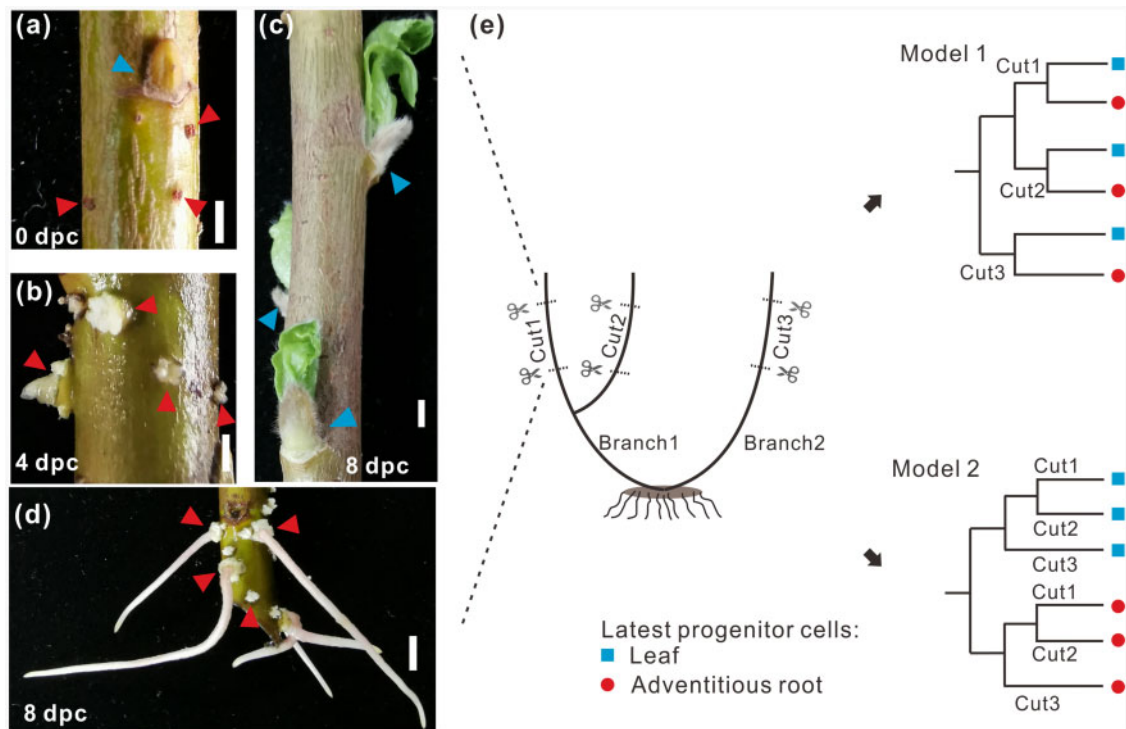
Assessing the timing of segregation is partly viable through careful tracking of different cell lineages (Poethig 1989; Irish 1991). A wealth of information on tissue identity formation in plant development has come from studies of plant chimeras (for reviews, see Poethig 1987, 1989; Szymkowiak and Sussex 1996; Frank and Chitwood 2016). Experiments using periclinal and mericlinal chimeras have revealed in angiosperms that the SAM is organized into three layers, termed as L1, L2, L3, where each layer generally forms epidermal, subepidermal, and internal tissues, respectively (Irish 1991; Szymkowiak and Sussex 1996). The cell fates in each layer are, however, thought to be largely determined by physical location rather than past lineage history (Irish 1991; Szymkowiak and Sussex 1996). For example, gametes are usually observed to arise from L2 cells, but they can also be formed by L1 and L3 cells (Szymkowiak and Sussex 1996). The adaptive significance of this stratified structure has been discussed through mathematical modeling (Klekowski and Kazarinova-Fukshansky 1984a, 1984b; Klekowski et al. 1985). With improved excision and imaging techniques, recent works have provided further insights on the gene regulatory programs within the meristems (Wang et al. 2018; Kitagawa and Jackson 2019), especially regarding their regeneration capability (Sena et al. 2009; Rahni et al. 2016; Ikeuchi et al. 2019).

Mutations are widely recognized as valuable molecular markers in these works for cell lineage tracing. Whereas traditional clonal analyses used induced mutations to mark cell lineages (Poethig 1987), the use of irradiation may itself change cell behavior and was difficult to control (Poethig 1987). This drawback could be overcome by directly tracking *de novo* somatic mutations through whole-genome

sequencing (WGS) of multiple parts of a plant. In our prior study (Wang et al. 2019), we confirmed the usability of this approach by tracking somatic mutations in the woodland strawberry. We found that several mutations were restricted to runners and were never been passed to daughter plants (Wang et al. 2019), giving further evidence for the segregation of stem cell lineages.

Here, we interrogate somatic mutations to test for segregated stem cell lineages in a perennial woody plant, the shrub willow (*Salix suchowensis* spp.). Shrub willow has strong regeneration ability and is an important crop for both bioenergy and environmental engineering (Volk et al. 2004). Like many willows (Carlson 1938; Sennerby-Forsse and Zsuffa 1995), shrub willow has strong resprouting ability, and its stem-cuttings root very readily when cultivated in favorable conditions. Histological analysis has revealed that most stem buds in willows are of axillary origin (Fink 1983; Sennerby-Forsse and Zsuffa 1995), whereas root buds are adventitious in origin (Carlson 1938; 1950; Haissig 1970; Fjell 1985, 1987; Sennerby-Forsse and Zsuffa 1995). Axillary buds/meristems are developed in or near the leaf axils (Wang and Jiao 2018), and are presumably directly derived from the SAM with possible minimized number of cell divisions (Burian et al. 2016). By contrast, aboveground adventitious buds/meristems most often initiate from cells neighboring vascular tissues (Bellini et al. 2014), and are likely derived from vascular cambial cell divisions (Steffens and Rasmussen 2016) or by reprogramming differentiated somatic cells (Díaz-Sala 2014). Though both meristems originate from the SAM (Lucas et al. 2013; De Rybel et al. 2016), the proliferation distance from apical stem cells and cell division patterns leading to the formation of them are likely to be distinct. However, neither the relationship of their underlying cell lineages nor the patterns of corresponded somatic mutations have been characterized previously in *Salix*, to the best of our knowledge.

We consider two models (fig. 1) that differ in whether meristematic stem cell lineages segregate late (model 1) or early (model 2) when forming the leaves and the adventitious roots. In model 1, the leaves and adventitious roots share same progenitor cell lineages until later differentiation. In this model, samples collected from the same branch or same stem-cutting have closer genealogy (fig. 1e) and somatic mutations that predate the formation of this branch will be propagated to both organs. In model 2, the leaves and adventitious roots have distinct progenitor cell lineages from the very beginning. In this model, the genealogy of two organs is not correlated with their physical locations (fig. 1e), and somatic mutations are propagated separately by these cell lineages. The models are tested by tracing somatic mutations in multiple stem-cuttings of the same willow, with shoot/leaves represent for organs differentiated from axillary meristems and adventitious roots represent for organs differentiated from adventitious meristems (fig. 1a–d). Our aim is to provide a landscape of somatic mutations in shrub willow and utilize these molecular markers to assess the



**Fig. 1.** Sampling strategies considering two possible segregation models. (a–d) Photos showing the two different meristem types. New leaves formed by axillary buds/meristems (blue arrows) and can be viewed around 6–8 days post cut (dpc), whereas new roots formed by adventitious buds/meristems (red arrows) can be viewed as early as 4 days post cut. For each cutting, three leaves from independent axillary buds (photo c) and three roots from independent adventitious buds (photo d) are sampled. Scale bars, 0.2 cm for a–c and 0.5 cm for d. (e) Conjectured cell lineage segregation models with expected presence of shared mutations. Model 1: late-segregated cell lineages in generating terminal leaves and adventitious roots. Under model 1, organs closer in physical locations tend to have closer genealogy therefore more mutations arise predate their formation. Model 2: the early segregation model. Under model 2, the physical locations of different organs do not predict their truly genealogy, and mutations are separately transmitted by segregated cell lineage to each organ.

possibility of any early-segregated stem cell lineages. Our results will contribute to the understanding of the role of somatic mutations in plant evolution in a more general sense than the intraorganismal hypothesis has conjectured.

## Results

### Sampling and Sequencing of the Genomes of a Shrub Willow

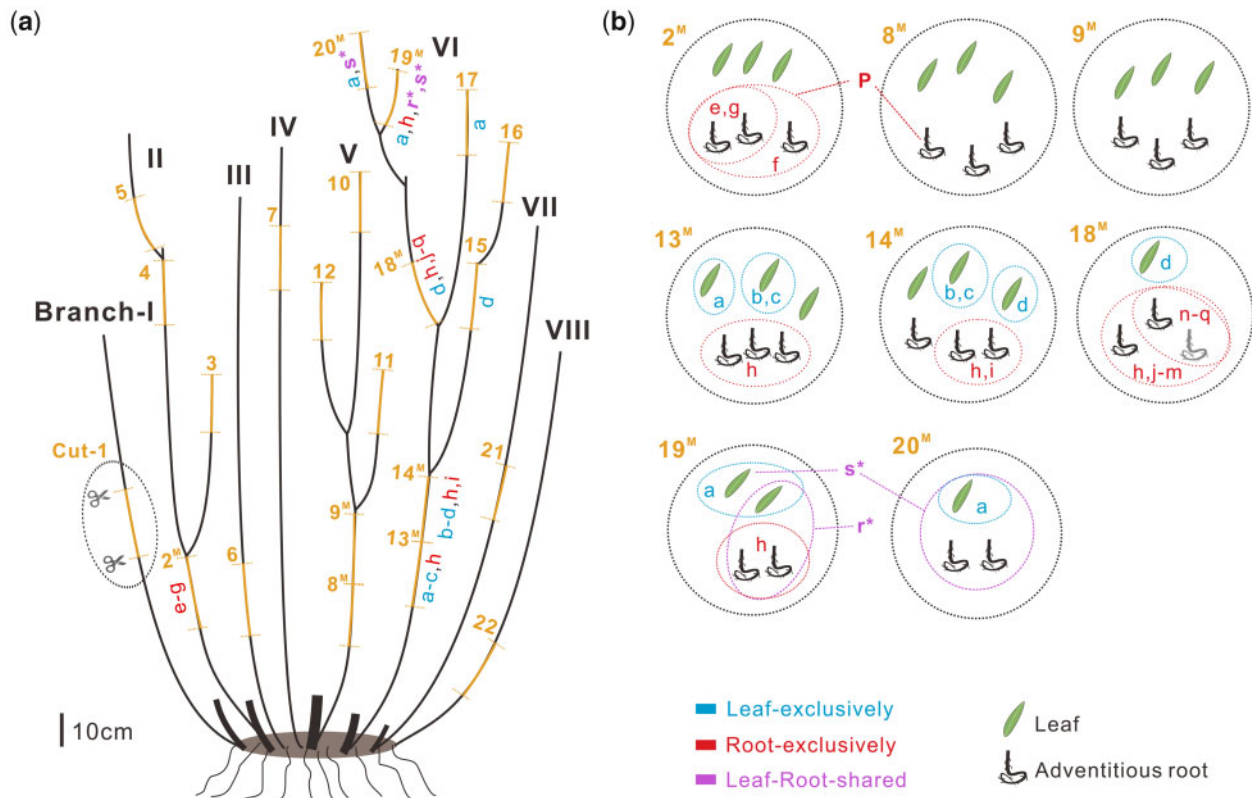
Original stem-cuttings were collected from the shrub willow (*S. suchowensis*, individual ID “YAF1”) in 2016 after new shoots regenerated from its trunk which was copied in 2015 (supplementary fig. S1, Supplementary Material online). When each cutting was made, new shoots/leaves were then generated from preformed axillary buds/meristems (fig. 1a and c), whereas adventitious roots were regenerated from preformed adventitious buds/meristems (fig. 1a, b, and d). The underlying cell lineages forming leaves and roots could either segregate late upon differentiation, or at an early stage, corresponding to our proposed segregation models #1 and #2, respectively (fig. 1e). Under model 1 in which no meristem is segregated until differentiation, mutations prior to the segregation are expected to have the same chance of

fixation in samples from both leaves and roots. In contrast, support for model 2 will be found if the fixation of mutations is mostly independent between leaves and roots.

To test between the models, we investigated somatic mutations in the 22 stem-cuttings (“Cut-1~22” in fig. 2a) from eight branches (I~VIII in fig. 2a). For 8 of 22 cuttings, each cutting was sequenced for up to three independent leaves, each from a separate axillary bud, and up to three independent adventitious roots, each from a separate adventitious bud (figs. 1a–d and 2b), yielding a total of 19 leaves and 22 adventitious roots from those eight cuttings (supplementary table S1, Supplementary Material online). For each of the remaining 14 cuttings, a single leaf sample was collected for accurate tracing of the origins of mutations.

In total, 33 leaves and 22 adventitious roots of the same shrub willow were whole-genome sequenced (supplementary table S1, Supplementary Material online), yielding over 2,000-fold sequencing depth of a single tree (supplementary table S1, Supplementary Material online). Each leaf sample was sequenced to an average of 126 million (M) cleaned reads (ranging from 92M to 163M), equivalent to around 44-fold raw depth of the 425 Mb estimated genome size (Dai et al. 2014). Each root sample was sequenced to around 119M cleaned reads or approximately 42-fold depth (ranging from 92M to 161M).





**FIG. 2.** Patterns of somatic mutations in the analyzed shrub willow. (a) Schematic of sampled branches and stem-cuttings. Eight major branches (labeled I to VIII) were sampled and shown here. One cutting (15–20 cm) per branch was made in branches I, III, IV, VII, and VIII, whereas four, five, and eight cuttings were made for branches II, V, and VI. Eight cuttings (labeled with superscript “M”) were measured for both leaf and adventitious root samples, otherwise only one leaf was sampled. Approximate locations where a cutting was sampled are shown in yellow. Lowercase letters “a–s” represent those BR-s mutations shared among different samples, which are colored by their appearance in leaves only (blue), roots only (red), and both organs (purple with asterisk). (b) Details of eight stem-cuttings with both leaves and adventitious roots sequenced. Samples collected from each cutting are enclosed within the black circle. Samples with shared BR-s mutations (lowercase letters) and a BR-m mutation (uppercase letter “P”) are indicated and colored correspondingly. The one poor sequenced root sample is shown in gray.

### Somatic Mutations Identified within the Single Shrub Willow

After mapping and variant calling, we searched for candidate somatic variant sites across all samples (see Materials and Methods and [supplementary fig. S2, Supplementary Material](#) online, for details). For each candidate variant site, the ancestral allele was inferred as the common allele presents in samples from most major branches (i.e., >4 of branches I~VIII here). A mutation is called at the variant site where one major branch carries an allele differed from the ancestral allele (hereafter denoted as BR-s mutations). The BR-s mutation therefore represents a somatic mutation that arose after formation of that major branch. For mutations raised before the formation of each major branch, samples from two or more major branches will share the same allele which differs from the ancestral alleles ([supplementary fig. S3, Supplementary Material](#) online, hereafter referred as BR-m mutations). The BR-m mutation lineage hence reflects the early stem cell lineages which form multiple major branches. We note that a few BR-m “mutations” may be generated by somatic recombination (i.e., gene conversion) in early cell lineages but not by mutation ([supplementary fig. S3, Supplementary](#)

[Material](#) online); we keep calling them mutations as 1) they are more likely early mutations (see later for details) and 2) we only use them to track cell lineages so does not matter whether these somatic variants are from mutation or recombination.

In total, we identified 199 reliable somatic mutations across all analyzed samples, including 182 BR-s mutations and 17 BR-m mutations ([table 1](#) and [supplementary table S2, Supplementary Material](#) online). The 182 BR-s mutations included 155 single-nucleotide variants (SNVs) and 27 insertion/deletions (INDELs), of which 177 mutations were distributed in assembled chromosomes and five were located in unanchored scaffolds ([fig. 3a](#)). Consistent with their origination as de novo mutations, nearly all identified BR-s mutations (180 of 182) were heterozygous, of which 176 were “homozygous->heterozygous” mutations (i.e., the ancestral genotype was homozygous, whereas the genotype of mutation was heterozygous) and four were “heterozygous->heterozygous” mutations. The only two exceptions (two “homozygous->homozygous” mutations in two root samples) were high likely due to sequencing bias that only one allele gets sequenced, as regions from which the two mutations were called and were associated with low read-depths

**Table 1.** Number of Mutations Identified in Shrub Willow YAF1.

Mutation Type	Predicted Effects	No. of BR-m Mutations				No. of BR-s Mutations				
		Sum	Leafa	Roota	Both	Sum	Leaf	Root	Both	Uncategorizedc
SNV	Total	14	12	1	1	155	36(4 <sup>b</sup> )	72(12)	2	45(4)
	Nonsynonymous	1	1	0	0	14	3	7(1)	0	4
	Synonymous	0	0	0	0	7	1	4(1)	0	2
	Intron	1	1	0	0	29	9(1)	14(3)	0	6(1)
	Intergenic	12	10	1	1	105	23(3)	47(7)	2	33(3)
INDEL	Total	3	3	0	0	27	3	14(1)	0	10(3)
	Frameshift	1	1	0	0	2	0	1	0	1(1)
	In-frame INDEL	0	0	0	0	2	0	1	0	1
	Intron	0	0	0	0	6	1	3(1)	0	2
	Intergenic	2	2	0	0	17	2	9	0	4(2)

<sup>a</sup>BR-m mutations present exclusively in leaves or in adventitious roots.

<sup>b</sup>BR-s mutations shared between multiple leaves or between multiple roots; the number of sample-specific mutations could be obtained by subtracting this number from the total number (number outside the parenthesis).

<sup>c</sup>For cuttings with only a single leaf was sequenced, the origin of these mutations could not be firmly assessed, so are left as uncategorized.

( $\leq 10$  reads). The proportion of “heterozygous- $>$ heterozygous” BR-s mutations ( $4/182 = 2.20\%$ ) was higher than the estimated heterozygosity (0.394%) of the genome (Fisher’s exact test,  $P = 0.006521$ ), implying possible mutagenic effects of heterozygosity per se, as has been reported in other studies (Amos 2010; Yang et al. 2015).

The 17 BR-m mutations included 14 SNVs and 3 INDELs, of which 16 mutations were distributed in assembled chromosomes and one was located in an unanchored scaffold (fig. 3a). Patterns of these 17 BR-m mutations and their implications are discussed in more detail below.

We selected 35 mutations, including all 17 BR-m mutations and 18 arbitrarily selected BR-s mutations, for Polymerase Chain Reaction (PCR) amplification followed by Sanger sequencing (supplementary table S3, Supplementary Material online). A total of 26 of these mutations (15 BR-m and 11 BR-s) were confirmed by Sanger sequencing (supplementary table S3, figs. S4 and S5, Supplementary Material online). The remaining nine nonanalyzable cases included six that failed PCR amplification (generally due to unavailability of suitable primers), and three that yielded poor Sanger sequencing signals (supplementary table S3 and fig. S4, Supplementary Material online). The absence of mutation alleles in control samples (e.g., some root samples were severed as control when testing BR-m leaf-exclusive mutations) is also confirmed (supplementary table S3, figs. S4 and S5, Supplementary Material online). Given a likelihood that the next mutation would fail validation, we estimated a false positive rate (FPR) of no more than ( $1/27 =$ ) 3.70%. Note that the FPR was even lower for mutations shared by two or more samples, for which the likelihood was less than 1%, the square of 3.70%, for a mutation being not validated in two or more independent samples.

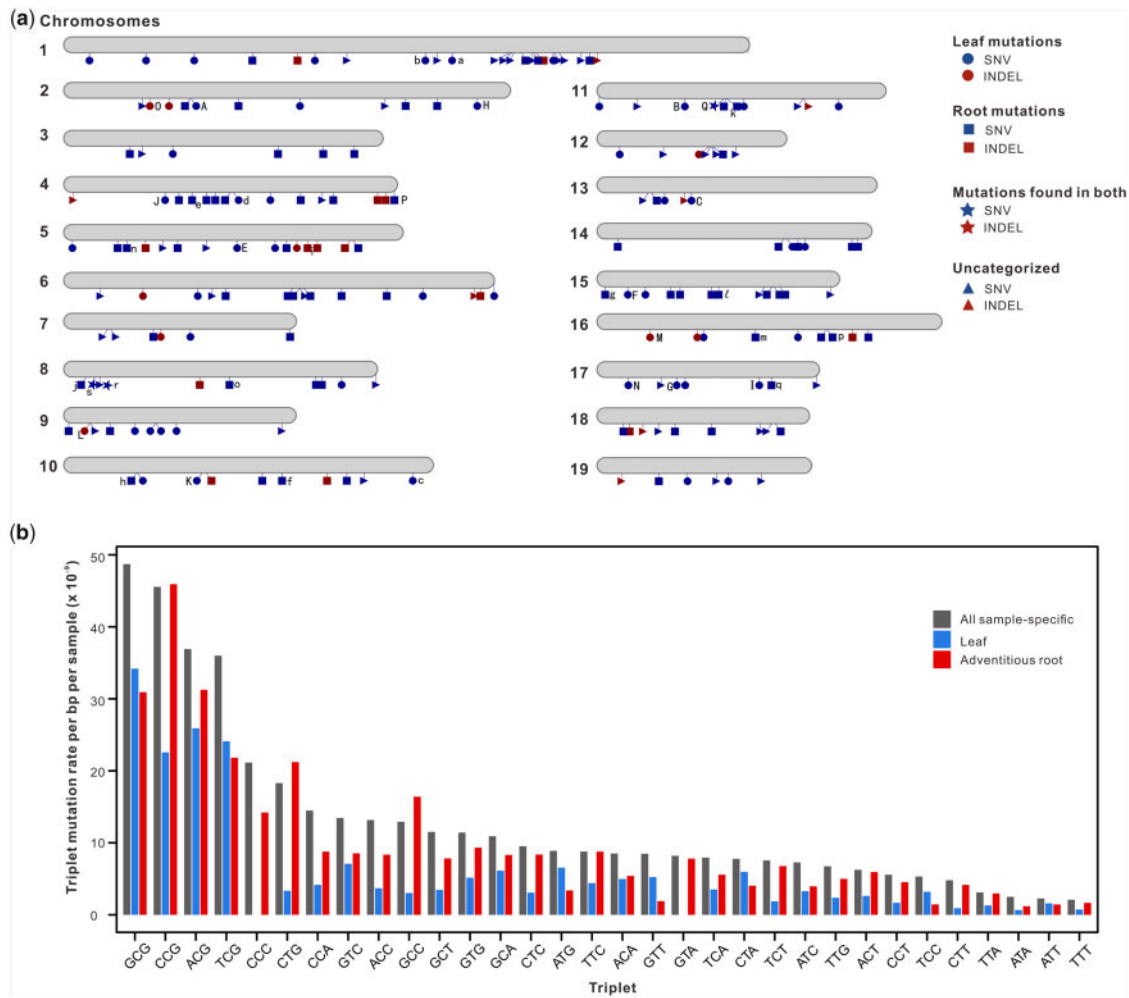
### Genomic Landscape and Profile of Somatic Mutations in Shrub Willow

The identified mutations allow us to depict the somatic genomic landscape of this woody plant. We mainly focus on BR-s mutations here (representing 91% of all identified mutations) to remove any uncertainty in ancestral inference that

might arise when dealing with BR-m mutations (supplementary fig. S3, Supplementary Material online).

We find more BR-s mutation events in chromosomes with larger size (Spearman’s  $Rho = 0.5568$ ,  $P = 0.01328$ ), with Chr01, the largest chromosome, having the most BR-s mutations (24 events identified). After accounting for chromosome size, the mutations are roughly distributed evenly across all chromosomes, with only a few regions (fewer than 9 Mb in size overall) as candidates to be mutation hotspots (supplementary table S4, Supplementary Material online, permutation test with 10,000 randomizations in 1 Mb windows,  $P < 0.05$ ). Most of the mutations are found within noncoding regions, whereas 21 SNVs and 4 INDELs reside in coding regions (table 1). The number of mutations in coding and noncoding regions are within the expectation from genomic coding areas (expected = 14.72% based on the coding regions of the reference genome, observed = 13.55% for SNVs, Chi-squared with Yates correction = 0.1682,  $P = 0.6817$ ; observed = 14.81% for INDELs, Chi-squared with Yates correction = 0.0002132,  $P = 1$ ), suggesting no apparent selection. Over 60% of mutations are transitions dominated by C- $\rightarrow$ T (or G- $\rightarrow$ A) changes and enriched in CpG sites (table 2 and fig. 3b), suggesting a transition bias when compared with the genomic expectation (genomic GC content =  $\sim 34.4\%$ ). This transition bias is frequently observed for spontaneous mutations in many plant species (Yang et al. 2015; Watson et al. 2016; Xie et al. 2016; Schmid-Siegert et al. 2017; Wang et al. 2019).

The generation of adventitious root is accompanied by callus formation (fig. 1b and d), which often introduces a large number of mutations when cultured in vitro (Phillips et al. 1994; Jiang et al. 2011; Zhang et al. 2014; Wang et al. 2019). Is callus-induced mutagenesis also happening in vivo and so might introduce some uncertainties here? The eight leaf-cutting groups allow us to directly assess this possibility. A BR-s mutation identified specific to a single leaf or an adventitious root (denoted as “sample-specific” mutation hereafter) is most probably only raised during the latest stem cell divisions. If these



**Fig. 3.** Properties of shrub willow somatic mutations. (a) Distribution of somatic mutations across the 19 chromosomes. Mutations found in one or more leaves (circles), in one or more roots (rectangles), or in both leaves and roots (stars) are distinguished. Uncategorized mutations are marked by triangles. SNV and INDEL mutations are colored in blue and red, respectively. BR-m mutations and BR-s mutations are labeled using uppercase and lowercase letters, respectively. (b) Triplet nucleotide context of sample-specific mutations. The triplet mutation rate per bp per sample is calculated as the number of mutated triplets (along with their complements, including mutation at first, second, and third positions) normalized by the overall abundances of the triplets in the reference genome. The rates are sorted by highest (leftmost) to lowest (rightmost) based on the overall sample-specific triplet mutation rates.

sample-specific mutations in adventitious roots have different profiles compared with those in leaves (Jiang et al. 2011; Zhang et al. 2014), the process of callus formation is likely to introduce substantial numbers of mutations in roots.

We find 108 BR-s sample-specific mutations within the eight leaf-root groups, including 35 leaf mutations and 73 root mutations (table 1 and supplementary table S2, Supplementary Material online). Of the 35 sample-specific leaf mutations, four are in coding sequences (CDS), whereas 11 of the 73 sample-specific root mutations are in CDS (table 1). These fractions of CDS numbers are not significantly different between two organs (two-sided Fisher's exact test,  $P = 0.7697$ ). Similarly, the non-synonymous/synonymous ratios of two organs (excess of nonsynonymous mutations over synonymous ones in one organ could indicate strong selection) are also not significantly different (table 1, two-sided Fisher's exact test,

$P = 1$ ). Both leaf and root mutations have a bias favoring transitions ( $\sim 65\%$ ) over transversions ( $\sim 35\%$ ), with most mutations as C $\rightarrow$ T or G $\rightarrow$ A changes (table 2). No significant differences in mutation spectra are observed between leaf and root mutations (two-sided Fisher's exact test,  $P = 0.4756$ ). Further investigation of triplet nucleotide context reveals that both leaf and root mutations are enriched in CpG sites (fig. 3b). There seems to be no significant difference with respect to per triplet mutation rate between leaf and root mutations (two-sided Fisher's exact test,  $P = 0.5477$ ), suggesting similar influences of trinucleotide context on mutations from both organs. A caveat here is that we have very limited number of mutations for each comparison, so more data are demanded for a solid conclusion. Nonetheless, at least we see no large difference in mutation profiles between leaf and root, implying no strong evidence for callus-induced mutagenesis here.

**Table 2.** Spectra of BR-s SNV Mutation Identified in Leaves and Adventitious Roots.

Type of Mutation	Overall (Fraction)	Sample-Specific	
		Leaf (Fraction)	Root (Fraction)
Transitions (Total)	106 (0.627)	22 (0.688)	39 (0.650)
A->G/T->C	17 (0.101)	7 (0.219)	5 (0.083)
G->A/C->T	89 (0.527)	15 (0.469)	34 (0.567)
Transversions (Total)	49 (0.290)	10 (0.313)	21 (0.350)
A>T/T->A	11 (0.065)	3 (0.094)	4 (0.067)
A->C/T->G	11 (0.065)	2 (0.063)	3 (0.050)
G->T/C->A	23 (0.136)	5 (0.156)	12 (0.200)
G->C/C->G	4 (0.024)	0 (0)	2 (0.033)

NOTE.—The fractions of each change are given in parentheses.

### Leaves and Adventitious Roots Are Likely Segregated Early

To test between model 1 and model 2 (fig. 1e), we take advantage of the somatic mutations that are shared among samples as molecular markers.

First, the 17 BR-m mutations allow us to trace back to the stage prior to the branching events that forming the eight sampled branches (I~VIII). We find from the 15 BR-m mutations (mutations “A~O” in fig. 4) that the “mutation alleles” propagate exclusively in leaves across all branches (referred to as “leaf-exclusive” mutation hereafter). At these mutation sites, all sequenced roots have homozygous genotypes, whereas most leaves have heterozygous genotypes (over 14 leaf-exclusive mutations [“A~N”] were present in  $\geq 28$  of 33 sequenced leaf samples). Given the low heterozygosity ( $\sim 0.394\%$ ) estimated for the genome, we are reasonable to consider them as leaf de novo mutations (e.g., a A/A->A/T mutation in leaf) rather than root mutations/recombination (e.g., a A/T->A/A mutation or gene conversion in root, supplementary fig. S3, Supplementary Material online). This observation suggests that the leaves and roots are initiated from a cluster of multiple founder cells, as these 15 mutations should already present in the initial founder cell niche. The near-fully absence of these mutation in roots could hardly be explained by their overall lower-sequencing coverage (supplementary note S1, Supplementary Material online). Further Sanger sequencing accompany PCR amplification confirmed their absences in roots (supplementary note S1, table S3, figs. S4 and S5, Supplementary Material online). The chance of seeing 15 such mutations which present in nearly all 33 leaves but absent in any of the 22 root samples is essentially nearly zero under model 1, after correcting for the putative influence of lower root coverage (see Materials and Methods for details), providing strong evidence for rejecting model 1. These 15 mutations suggest that at least one cell lineage (the lineage propagate mutations “A~O”) forming the leaves is segregated from the adventitious stem cell lineage forming the roots and that this segregation predates the formation of all sampled branches (fig. 5a and b). The relatively large number of mutations further leads us to speculate that the mutations “A~O” mark putative rapidly dividing cell lineages, also distinct from the stem cell lineages of the axillary meristem (see Discussion for details).

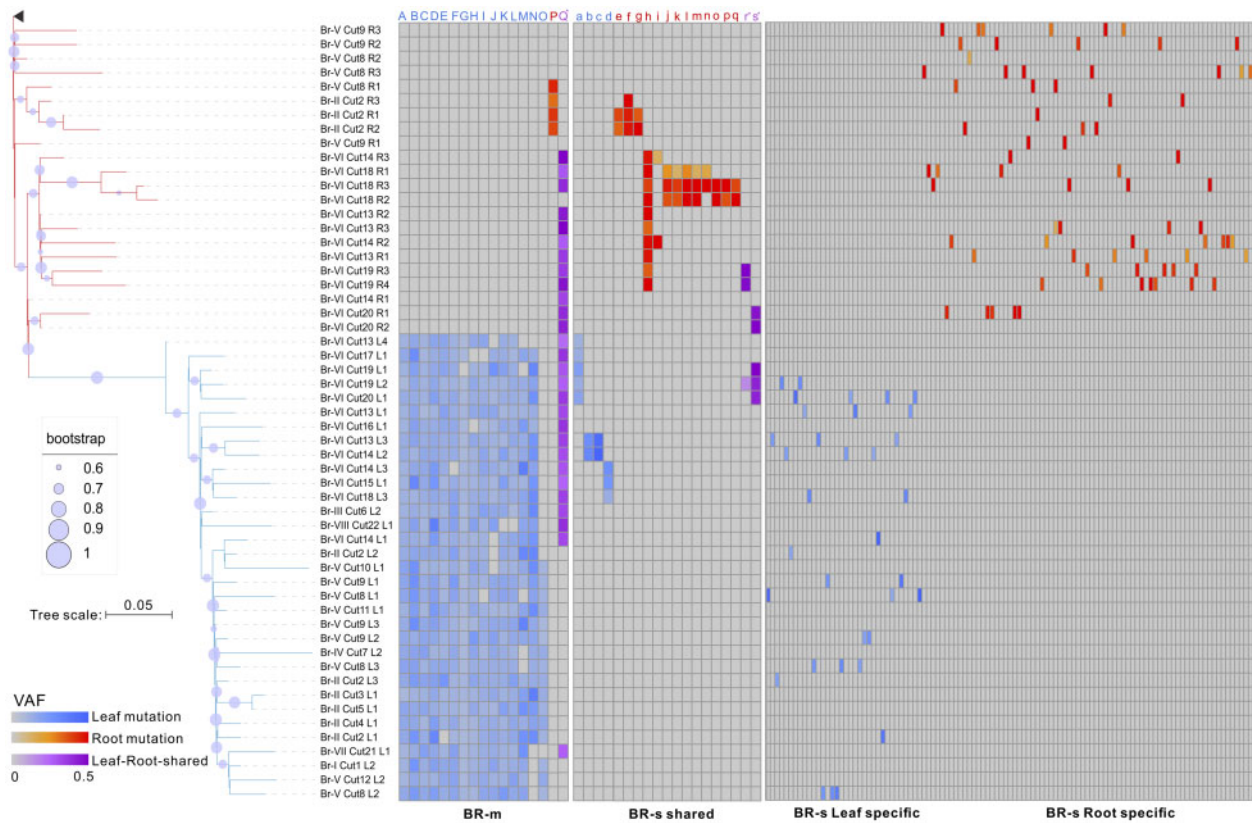
Second, we consider a shared de novo mutation that we observe in roots of cuttings from two branches. This mutation (“P” in fig. 4) is found in roots from branches II and V, whereas it is absent in all leaves (referred to as “root-exclusive” mutation hereafter). For mutation “P” to exclusively propagate through adventitious roots but no leaves across branches, the expected chance is low ( $< 0.0352$ ) under model 1 (i.e., the chance that none of the 15 leaves in branches II and V would have this mutation). Thus, this root-exclusive mutation “P” is also not consistent with mode 1.

Third, we consider more recent mutations which are de novo to specific major branches. Of the 182 BR-s mutations, 19 mutations, which are from cuttings with both leaves and roots sequenced, are shared among different samples in a branch (fig. 4 and supplementary table S2, Supplementary Material online). There are four BR-s mutations (“a~d” in branch VI) present in multiple leaves and propagated through two or more cuttings (usually consecutive cuttings), indicating their origination prior to the formation of the corresponded cuttings. None of the four mutations is found in roots within the same cutting, which is with low expectation under model 1 ( $< 0.168$  for all three roots within the same cutting to lack this mutation).

Fourth, another 13 BR-s shared mutations (mutations “e~g” in branch II and “h~q” in branch VI) are de novo to roots of cuttings within a branch, of which 12 (“e~g” and “i~q”) with the exception of “h” are present only in single cuttings (figs. 2a, 2b, and 4). This contrasts with the leaf specific mutations “a~d,” which are all present in multiple cuttings (fig. 4), suggesting that these root mutations possibly arise later. As before, the finding that all 13 root de novo mutations would not be found in leaves has a very low probability under model 1 ( $< 0.055$ ).

In the above, we ignore three mutations, one BR-m mutation (“Q”) and two BR-s mutations (“r” and “s”), which are found both in leaves and roots (fig. 4). Mutation “Q” is present in nearly all sequenced leaf and root samples from branches III, VI, VII, and VIII, but is absent from all sequenced samples from branches I, II, IV, and V. This mutation implies a possible closer relationship of branches III, VI, VII, and VIII to each other relative to the other four branches considering the extremely low possibility that we would see the same mutation





**Fig. 4.** Reconstructed ontogenetic tree of sequenced samples. The right four panels show the presence of 17 BR-m (“A~Q”), 19 BR-s shared (“a~s”), and 108 BR-s sample-specific mutations in each sequenced sample. Mutations in each sample are marked by different colors following: blue—mutation only observed in leaf, red—mutation only observed in adventitious root, purple (also marked by asterisk)—mutation observed in both leaf and root, and gray—ancestral alleles without mutation. Uncategorized mutations are not shown here. VAF of each mutation in each sample is indicated by the saturation of the color. The leftmost panel is a maximum-likelihood ontogenetic tree which is constructed using base mutations identified. The black triangle represents the inferred ancestral sequence. Only bootstrap values over 0.6 are shown (1,000 replicates bootstrap test). Corresponding sample IDs are given between two parts with the format: “Branch ID” (e.g., Br-I, Br-II, . . .), “Cutting ID” (e.g., Cut1, Cut2, . . .), and “Sample ID” (ID with “L” initial represents leaf, ID with “R” initial represents adventitious root).

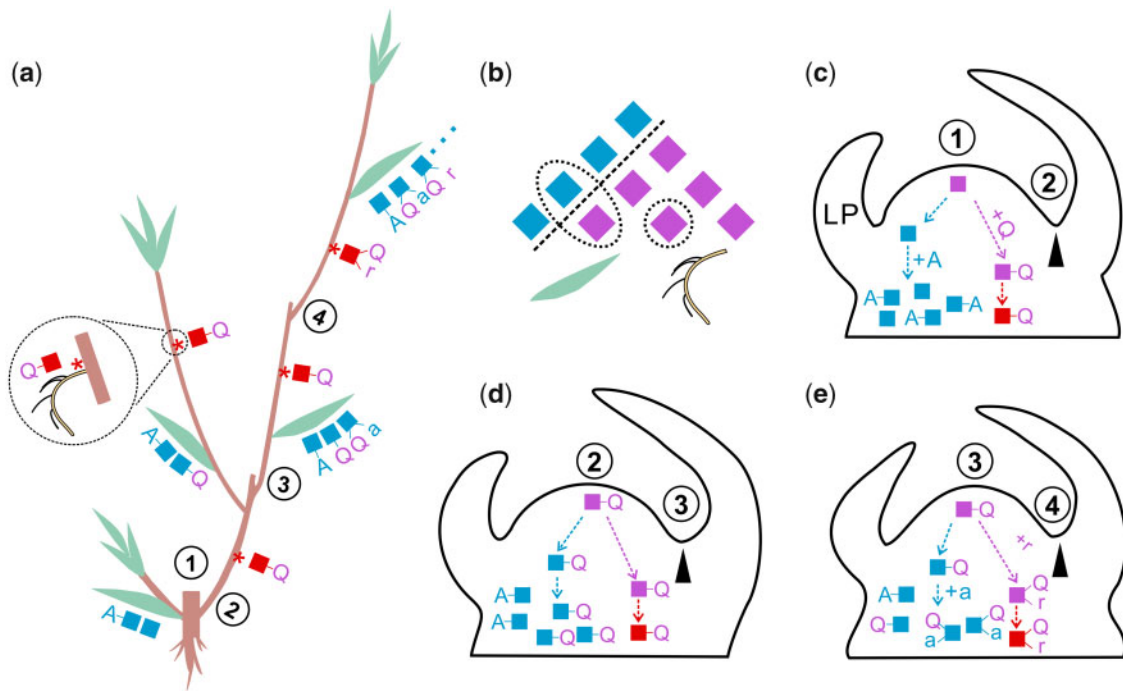
independently arise in four branches. The mutation “r” appears in one leaf and two roots from a single cutting (fig. 2b), and the mutation “s” presents in two nearby cuttings and are found in two leaves of cutting-19 while in one leaf and two roots of cutting-20 (fig. 2b). Given the fact that the dominate part of mutations favors model 2 over model 1, observation of these three mutations in both leaves and adventitious roots is less likely to suggest any possibility of model 1. Rather, these three mutations are more likely to imply putative cell lineages, which could replenish both leaf and root cell lineages after their early segregation (fig. 5).

In summary, the majority of shared BR-m (16 of 17) and BR-s (17 of 19) mutations that propagated separately between leaves and adventitious roots gives clear evidence that there exist different cell lineages between two organs segregated prior to the formation of all branches. The constructed ontogenetic tree using somatic mutations (fig. 4) matches well with the expectation of model 2. Beyond the segregation, the three mutations shared between leaves and roots also imply the existence of shared replenishing cell lineages for both organs (fig. 5).

### Leaf Might Be Differentiated from Multiple Founder Cells whereas Root Is Differentiated from One

Are leaves and adventitious roots differentiated from multiple founder cells or single founder cell? A mutation presents within a single founder cell will be passed to nearly all cells whose DNA were sequenced, whereas a mutation from one of multiple founder cells will be passed to only part of the sequenced cells, leading to varied allele frequencies (fig. 6a). Assuming a low chance of somatic recombination, all somatic mutations are expected to remain in a heterozygous state during propagation (i.e., allele frequency = 50%). We consider the variant allele frequency (VAF, estimated as “read-depth of the mutation allele”/“overall read-depth at this site” for a certain mutation of each sample) as a surrogate of the allele frequency of mutation in terminal cell population of different organs (Yu et al. 2020). In real sequencing, the VAF is expected to follow a binomial distribution (fig. 6a) determined by both allele frequency and the read-depth at a given locus (Yu et al. 2020). The distribution of VAF for mutations from single founder cell is predicted to be similar to that derived from the intrinsic heterozygous sites (i.e., pre-existing heterozygous variants) in the genome, whereas





**FIG. 5.** A proposed model for development of leaves and adventitious roots in shrub willow. (a, b) Progenitor cells (purple squares) of axillary meristems (arrowhead) are segregated (dashed lines) early in SAM from rapidly dividing cells (blue squares). The progenitor cells (red squares) of adventitious meristems (asterisks) are likely derived from the segregated axillary meristems. Each leaf is differentiated from the leaf primordia (LP) which are formed by a chimeric cluster of rapidly dividing cells (blue squares aside each leaf, numbers are not scaled to actual proportions of different sources). Each adventitious root is differentiated from a single founder cell within the adventitious meristems (red squares aside the asterisks). (c, d, e) Three branching events are indicated (①→②, ②→③, ③→④). The earliest rapidly dividing cells marked by mutation “A” are pushed away together with the meristematic stem cells and are recruited to form leaves after iterative branching. The axillary meristems replenish new progenitor cells for leaf (e.g., blue with “Q”) and adventitious meristems (e.g., red with “Q”). Mutations could happen frequently within rapid dividing cell lineages (e.g., mutation “a”), but will only present with low VAF in leaves. Only mutations arise in meristematic cells (e.g., mutation “Q” and “r”) could reach a high VAF in leaves of subsequent branches. Cell divisions from some progenitor cells are indicated by dashed lines with arrows.

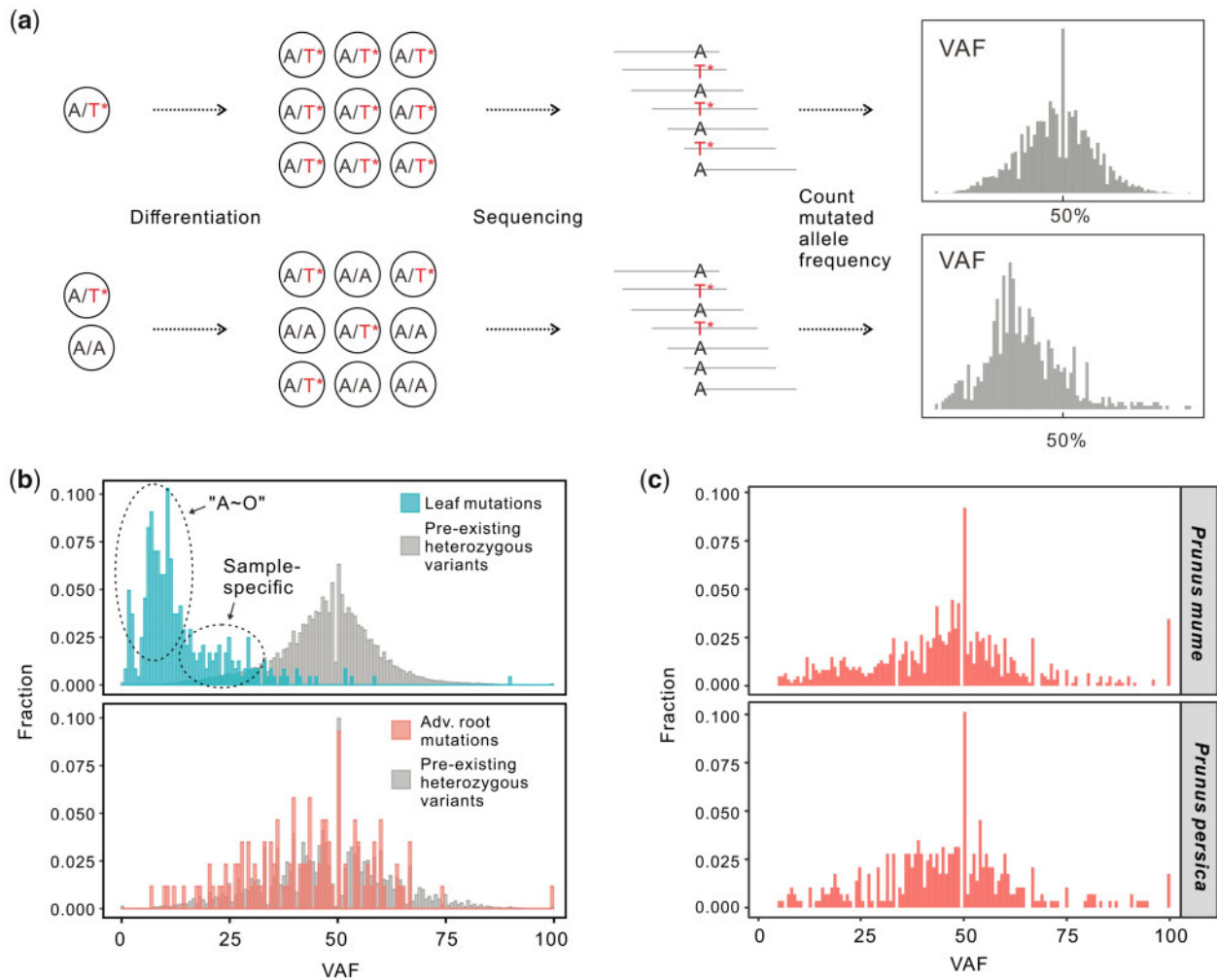
mutations from multiple founder cells tend to have VAFs  $\ll$  50% (fig. 6b).

We found most leaf mutations have VAFs  $< 30\%$  (figs. 4 and 6b), apparently deviating from the distribution of VAF for pre-existing heterozygous variants (unequal variances *t*-test,  $P < 2.2e-16$ ). It’s noteworthy that the leaf-exclusive BR-m mutations generally have lowest VAFs (mean = 10.6%, SD = 6.42%, fig. 6b), as validated by PCR-Sanger results (mean = 11.2%, SD = 8.39%, supplementary figs. S4 and S5, Supplementary Material online), suggesting these mutations only present in a limited portion of cells within each leaf. Contrary to leaf, the distribution of VAF for root mutations matches well with the distribution of pre-existing heterozygous variants (unequal variances *t*-test,  $P = 0.056$ ) with a peak around 50% (figs. 4 and 6b). The distinct distributions between leaf mutations and root mutations suggest that leaf samples are more likely derived from multiple founder cells (fig. 6a and b), whereas an adventitious root may be differentiated from a single founder cell. Both findings are consistent with the previous chimera studies (Marcotrigiano and Stewart 1984; Broertjes and van Harten 1985; Poethig 1987; Furner and Pumfrey 1992; Irish and Sussex 1992).

Secondly, the different number of founder cells in initiating leaf and adventitious root predicts that mutations are more

easily fixed in root than in leaf owing to the cell population “bottleneck.” Therefore, we expect higher observable mutation rate in root than in leaf. The prior-mentioned 108 BR-mutations specific to each leaf or adventitious root allowed us to test this prediction. The sample-specific mutations share the same timespan between leaf and root during differentiation from their latest progenitor cells. Based on these sample-specific mutations, we estimated a normalized per site per sample (see Discussion on interpretation of the unit) rate of  $4.32 \times 10^{-9}$  ( $\pm 0.786 \times 10^{-9}$  SEM) for leaf SNV mutations, and  $8.15 \times 10^{-9}$  ( $\pm 1.42 \times 10^{-9}$  SEM) for root SNV mutations. The INDEL mutation rate is  $4.05 \times 10^{-10}$  ( $\pm 2.21 \times 10^{-10}$  SEM) per site per sample for leaf and  $1.76 \times 10^{-9}$  ( $\pm 0.501 \times 10^{-9}$  SEM) per site per sample for root. The observed somatic mutation rate in root is around 2-fold of that in leaf (two-sided Brunner–Munzel [BM] test,  $P = 0.033$  for SNV mutations and  $P = 0.008$  for INDEL mutations), confirming the prediction.

Given the same read-depth cutoff for mutation calling and both organs are sequenced at  $\sim 40$ -fold, an issue here is that more leaf mutations will be below the read-depth threshold (and become undetectable) if the leaf is initiated from multiple founder cells although root is only from a single founder cell (fig. 6a). This offers another validation of whether the



**FIG. 6.** Estimated VAFs from somatic mutations for different developmental stages. (a) Read-depth fractions as the indicator of mutation allele frequency from single or multiple founder cells. Here “A” represents the ancestral allele, whereas “T\*” stands for de novo mutation presents in one recent founder cell. Only when mutation “T\*” is from a single founder cell can it be propagated to all later cells, leading to a binomial distribution of VAF peaked at 50%, otherwise the VAF distribution will be skewed toward a lower peak depending on the initial fraction of mutation “T\*” in founder cells. (b) Distributions of VAFs for leaf and root mutations. The gray histograms are drawn from pre-existing variant sites randomly sampled ( $n = 10,000$ ) from fully heterozygous sites across all sequenced leaf and root samples to measure the variance of VAFs owing to sequencing bias. Only mutation and variant sites with read-depth no less than 10 were presented here to reduce the bias in low-depth regions. The peaks contributed by BR-m “A~O” mutations (around 12.5%) and BR-s sample-specific mutations (around 25%) in leaves are marked by dashed circles. (c) Distribution of VAFs estimated for root mutations from woody species *Prunus mume* and *Prunus persica*. Data collected from Wang et al. (2019). For all figures, “adv. root” is abbreviation for adventitious root.

higher mutation rate observed in adventitious root is due to a higher fixation rate (i.e., because root is from a single founder cell, whereas leaf is from multiple founder cells) or a more mutagenic process (e.g., more mutations are induced in roots through callus formation). Once we use a fairly depth cutoff for two organs, that is, lower cutoff in leaf than in root, we expect to see similar mutation rates if the fixation chance is the main factor, whereas persistent higher mutation rate in root, if any mutagenic process, is the main factor. Therefore, we require  $\geq 10$  supporting reads for root mutations although keeps requiring only  $\geq 5$  supporting reads for leaf mutations (2-fold requirement for root mutations compared with leaf mutations considering leaf is from at least two founder cells). The re-estimated root mutation rate is  $3.53 \times 10^{-9}$  ( $\pm 1.08 \times 10^{-9}$  SEM) and  $2.72 \times 10^{-10}$  ( $\pm 1.87 \times 10^{-10}$

SEM) per site per sample for SNV and INDEL mutations, respectively, which is similar to that of the leaf’s (two-sided BM Test,  $P = 0.4104$  for SNVs and  $P = 0.6639$  for INDELS). The similar mutation rate of two organs at more comparable depth cutoff confirms that the fixation chance is the major factor leading to the observed higher mutation rate in adventitious root, and further confirms that root forms from fewer founder cells than leaf.

## Discussion

The canonical plant development model posits that SAM contains a self-replenishing population constituted by a few slowly dividing initial cells whereas some of their descendants divide actively when differentiating into leaves (Evans and

Barton 1997). The initial cells only divide when forming new axillary meristems upon iterative branching (Burian et al. 2016). It is less clear however whether a leaf in a high-order branch also contains cells from the descendants of the earliest SAM or is solely derived from the descendants of the latest axillary meristems. Our results suggest that both the descendants from earliest SAM (rapidly dividing “somatic” cell lineage propagate mutations “A~O”) and the replenishing cell lineages (slowly dividing stem cell lineages propagate mutations “Q” “r,” and “s”) contribute to the formation of the leaf (fig. 5). Considering that adventitious meristems/buds mostly form from cambial cell divisions (Steffens and Rasmussen 2016), which are also derived from SAM (Nieminen et al. 2015), we can guess the replenishing cell lineages most likely correspond to the axillary meristems (fig. 5). Our results hence confirm that the axillary meristems (together with the adventitious meristems) are segregated from other rapidly dividing cell lineages (e.g., the cell lineage marked by “A~O”) as early as the cell divisions prior to the formation of all major branches (Klekowski et al. 1985; Poethig 1987; Irish 1991; Szymkowiak and Sussex 1996; Evans and Barton 1997; Burian et al. 2016), with the state-of-the-art WGS-based clonal analysis.

Histological analysis in shrub willow shows that the apical meristem is surrounded by multiple putative leaf primordia (supplementary fig. S6, Supplementary Material online), a structure similar to many plants including other willow species (Berggren 1984) and Arabidopsis (Evans and Barton 1997; Burian et al. 2016). Since the axillary meristems arise later from the axils of leaves developed from these primordia (fig. 5), it is highly likely that the axillary meristems have already been specified or segregated there (Burian et al. 2016). However, only with marker-based analysis could we confirm this segregation.

Beyond the early segregation, our results further reveal that these rapidly dividing “somatic” cell lineages (carrying BR-m mutations A~O) can survive leaf and branch formation, and are present even after several iterative branching without being replaced by other lineages (fig. 5). The persistent presence of these BR-m mutations with low VAFs implies that the independent branches are likely recruited from a chimeric cluster of cells where only a portion of the cells retained these mutations. Furthermore, the cells containing these BR-m mutations may be predetermined to contribute exclusively to the formation of leaves in the new branch, but not the growth of stems and vascular tissues. Therefore, initiation of axillary meristem from SAM seemingly involves concomitantly cell divisions of both slowly and rapidly dividing cells (fig. 5). Preservation of such rapidly dividing cells along with iterative branching provides another possible strategy to reduce the number of cell divisions in meristematic stem cells when forming new leaves in subsequent branches.

The intraorganismal selection hypothesis posits that somatic mutations may be selectively advantageous in generating “mosaic sectors” against natural enemies, such as herbivores and pathogens (Whitham and Slobodchikoff 1981; Whitham 1983; Antolin and Strobeck 1985; Suomela and Ayres 1994; Simberloff and Leppanen 2019). This hypothesis assumes that selective pressures could help to fix the

newly arisen mutations (Whitham and Slobodchikoff 1981; Simberloff and Leppanen 2019), which in principal requires that 1) the new mutation is directly selection-favored within the meristem to ensure its fixation in later development and 2) the new mutation is advantageous in the heterozygous state and provides phenotypic effects at both moderate- and low allele frequency. However, these assumptions are poorly supported by our results, as we found that the chance for a somatic mutation to be fully fixed in branch is generally very low as supposed previously (Burian et al. 2016). This is evidenced by: 1) no BR-s mutations were detected in either all leaves or all roots within a major branch in this study, suggesting later mutations can rarely fully fix within a whole branch; and 2) most mutations may only have a chance to be fixed within a single-cell lineage derived from meristems, as the majority of mutations detected in this study are sample-specific mutations; and 3) though early mutations may have higher chance to be propagated to more branches, the iterative replenishing process by low-dividing axillary meristems might eventually reduce their fixation proportion as seen for BR-m mutations (fig. 6b). The replenishing process seems to introduce very limited number of mutations to axillary meristems, as witnessed by only three mutations shared between leaves and roots, adding further support for the notion that plants protect their meristems from mutational burden by segregating these meristems early (Burian et al. 2016; Schmid-Siegert et al. 2017; Lanfear 2018; Plomion et al. 2018).

We note here that our resolution in mutation detection is limited by moderate sequencing depth (~40-fold) so only mutations with VAF no less than 12.5% (fig. 6b) could be reliably detected (five supporting reads/40-fold). Thereby, the mutations that accompany rapidly dividing cells might be much more than those we detected. The rate estimated from the sample-specific mutations is likely a more accurate estimator. Since these mutations are detected at moderate sequencing depth, only early mutations from one or a few early cell divisions within the meristem could reach this high fraction in the terminal cell population. Thereby, the unit might be most suitably interpreted as “per site per (few) cell division(s)” instead of “per site per sample.” Given the same reason, the “per site per sample” rate is possibly not directly comparable to those “per site per year” rates estimated for many old trees (Schmid-Siegert et al. 2017; Plomion et al. 2018; Hanlon et al. 2019; Hofmeister et al. 2020; Orr et al. 2020), as these “per site per year” rates represent “observable” (or fixable) mutations after many years of accumulation, rather than the yearly mutation rate.

The low fixation chance but not necessarily low rate of somatic mutations we see here is most possibly a direct consequence of early segregation of axillary meristems, a conjecture that has been proposed for a century (Laux 2003; Lanfear 2018). The early-segregated meristems also suggest the possibility of an early-segregated plant germline (Lanfear 2018). The estimated mutation rate for axillary meristems is much lower than that of other cell lineages, which is likely consistent with the prior observation that the somatic reversion rate is 4-fold lower in the commonly assumed gamete-bearing L2



layer than in L1 layer in peach (Chaparro et al. 1995). Tracking heritable somatic mutations may provide more clues as to which cell lineage directly contributes to the formation of gametes in woody plants.

Does adventitious root contain another independent cell lineage, similar to the one marked by “A~O” in leaf, besides the replenishing cell lineages? Currently, we found no strong evidence. The root BR-m mutation “P” is present in all three root samples from cutting-2 but only one of three root samples from cutting-8 (fig. 2b). Therefore, the mutation “P” could have occurred in a cell downstream of its original meristematic stem cell but ended up in a limited number of closely related adventitious meristems, distinct from mutations “A~O.” This partly explains the contradictory observation that there are many BR-m mutations present in nearly all leaves and absent in all roots (e.g., mutations “A~O”), but no BR-m mutation is present in most adventitious roots and absent in all leaves.

The low fixation chance seems contrary to the empirical prevalence of bud sport mutants (Shamel and Pomeroy 1936; Foster and Aranzana 2018). However, our finding that the adventitious root is likely differentiated from a single founder cell (fig. 6a) suggests a possibility that some mutations can be fixed through a cellular population “bottleneck” effect within certain meristems. By establishing an organ from a single ancestral cell, a mutation that predates this ancestral cell can be fully fixed within the organ. Furthermore, this process seems to be a universal phenomenon for the development of plant roots, as the VAF distribution of somatic mutations identified in underground roots (i.e., roots differentiated from root apical meristems) of some plants also peaked at around 50% (fig. 6c). The varied fixation chance of somatic mutations within different organs emphasizes a strategy that not only helps reducing somatic mutations fixed in reproduction-related meristems but also maintains a substantial level of somatic mosaicism within other meristems.

Our study also demonstrates the power of somatic mutation as a hallmark in elucidating cell lineages (Wang et al. 2019). However, compared with the standard methods using chimeras which operate at the level of the cell (Poethig 1987; 1989; Szymkowiak and Sussex 1996; Frank and Chitwood 2016), our approach has limitations in explicitly determining which cell lineage corresponds to the specified cell layer of the SAM, since the somatic mutations are called from bulk sequencing of multiple cell populations (Wang et al. 2019). Future studies combining single-cell profiling (Woodworth et al. 2017) as well as DNA barcoding technologies (Kebschull and Zador 2018) will give a more complete picture about the cellular development of woody plants.

## Materials and Methods

### Sample Collection, DNA Extraction, and Whole-Genome Resequencing

The shrub willow (*S. suchowensis*) YAF1 (male) was kindly provided by Jiangsu Academy of Forestry, China, and was transplanted in Nanjing University. This tree was started as a sapling from cutting breeding (single cutting in origin) in

year 2014 and was coppiced in 2015 to stimulate new shoots in 2016 (supplementary fig. S1, Supplementary Material online). The sampling was performed in 2016. A total of 22 stem-cuttings from eight major branches were sampled from this tree (fig. 2a). The stem-cuttings we collected are from different developmental stages (fig. 2a), including postcoppice stem tissue (e.g., cuttings 6, 8, 13), sylleptic (e.g., cuttings 3, 15), and doubly sylleptic branches (e.g., cuttings 19, 20), to represent iterative branching processes. The collected cuttings were first cultivated in 1/2 Murashige and Skoog (MS) fluid medium for 1 week to obtain sufficient nutrients and were then transplanted into fresh water and cultivated under ~25 °C room temperature. Leaf and root DNA were extracted using the Cetyltrimethyl Ammonium Bromide method (Clarke 2009) after growing for about 3 weeks. Quality of DNA samples was tested by microplate reader and agarose gel electrophoresis to ensure sufficient quantity and integrity for whole-genome sequencing. Qualified DNA samples were fragmented into an insert size of about 300–350 bp by sonication and sequenced on the Illumina HiSeq4000 platform with 150-bp paired-end reads. All quality testing, library construction, and sequencing steps were performed at BGI-Shenzhen. Each sample was sequenced to around 18 Gb (averaged 44×, supplementary table S1, Supplementary Material online) after cleaning for low quality reads, either >5 bp Ns or more than 30% base calls with quality score below 20.

### Alignment and Initial Variant Discovery

Detailed procedures for read alignments and initial variant discovery followed the same pipelines described in Wang et al. (2019). In brief, cleaned reads were mapped to the pseudomolecule-level assembly of *S. suchowensis* (Dai et al. 2014) downloaded from PopGenIE database (ftp://plantgenie.org/Data/PopGenIE/Salix-suchowensis, v4.1) using BWA-mem algorithm (Li 2013). This assembly contains ~229-Mb sequences anchored in 19 chromosomes, with an additional 6.7-Mb sequences assembled into large scaffolds longer than 10 kb. The BWA (version 0.7.10-r789) was run with option “-M” to keep the resulted SAM file (Li et al. 2009) compatible with downstream processes. Picard package version 1.114 (https://broadinstitute.github.io/picard/, last accessed September 30, 2021) was used to mark noninformative PCR duplicates in mapping results by MarkDuplicates function bundled inside. GATK package (version 3.5) was further used to perform local realignment using RealignerTargetCreator and IndelRealigner functions to minimize false variant calls due to alignment errors around INDEL locations (DePristo et al. 2011).

After mapping to the reference assembly, the leaf samples covered ~90.9% of the reference genome with at least one reliable read (i.e., reads with mapping quality score [MAPQ] ≥20, indicating a mismapping rate ≤1%), from the lowest 87.0% to the highest 91.3% (supplementary table S1, Supplementary Material online). This proportion dropped slightly to ~81.0% in average when only considering genomic regions covered by at least ten reliable reads (supplementary table S1, Supplementary

Material online). The root samples contained slightly lower coverage than leaf samples, ranging from 78.6% to 90.7% (86.6% in average, [supplementary table S1, Supplementary Material](#) online), but the reduction was stronger when only considering regions with  $\geq 10$  reliable reads ( $\sim 52.8\%$  in average, [supplementary table S1, Supplementary Material](#) online). The low coverage in root samples is due to higher sequencing yield loss caused by putative bacterial contaminants in roots ([supplementary note S2 and fig. S7, Supplementary Material](#) online), especially in one sample “Cut18-R2,” which yielded a coverage of  $\sim 1.4\%$  with over ten reliable reads. Therefore, we excluded this sample from identification of mutations, but used it for confirmation of mutations identified in other samples. We also paid special attention to remove any putative contaminants when performing mutation calling ([supplementary note S2 and fig. S2, Supplementary Material](#) online).

Variants, including SNVs and small-sized INDELs (1–100 bp), were called for all samples using two algorithms, UnifiedGenotyper (UG) and HaplotypeCaller (HC), both implemented in GATK. A union set of called variants from both algorithms were used for downstream mutation identification to reduce the false negative rate (FNR; [Xie et al. 2016; Wang et al. 2019](#)). Only reads with MAPQ over 20 (Phred scaled, equivalent to less than 1% mismatching rate) were used for variant calling here. Variants that resided in short scaffolds (size  $< 10$  kb) were discarded for further mutation calling, as these unanchored fragments (average size = 666 bp, median size = 372 bp) were generally highly repetitive in nature.

### Somatic Mutation Calling

Since spontaneous somatic mutations generally arise and are fixed during cell divisions, the distribution of mutations is expected to reflect their historical origin in the plant's development, that is, follow the ontogeny. For example, a mutation only present in a single branch (denoted as a BR-s mutation) most likely emerged during or after the separation of that branch. In contrast, a mutation present in multiple branches (denoted as a BR-m mutation) would indicate an origination predating the separation of these branches (the underlying ontogeny not observed). The BR-s mutations could be easily identified through comparing one branch against all other branches. For BR-m mutations, we considered two possible situations ([supplementary fig. S3, Supplementary Material](#) online): 1) part (denoted as “ $m$ ,”  $m > 1$  and  $m < 8$ ) of the eight sampled branches was originated from a single branch ([supplementary fig. S3a and c, Supplementary Material](#) online), for which we should expect mutations present in  $m$  branches but absent in all other branches (since these mutations are earlier than the separation of the  $m$  branches, we should also expect them to be present in nearly all samples obtained from the  $m$  branches); (2) if the early segregation model (model 2) is correct, that is, the two organs could have segregated cell lineages at the beginning, there might exist some mutations present exclusively in leaves or in roots (similarly, the mutations are also expected to be present in nearly

all leaves/roots according to when they are generated, [supplementary fig. S3b and d, Supplementary Material](#) online).

To detect BR-m mutations under the first situation, we compared all combinations of seven branches against the remaining branch and ensuring that the mutation was absent in the one branch but present in more than one of the seven branches ([supplementary note S3, Supplementary Material](#) online). Note that this corresponds to all situations (i.e.,  $m = 7, 6, 5, 4, 3, 2$ ) as we do not require the mutations to be present in all seven branches. For mutated branches with multiple samples sequenced (branches II, V, and VI), these mutations are expected to be present in all of them as mentioned above. However, amplification or sequencing bias could cause the mutated allele to be missed in a few samples. Therefore, we used a lenient requirement that the mutation should be present in at least 90% of the samples in these three branches, that is, at least 8 samples in branch II, 13 samples in branch V, and 22 samples in branch VI ([supplementary note S3 and table S5, Supplementary Material](#) online). To detect BR-m mutations under the second situation, we directly compared leaves with roots. A mutation was determined to be exclusive to leaves if it was only present in leaves from two or more branches but was absent in all roots. A mutation was determined to be exclusive to roots if it was only present in roots from two or more branches but was absent from all leaves.

All mutation candidates were called first using a parallel comparison strategy ([Sung et al. 2012; Wang et al. 2019](#)) to remove sequencing or mapping errors which show up repeatedly between focal samples which are supposed to carry the mutation and control samples which are supposed not to carry the mutation ([Li and Stoneking 2012](#)). Candidate mutations were further filtered by 1) removing sites with a low variant quality score ( $< 50$ ) or with many noninformative calls ( $> 5$  samples applied here as sites failed this criterion were frequently associated with other sequencing or mapping issues); 2) requiring  $\geq 5$  supporting reads for the mutation in at least one focal sample, and requiring at least one reliable read carrying the identical mutation (with base quality  $\geq 30$  to avoid sequencing errors) for the remaining focal samples; and 3) requiring presence of both forward and reverse strands for the supporting reads, which minimizes mismatching caused by homology sequences ([Xie et al. 2016; Wang et al. 2019](#)). For BR-m mutations under situation 1, more rigorous criteria were used to reduce FPR after extensive assessment ([supplementary note S3 and table S5, Supplementary Material](#) online). A flowchart of the mutation screening is included to show the number of candidates at each step ([supplementary fig. S2, Supplementary Material](#) online). All filtered mutations were assessed manually using Integrative Genomics Viewer ([Thorvaldsdóttir et al. 2013](#)) to remove ambiguous cases such as mutations found in regions with extremely high-sequencing errors, mutations from possible exogenous contaminations, or mutations from reads which are poorly aligned. Only candidates that passed all criteria were retained for the analyses ([supplementary table S2, Supplementary Material](#) online).

### Evaluation of Identified Somatic Mutations

The overall strategy and criteria have been extensively tested across various plant taxa and were estimated previously to have a FPR <5% within callable regions when sufficient control samples ( $\geq 5$ ) are provided (Wang et al. 2019). For example, when calling BR-s mutations in one branch, all other branches serve as control samples, whereas when calling BR-m mutations exclusively present in leaves (i.e., situation 2), all root samples serve as controls. The analytical artifacts could thus be removed efficiently as the majority of them have the same chance to be found in focal and control samples (Li and Stoneking 2012).

Further, we assessed all 17 BR-m mutations and a random subset of 18 BR-s mutations with PCR amplification followed by Sanger sequencing (supplementary table S3, Supplementary Material online). Samples for validation were randomly picked from those with sufficient DNA. PCR amplification and Sanger sequencing were performed for each sample and each mutation independently. The VAFs of BR-m mutations were assessed using two approaches. The first approach, for each BR-m mutation in “A, C, D, F, H, J, K, L, N, P, Q,” including roughly ten mutated samples (samples supposed to carry a certain mutation) and four control samples, was PCR amplified (supplementary table S3 and fig. S4, Supplementary Material online). The PCR products (representing pooled cell populations) for each sample were Sanger sequenced independently, yielding a total of 110 and 41 analyzable results for mutated and control samples, respectively. The obtained Sanger chromatogram traces were subsequently decomposed using Indigo (<https://www.gear-genomics.com/indigo/>, last accessed September 30, 2021) to calculate the allelic fractions of mutations (Rausch et al. 2020). This approach verified all 11 BR-m mutations here and confirmed the mutations in 105 of the 110 samples (supplementary table S3 and fig. S4, Supplementary Material online). The five unconfirmed cases (two from mutation “A” and three from mutation “J”) were those with no reliable signal found for the mutation allele. These are not unexpected due to the low frequency presence of these mutations. Consistent with this, we also observed two cases where the mutation allele could be detected by PCR but not previously reported by WGS analysis (supplementary fig. S4, Supplementary Material online, mutation “K” and “L” in sample Cut22-L1 which has lowest average depth in leaves). All 41 control samples were also confirmed to not contain the mutations (supplementary table S3 and fig. S4, Supplementary Material online).

In second approach, the PCR products were first cloned using TA cloning (also known as rapid cloning) prior to Sanger sequencing. The PCR products amplified by DNA polymerase (P312-01; Vazyme Biotech Co.) were integrated into the pMD20-T vector (6028; Takara Bio Inc.) by ligation reaction. Then, the plasmids were transformed into competent cells of *Escherichia coli*. For each BR-m mutation in “B, D, E, F, I, J, M, N,” two or three mutated samples were PCR

amplified independently. For each sample, more than 32 monoclonal colonies were selected and validated by colony PCR. Positive clones were subsequently sequenced by Sanger reaction, yielding a total of 362 analyzable results from 20 mutated samples (supplementary table S3 and fig. S5, Supplementary Material online). The mutation alleles were verified in all 20 samples (at least one clone contains the mutation alleles for each sample), with 40 of 362 clones (i.e., overall mutation allelic fraction = 11.11%) contain the mutation alleles (supplementary fig. S5, Supplementary Material online).

For the remaining two BR-m mutations (“G” and “O”), no suitable primer is available (supplementary fig. S4, Supplementary Material online). We did not further assess those two mutations after failing several rounds of trials.

The FPR was estimated as:

$$\text{FPR} = 1 - \frac{\text{no. of validated mutations}}{\text{(no. of analyzable mutations for validation + 1)}}$$

assuming if we verify one more mutation by Sanger sequencing, it might be failed to validate.

The FNR was estimated using a simulation method described before (Keightley et al. 2015; Wang et al. 2019). Briefly, a total of 1,000 synthetic mutation sites were generated from the same sequencing data by replacing 1,000 randomized sites in them. The read-depth of each synthetic mutation was sampled from the distribution of the real mutations identified. The leaf and root samples were simulated separately considering the differences in their coverages and read-depth distributions. The FNR was calculated as:

$$\text{FNR} = \frac{\text{No. of uncallable mutations}}{\text{No. of valid synthetic mutations}}$$

Based on the simulation results, around 79.4% and 71.5% (calculated as “100% - FNR”) of the reference genome were estimated to be callable for leaf and root samples, respectively (supplementary table S6, Supplementary Material online). The normalized per site per sample somatic mutation rate ( $\tau$ ) is calculated as:

$$\tau = m / [G \times (1 - \text{FNR}) \times 2 \times S] \times (1 - \text{FPR}),$$

where  $m$  is the number of somatic mutations called in a certain organ,  $G$  is the haploid genome size, and  $S$  is the number of analyzed samples. Only those samples from cuttings with both leaf and root sequenced were analyzed here, hence  $S = 19$  for leaf samples,  $S = 21$  for root samples.

### Downstream Analysis

For prediction of effects of identified mutations, we used SnpEff version 4.0 (De Baets et al. 2012) based on the gene models of *S. suchowensis* (v4.1, also downloaded from the PopGenIE database). BM test was performed using R (R Development Core Team 2013) package “lawstat” (Hui et al. 2008).



Heterozygosity of the genome was estimated as  $H_f/G_i$ , where  $H_f$  is the number of fully heterozygous variants (i.e., variants with heterozygous genotype across all called samples, given as heterozygous genomic differences in [supplementary fig. S2, Supplementary Material](#) online) and  $G_i$  is the overall size of informative genomic regions used for variants calling, which is equivalent to the size of the analyzed genomic regions covered by no fewer than five reads ([supplementary table S1, Supplementary Material](#) online).

To assess whether somatic mutations are evenly distributed along the genome, we divided the whole genome in nonoverlapping 1 Mb windows and counted the mutation events for each window (observations). All windows were then tested using a Monte Carlo process, with 10,000 randomizations of shuffling all mutation events across the whole genome to derive the expectations. The unbiased estimation of empirical  $P$  (expected type I error rate) was derived as  $(n + 1)/(m + 1)$  for each window, where  $n$  is the number of seeing more in randomization than observation and  $m$  is the number of randomization (North et al. 2003). Regions with  $P < 0.05$  were defined as hotspot regions, which will indicate nonrandom occurrences of somatic mutations within these regions.

An ontogenetic tree was constructed using an approximately maximum-likelihood approach implemented by FastTree (Price et al. 2010) (version 2.1.10) with generalized time-reversible model. Interactive Tree Of Life (Letunic and Bork 2019) was used to annotate and display the tree. Plot of chromosome distributions was generated using Rldeogram package (Hao et al. 2019).

To derive the expected chance of leaf-exclusive or root-exclusive mutations under model 1, we consider “ $p$ ” as the real chance we can finally see a mutation in a particular sequenced sample (since  $p$  is a probability here,  $0 < p \leq 1$ ). Under model 1, the propagation of a mutation to each sample is completely random, so  $p$  is the same in both leaf and root (meaning a mutation has the same independent chance to be detected in any of the samples formed after this mutation). Considering the possibility that the lower coverage does lead to a lower chance to see a same mutation in root than in leaf, we can correct the root chance by multiplying the original chance with a factor “ $r$ ,” where  $0 < r < 1$ . Therefore, the overall chance we see an early mutation to be only detectable in  $n$  leaf samples but not detectable in  $m$  root samples is  $P = p^n \times (1 - rp)^m$ . Let's assume  $r = 0.2$ , which means a mutation could be detected with 5-fold higher chance in leaf than in root though leaf is generally only 2-fold higher ([supplementary table S1, Supplementary Material](#) online). Both  $n$  and  $m$  are constants, for example, for most of 15 BR-m leaf-exclusive mutations, we have  $n = 33$  and  $m = 21$ . By analyzing the derivative and the second derivative of  $P$ , we can know that when  $0 < p < \frac{n}{r(m+n)}$ ,  $P$  increases monotonically with  $p$ . For all leaf-exclusive mutations, we found  $\frac{n}{r(m+n)} > 1$ , hence the maximum possible value of  $p$  is 1. The maximum chance  $P$  could thus be obtained as  $P_{\max} = (1 - r)^m$ , which means that under this situation the absence of mutations in roots is solely due to the reduced detection power in roots. For most of 15 BR-m leaf-exclusive

mutations,  $P_{\max} = (1 - 0.2)^{21} = 9.22 \times 10^{-3}$ . The chance we see over ten such mutation sites is essentially near zero [ $(9.22 \times 10^{-3})^{10} = 4.46 \times 10^{-21}$ ]. This again suggested that these mutations cannot be explained by imbalance of mutation-detection power. The chances of other mutations could be derived using the same approach with different  $n$  and  $m$  numbers.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Grant Nos. 31970236, 91631104, 31970517, and 31900203). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

L.W., Y.J., D.T., and S.Y. planned and designed the research. Y.R., Z.H., P.L., and L.W. performed experiments, conducted fieldwork, and analyzed data. L.W., Y.J., S.Y., S.S., and B.T. wrote the manuscript. Y.R. and Z.H. contributed equally.

## Data Availability

The sequencing reads have been submitted to the National Center for Biotechnology Information (NCBI) under BioProject PRJNA497989. Codes for resequencing preprocessing, mutation calling as well as downstream analysis are available at GitHub page <https://github.com/ArthurSilver/willow-workflows> (last accessed September 30, 2021). Original Sanger sequencing results (in ab1 format) and other related miscellaneous data have been deposited in <https://doi.org/10.6084/m9.figshare.14746179>.

## References

- Amos W. 2010. Heterozygosity and mutation rate: evidence for an interaction and its implications. *Bioessays* 32(1):82–90.
- Antolin MF, Strobeck C. 1985. The population genetics of somatic mutation in plants. *Am Nat.* 126(1):52–62.
- Bellini C, Pacurar DI, Perrone I. 2014. Adventitious roots and lateral roots: similarities and differences. *Annu Rev Plant Biol.* 65:639–666.
- Benedict RC. 1923. New bud sports in nephrolepis. *Genetics* 8(1):75–95.
- Berggren B. 1984. Ultrastructure of the histological zones in growing vegetative buds of *Salix* spp. *Nord J Bot.* 4(6):771–790.
- Broertjes C, van Harten AM. 1985. Single cell origin of adventitious buds. *Euphytica* 34(1):93–95.
- Burian A, Barbier de Reuille P, Kuhlemeier C. 2016. Patterns of stem cell divisions contribute to plant longevity. *Curr Biol.* 26(11):1385–1394.
- Carlson MC. 1938. The formation of nodal adventitious roots in *Salix cordata*. *Am J Bot.* 25(9):721–725.
- Carlson MC. 1950. Nodal adventitious roots in willow stems of different ages. *Am J Bot.* 37(7):555–561.
- Chaparro JX, Werner DJ, Whetten RW, O'Malley DM. 1995. Characterization of an unstable anthocyanin phenotype and estimation of somatic mutation rates in peach. *J Hered.* 86(3):186–193.
- Clarke JD. 2009. Cetyltrimethyl ammonium bromide (CTAB) DNA mini-prep for plant DNA Isolation. *Cold Spring Harb Protoc.* 2009(3):pdb.prot5177.

- Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, Chen Y, Wan Z, Wang Z, et al. 2014. The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res.* 24(10):1274–1277.
- De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, Schymkowitz J, Rousseau F. 2012. SNPeff 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res.* 40(Database issue):D935–D939.
- De Rybel B, Mähönen AP, Helariutta Y, Weijers D. 2016. Plant vascular development: from early specification to differentiation. *Nat Rev Mol Cell Biol.* 17(1):30–40.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 43(5):491–498.
- Díaz-Sala C. 2014. Direct reprogramming of adult somatic cells toward adventitious root formation in forest tree species: the effect of the juvenile–adult transition. *Front Plant Sci.* 5:310.
- Diwan D, Komazaki S, Suzuki M, Nemoto N, Aita T, Satake A, Nishigaki K. 2014. Systematic genome sequence differences among leaf cells within individual trees. *BMC Genomics* 15:142.
- Evans MMS, Barton MK. 1997. Genetics of angiosperm shoot apical meristem development. *Annu Rev Plant Physiol Plant Mol Biol.* 48:673–701.
- Fink S. 1983. The occurrence of adventitious and preventitious buds within the bark of some temperate and tropical trees. *Am J Bot.* 70(4):532–542.
- Fjell I. 1985. Preformation of root primordia in shoots and root morphogenesis in *Salix viminalis*. *Nord J Bot.* 5(4):357–376.
- Fjell I. 1987. Structure and development of adventitious roots in *Salix caprea* and *S. caprea* × *S. viminalis*. *Nord J Bot.* 7(3):311–323.
- Foster TM, Aranzana MJ. 2018. Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Hortic Res.* 5:44–13.
- Frank MH, Chitwood DH. 2016. Plant chimeras: the good, the bad, and the 'Bizzaria'. *Dev Biol.* 419(1):41–53.
- Furner IJ, Pumfrey JE. 1992. Cell fate in the shoot apical meristem of *Arabidopsis thaliana*. *Development* 115(3):755–764.
- Groot EP, Laux T. 2016. Ageing: how do long-lived plants escape mutational meltdown? *Curr Biol.* 26(13):R530–R532.
- Haissig BE. 1970. Preformed adventitious root initiation in brittle willows grown in a controlled environment. *Can J Bot.* 48(12):2309–2312.
- Hanlon VCT, Otto SP, Aitken SN. 2019. Somatic mutations substantially increase the per-generation mutation rate in the conifer *Picea sitchensis*. *Evol Lett.* 3(4):348–358.
- Hao Z, Lv D, Ge Y, Shi J, Weijers D, Yu G, Chen J. 2019. RIdiogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. Available from: <https://peerj.com/preprints/27928>. Accessed September 30, 2021.
- Hofmeister BT, Denkena J, Colomé-Tatché M, Shahryary Y, Hazarika R, Grimwood J, Mamidi S, Jenkins J, Grabowski PP Sreedasyam A, et al. 2020. A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial *Populus trichocarpa*. *Genome Biol.* 21:259.
- Hui W, Gel YR, Gastwirth JL. 2008. lawstat: an R Package for law, public policy and biostatistics. *J Stat Softw.* 28:1–26.
- Ikeuchi M, Favero DS, Sakamoto Y, Iwase A, Coleman D, Rymer B, Sugimoto K. 2019. Molecular mechanisms of plant regeneration. *Annu Rev Plant Biol.* 70:377–406.
- Irish VF. 1991. Cell lineage in plant development. *Curr Opin Genet Dev.* 1(2):169–173.
- Irish VF, Sussex IM. 1992. A fate map of the *Arabidopsis* embryonic shoot apical meristem. *Development* 115(3):745–753.
- Jiang C, Mithani A, Gan X, Belfield EJ, Klingler JP, Zhu J-K, Ragoussis J, Mott R, Harberd NP. 2011. Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. *Curr Biol.* 21(16):1385–1390.
- Kebschull JM, Zador AM. 2018. Cellular barcoding: lineage tracing, screening and beyond. *Nat Methods.* 15(11):871–879.
- Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW, Jiggins CD. 2015. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Mol Biol Evol.* 32(1):239–243.
- Kitagawa M, Jackson D. 2019. Control of meristem size. *Annu Rev Plant Biol.* 70:269–291.
- Klekowski EJ, Kazarinova-Fukshansky N. 1984a. Shoot apical meristems and mutation: fixation of selectively neutral cell genotypes. *Am J Bot.* 71(1):22–27.
- Klekowski EJ, Kazarinova-Fukshansky N. 1984b. Shoot apical meristems and mutation: selective loss of disadvantageous cell genotypes. *Am J Bot.* 71(1):28–34.
- Klekowski EJ, Kazarinova-Fukshansky N, Mohr H. 1985. Shoot apical meristems and mutation: stratified meristems and angiosperm evolution. *Am J Bot.* 72(11):1788–1800.
- Klekowski EJ Jr, Godfrey PJ. 1989. Ageing and mutation in plants. *Nature* 340(6232):389–391.
- Lanfear R. 2018. Do plants have a segregated germline? *PLoS Biol.* 16(5):e2005439.
- Laux T. 2003. The stem cell concept in plants: a matter of debate. *Cell* 113(3):281–283.
- Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47(W1):W256–W259.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv13033997 Q-Bio [Internet]. Available from: <http://arxiv.org/abs/1303.3997>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Li M, Stoneking M. 2012. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol.* 13(5):R34.
- Lucas WJ, Groover A, Lichtenberger R, Furuta K, Yadav S-R, Helariutta Y, He X-Q, Fukuda H, Kang J, Brady SM, et al. 2013. The plant vascular system: evolution, development and functions. *J Integr Plant Biol.* 55(4):294–388.
- Marcotrigiano M, Stewart RN. 1984. All variegated plants are not chimeras. *Science* 223(4635):505–505.
- Michel A, Arias RS, Scheffler BE, Duke SO, Netherland M, Dayan FE. 2004. Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). *Mol Ecol.* 13(10):3229–3237.
- Nieminen K, Blomster T, Helariutta Y, Mähönen AP. 2015. Vascular cambium development. *Arab Book [Internet]*. Available from: <https://bioone.org/journals/the-arabidopsis-book/volume-2015/issue-13/tab.0177/Vascular-Cambium-Development/10.1199/tab.0177.full>. Accessed September 30, 2021.
- North BV, Curtis D, Sham PC. 2003. A note on the calculation of empirical *P* values from Monte Carlo procedures. *Am J Hum Genet.* 72(2):498–499.
- O'Connell LM, Ritland K. 2004. somatic mutations at microsatellite loci in western Redcedar (*Thuja plicata*: cupressaceae). *J Hered.* 95(2):172–176.
- Orr AJ, Padovan A, Kainer D, Külheim C, Bromham L, Bustos-Segura C, Foley W, Haff T, Hsieh J-F, Morales-Suarez A, et al. 2020. A phylogenomic approach reveals a low somatic mutation rate in a long-lived plant. *Proc Biol Sci.* 287(1922):20192364.
- Phillips RL, Kaeppeler SM, Olhoft P. 1994. Genetic instability of plant tissue cultures: breakdown of normal controls. *Proc Natl Acad Sci U S A.* 91(12):5222–5226.
- Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonnet N, Labadie K, Le Provost G, et al. 2018. Oak genome reveals facets of long lifespan. *Nat Plants.* 4(7):440–452.
- Poethig RS. 1987. Clonal analysis of cell lineage patterns in plant development. *Am J Bot.* 74(4):581–594.
- Poethig S. 1989. Genetic mosaics and cell lineage analysis in plants. *Trends Genet.* 5(8):273–277.

- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5(3):e9490.
- R Development Core Team. 2013. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available from: <http://www.R-project.org>. Accessed September 30, 2021.
- Rahni R, Efroni I, Birnbaum KD. 2016. A case for distributed control of local stem cell behavior in plants. *Dev Cell*. 38(6):635–642.
- Rausch T, Fritz MH-Y, Untergasser A, Benes V. 2020. Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files. *BMC Genomics* 21(1):230.
- Roest S, Berkel MAE, van Bokelmann GS, Broertjes C. 1981. The use of an in vitro adventitious bud technique for mutation breeding of *Begonia x hiemalis*. *Euphytica* 30(2):381–388.
- Schmid-Siegert E, Sarkar N, Iseli C, Calderon S, Gouhier-Darimont C, Chrast J, Cattaneo P, Schütz F, Farinelli L, Pagni M, et al. 2017. Low number of fixed somatic mutations in a long-lived oak tree. *Nat Plants*. 3(12):926–929.
- Sena G, Wang X, Liu H-Y, Hofhuis H, Birnbaum KD. 2009. Organ regeneration does not require a functional stem cell niche in plants. *Nature* 457(7233):1150–1153.
- Sennerby-Forsse L, Zuffa L. 1995. Bud structure and resprouting in coppiced stools of *Salix viminalis* L., *S. eriocephala* Michx., and *S. amygdaloides* Anders. *Trees* 9(4):224–234.
- Shamel AD, Pomeroy CS. 1936. Bud mutations in horticultural crops. *J Hered*. 27(12):487–494.
- Simberloff D, Leppanen C. 2019. Plant somatic mutations in nature conferring insect and herbicide resistance. *Pest Manag Sci*. 75(1):14–17.
- Steffens B, Rasmussen A. 2016. The physiology of adventitious roots. *Plant Physiol*. 170(2):603–617.
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A*. 109(45):18488–18492.
- Suomela J, Ayres MP. 1994. Within-tree and among-tree variation in leaf characteristics of mountain birch and its implications for herbivory. *Oikos* 70(2):212–222.
- Sutherland WJ, Watkinson AR. 1986. Somatic mutation: do plants evolve differently? *Nature* 320(6060):305–305.
- Szymkowiak EJ, Sussex IM. 1996. What chimeras can tell us about plant development. *Annu Rev Plant Physiol Plant Mol Biol*. 47:351–376.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14(2):178–192.
- Volk TA, Verwijst T, Tharakan PJ, Abrahamson LP, White EH. 2004. Growing fuel: a sustainability assessment of willow biomass crops. *Front Ecol Environ*. 2(8):411–418.
- Wang B, Smith SM, Li J. 2018. Genetic regulation of shoot architecture. *Annu Rev Plant Biol*. 69:437–468.
- Wang L, Ji Y, Hu Y, Hu H, Jia X, Jiang M, Zhang X, Zhao L, Zhang Y, Jia Y, et al. 2019. The architecture of intra-organism mutation rate variation in plants. *PLoS Biol*. 17(4):e3000191.
- Wang Y, Jiao Y. 2018. Axillary meristem initiation—a way to branch out. *Curr Opin Plant Biol*. 41:61–66.
- 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*. *Proc Natl Acad Sci U S A*. 113(43):12226–12231.
- Weismann A. 1892. Das Keimplasma: eine Theorie der Vererbung. 1. Auflage. Jena (Germany): Fischer.
- Whitham TG. 1983. Host manipulation of parasites: within-plant variation as a defense against rapidly evolving pests [Insect pests, *Pemphigus* gall aphids, plant genotypes, *Populus angustifolia*]. Available from: <http://agris.fao.org/agris-search/search.do?recordID=US8294052>. Accessed September 30, 2021.
- Whitham TG, Slobodchikoff CN. 1981. Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability: the adaptive significance of somatic mutations in plants. *Oecologia* 49(3):287–292.
- Woodworth MB, Girsakis KM, Walsh CA. 2017. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat Rev Genet*. 18(4):230–244.
- Xie Z, Wang L, Wang L, Wang Z, Lu Z, Tian D, Yang S, Hurst LD. 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 R Soc B*. 283:20161016.
- Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen J-Q, Hurst LD, Tian D. 2015. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* 523(7561):463–467.
- Yong E. 2012. Tree's leaves genetically different from its roots. *Nat News [Internet]*. Available from: <http://www.nature.com/news/tree-leaves-genetically-different-from-its-roots-1.11156>. Accessed September 30, 2021.
- Yu L, Boström C, Franzenburg S, Bayer T, Dagan T, Reusch TBH. 2020. Somatic genetic drift and multilevel selection in a clonal seagrass. *Nat Ecol Evol*. 4:952–962.
- Zhang D, Wang Z, Wang N, Gao Y, Liu Y, Wu Y, Bai Y, Zhang Z, Lin X, Dong Y, et al. 2014. Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. *PLoS One* 9(5):e96879.
- Zhang L, Vijg J. 2018. Somatic Mutagenesis in mammals and its implications for human disease and aging. *Annu Rev Genet*. 52:397–419.