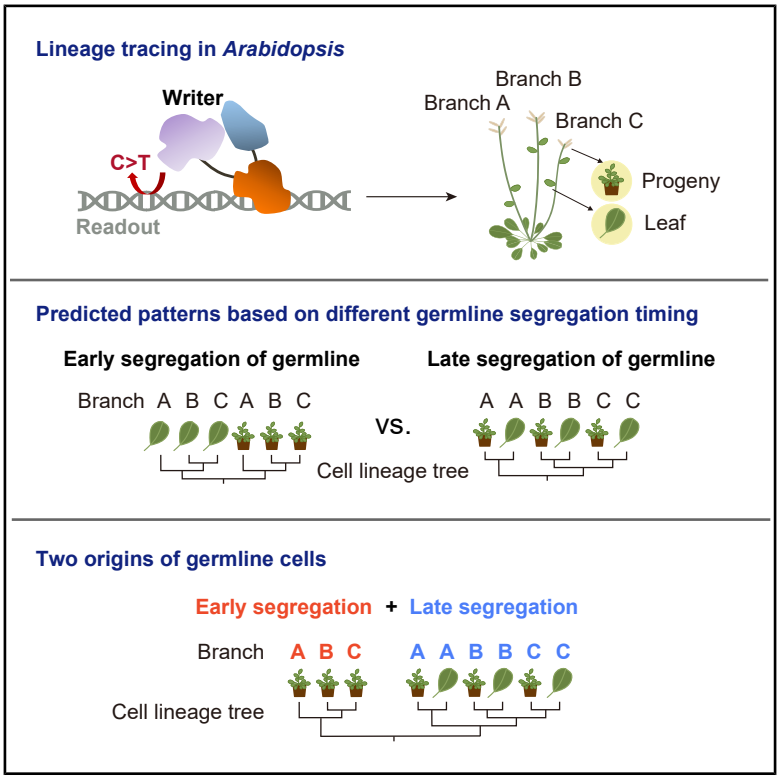
Article

Current Biology

Testing Weismann's germ plasm theory in *Arabidopsis*

Graphical abstract



Highlights

- Dynamic editing-based lineage tracing is applied to *Arabidopsis*
- Some progenies from different plant branches cluster in the cell lineage tree
- Green progenies can arise from albino siliques in the *Atmsh1* mutant
- Late-arising mutations are infrequently detected in progeny

Authors

Haotian Guo, Geyu Zhang, Lu Gao, ..., Li Liu, Qing Huan, Wenfeng Qian

Correspondence

qhuan@genetics.ac.cn (Q.H.), wfqian@genetics.ac.cn (W.Q.)

In brief

Guo et al. employ a dynamic editing approach to trace cell lineages in *Arabidopsis*. Their findings reveal two origins of germline cells: some germline cells segregate during inflorescence meristem formation, while others segregate earlier than branch formation—akin to the Weismann barrier traditionally associated with animal development.

Guo et al., 2026, Current Biology *36*, 1–14
April 20, 2026
https://doi.org/10.1016/j.cub.2026.03.003

CellPress

Current Biology

CellPress

Article

Testing Weismann's germ plasm theory in *Arabidopsis*

Haotian Guo,[1,2,5] Geyu Zhang,[1,2,5] Lu Gao,[1,2,5] Yang Liu,[1] Yu Wang,[1,2] Zhan Liu,[3] Chunmei Guan,[4] Chun Zhang,[1,2] Li Liu,[3] Qing Huan,[1,*] and Wenfeng Qian[1,2,6,*]

[1]State Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China

[4]State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Peking University, Beijing 100871, China

[5]These authors contributed equally

[6]Lead contact

*Correspondence: qhuan@genetics.ac.cn (Q.H.), wfqian@genetics.ac.cn (W.Q.)

https://doi.org/10.1016/j.cub.2026.03.003

SUMMARY

Weismann's germ plasm theory proposed that germline cells are set aside early in development, ensuring that mutations arising in somatic tissues during an organism's lifetime are not inherited. While this principle has been well supported in animals, it has not been rigorously tested in plants. Plants differ fundamentally from animals in that their meristematic cells drive continuous growth and organ formation throughout life, leading to the long-held view that the plant germline segregates late. Here, we used a dynamic genome-editing lineage tracing system to construct cell lineages in *Arabidopsis thaliana*, including both somatic and germline cells. Our analysis of the cell lineage tree revealed two distinct germline segregation patterns. While some germline cells clustered with somatic cells from their branch of origin (consistent with late segregation), others from different branches shared a recent common ancestry (indicative of early segregation). This supports a dual-origin model for germline cells in *A. thaliana*: early-segregated germlines represent a plant counterpart to Weismann's barrier, reducing the risk of transmitting excessive mutations across generations, whereas late-segregated germlines can inherit beneficial mutations acquired during development, potentially facilitating adaptation.

INTRODUCTION

As multicellular organisms develop, they inevitably acquire somatic mutations over time, posing a genetic burden because some of these mutations are deleterious. August Weismann's germ plasm theory, proposed in 1892, posited that only germ cells (eggs and sperm) transmit genetic information to the next generation, whereas somatic mutations represent an evolutionary dead end.[1] This idea was later confirmed in many animals, such as *Caenorhabditis elegans*, mice, and humans, in which the germline is set aside early in development.[2–4] Weismann's theory laid foundational concepts for modern genetics and evolutionary biology because it separates heredity from environmental influences, paving the way for understanding inheritance through stable genetic materials.

However, multicellularity has evolved independently in plants and animals, leading to distinct developmental strategies.[5] Rooted and immobile, plants adapt to environmental changes through developmental plasticity enabled by meristematic cells.[6,7] These cells facilitate continuous growth and new organ formation throughout their lives, contrasting with the rapid, early organ development commitment seen in animals.[8] Because plants establish their reproductive organs relatively late, it is widely believed that segregation of the germline occurs only after the transition from vegetative to reproductive growth. This idea challenges the applicability of Weismann's germ plasm theory to plants.

Indeed, it has been reported that mutations accumulated during the vegetative growth phase could be transmitted to progeny in plants.[9–11] For example, a study showed that between 2% and 6% of detected somatic mutations in *Arabidopsis*, rice, or *Brachypodium* were inherited by the progeny.[9] This rate of transmission was surprisingly higher in tree species. For instance, research indicated that 47% of detected somatic mutations in oak[11] and 52% in peach[9] were also detected in subsequent generations. These observations raise a fundamental question: how do plants, which inevitably accumulate mutations during their vegetative growth, manage to prevent excessive, environmentally induced, and potentially harmful mutations from overwhelming their progeny?

One proposed answer is the concept of a functional germline.[12] One study found that the number of DNA replications leading to gametes in *A. thaliana* was not correlated with lifespan, indicating that older plants don't necessarily pass on more mutations to progeny than younger plants do.[13] Another study demonstrated that apical meristems in plants remained relatively inactive, undergoing only seven to nine cell divisions before forming an axillary meristem.[14] This quiescence contrasts

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

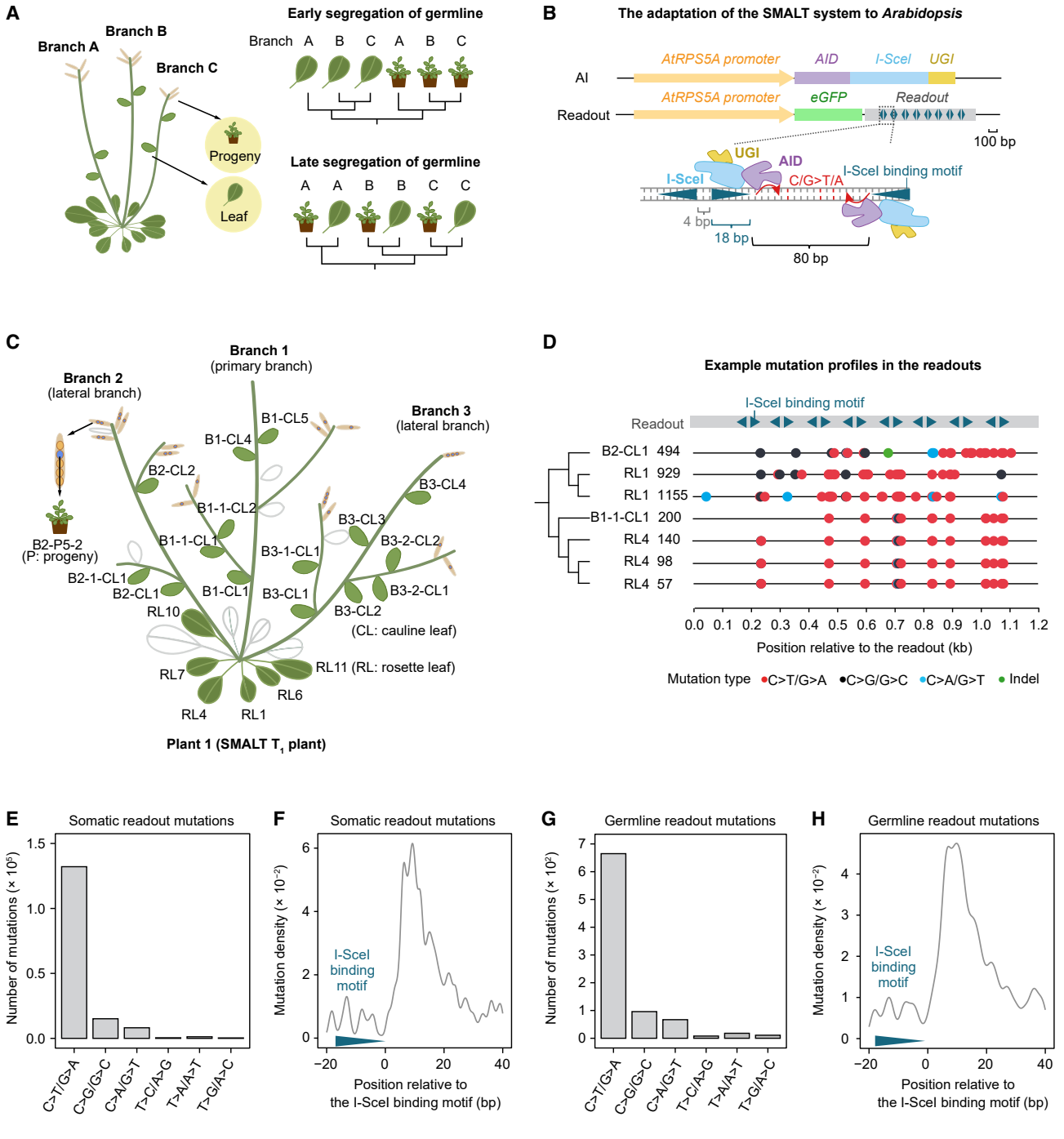


Figure 1. Application of the SMALT system to trace cell lineage in *A. thaliana*

(A) Schematic cell lineage trees contrasting early and late segregation of germlines in plants. If germline cells segregate earlier than branch (primordium) formation, leaves (representing somatic tissues) and progeny (representing parental germline) should form their respective clusters in the cell lineage tree, independent of their branch origin. Conversely, if germline segregation occurs after branch formation, leaves and progeny should cluster within their originating branches.

(B) Diagram illustrating the adaptation of the SMALT system to *Arabidopsis*. The SMALT system is composed of two components, the "writer" AI and the readout. AI is a fusion protein, consisting of AID (activation-induced cytidine deaminase), I-SceI (an intron-encoded endonuclease devoid of catalytic activity), and UGI. The readout is composed of 16 I-SceI binding motifs (labeled in blue triangles in a back-to-back arrangement), with 80-bp editable sequences between each pair of facing motifs. The readout serves as the 3′ untranslated region of *eGFP* (enhanced green fluorescent protein), leveraging single-stranded DNA regions to enhance editing efficiency. Both *AI* and *eGFP* were driven by the *A. thaliana RPS5A* promoter.

(C) The SMALT system was introduced into *A. thaliana* through *Agrobacterium*-mediated transformation, and 21 leaves (labeled with black letters) were collected from the T_1 plant. Sample labeling follows these conventions: B, branch; CL, cauline leaf; RL, rosette leaf; and P, progeny. Cauline leaf sample IDs consist of a

(legend continued on next page)

with the vigorous division of neighboring cells, helping to slow mutation accumulation in axillary buds.[14] Although these slowly dividing stem cells can later produce both reproductive and somatic tissues (and thus differ from an early-segregated germline), they effectively restrict the number of mutations transmitted to the next generation.[12,15]

However, the presence of a functional germline does not exclude the possibility that plants also possess an early-segregated germline (yet not differentiated), similar to animals.[12] Such cells may have escaped notice because they lack obvious morphological features and cannot be identified under a microscope. Without seeing them, an early-segregated germline can be informed by the mutation patterns observed in progeny from various branches.[12] If germline cells segregate early and stay in individual axillary meristems, then progenies (representing mutation patterns of their parental germline cells) from different branches should share highly similar mutation patterns, pointing to a common ancestral lineage that predates branch formation. By contrast, if the germline segregates late, progeny would share more mutations with the somatic tissues of their own branch (Figure 1A), implying a recent shared lineage with nearby somatic cells.

Tracing cell lineages through somatic mutations in plants is challenging because they occur infrequently and are difficult to detect.[9,16–18] Recently, dynamic editing approaches—combining genome-editing technologies with high-throughput sequencing—have enabled the creation of vast arrays of evolving sequences connected through common mutations.[19–23] This provides a powerful, detailed view of developmental processes across broad temporal and spatial scales. Such strategies have also been applied to plants. For instance, one study demonstrated that regenerated whole plants can originate from a single somatic progenitor cell, suggesting a competitive bottleneck during regeneration.[24] Another, using a high-resolution dynamic editing system, revealed that the somatic tissues of each branch in *Arabidopsis* develop from exactly three founder cells, establishing a new model for plant organogenesis.[25]

Here, we applied a dynamic editing system, originally developed in *Drosophila melanogaster*,[26] to *A. thaliana*. We built a high-resolution mutation map that links somatic tissues to progeny, allowing us to test directly when germline segregation occurs. We further explored if plants possess an early-segregated germline using the *A. thaliana msh1* mutant. We identify a dual-origin germline in *A. thaliana*—early-segregated and late-segregated lineages. Our findings reveal a developmental mechanism by which plants balance the inheritance of new genetic variation with the safeguarding of genomic integrity across generations.

RESULTS

Advancing cell lineage tracing in *A. thaliana* with the SMALT system

To construct a high-resolution cell lineage tree, we enhanced the local mutation rate by adapting the SMALT system established in *Drosophila* to *A. thaliana*. Briefly, the system comprises two components: the "writer" protein AI and a 1.2-kb synthetic DNA sequence for recording mutations, referred to as the "readout." AI is a fusion protein of an activation-induced cytidine deaminase (which can introduce C>T or G>A mutations in the readout sequence), an inactive version of intron-encoded endonuclease I-SceI (which can recognize and bind to the I-SceI binding motifs in the readout), and a UGI (uracil-DNA glycosylase inhibitor to increase C>T mutation rate). The readout contained 16 I-SceI binding motifs separated by editable sequences, which together comprised 595 editable C and G bases—providing a vast genotype space for high-resolution cell lineage recording (Figures 1B and S1A).

Both the *AI* and the readout (which was in the 3′ untranslated region of an enhanced green fluorescent protein [*eGFP*]) were expressed from the *A. thaliana RPS5A* promoter (Figures 1B and S1A). This promoter is active from the embryo stage onward and persists in proliferating cells[27–30] (Figures S1B–S1D), enabling mutation recording across the entire developmental timeline. The AI and the readout constructs were introduced into *A. thaliana* using *Agrobacterium*-mediated transformation (STAR Methods). Mutations emerge in the readout during plant development and will be passed to daughter cells during cell division. Based on readout mutations in individual cells, the cell lineage tree can be computationally constructed.

To confirm that the mutations indeed occur in the readout as expected, we collected DNA from 6 rosette leaves and 15 cauline leaves from three branches of a T_1 plant (named "plant 1"; Figures 1C and S2A), representing a diverse spatial distribution of somatic tissues. In this plant, branch 1 was primary, emerging directly from the shoot apical meristem, whereas branches 2 and 3 were lateral, developing from axillary meristems associated with subtending rosette leaves. For precise genotyping of readouts of individual cells, we utilized UMIC-seq, a method introducing a unique molecular identifier (UMI) to each initial template of a readout during polymerase chain reaction (PCR).[31] This method obtained consensus sequences for

branch ID and a leaf ID, separated by a hyphen. Progeny sample IDs start with the branch ID, followed by a progeny ID prefixed with "P," which includes a silique ID (numbered successively ascending a branch starting from "1"), and a seed ID within that silique, all separated by hyphens. Seeds used for analysis are highlighted in blue within their corresponding siliques.

(D) Example mutation profiles in the readout, with their sample IDs and UMIs labeled. Each line represents a readout sequence, with mutations indicated by marks proportional to their position along the readout, and colors correspond to mutation types. The tree constructed from these mutation profiles is displayed on the left.

(E) Histogram showing the molecular spectrum of mutations identified in the readout sequences in somatic tissues.

(F) Plot showing the density of mutations identified in somatic tissue in relation to the position of the I-SceI binding motif. Distances were adjusted according to the orientation of the binding motif.

(G and H) Similar to (E) and (F), parental germline mutations identified from progeny samples.

See also Figures S1 and S2 and Table S5.

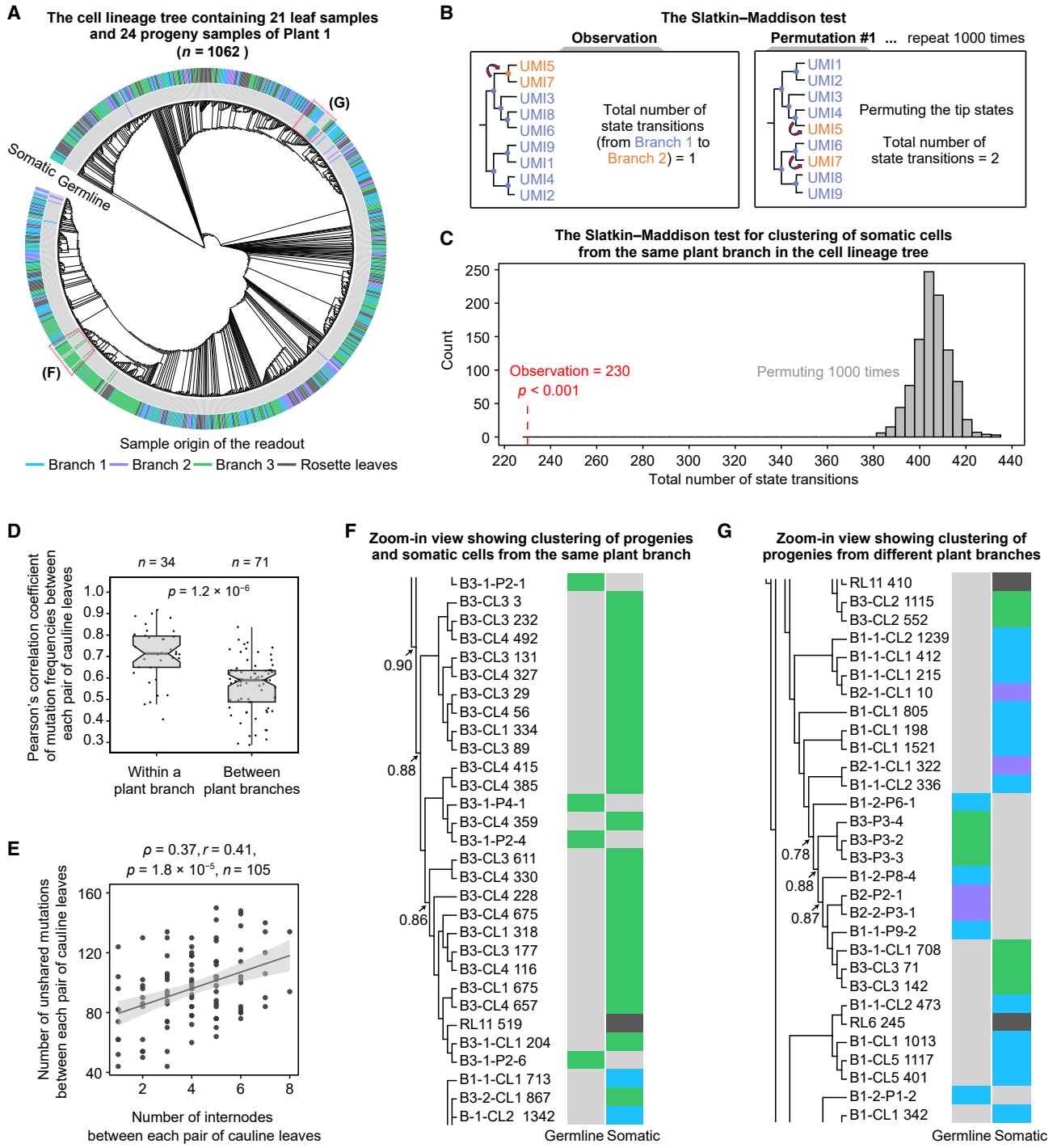


Figure 2. Cell lineage tree including both leaf and progeny samples

(A) Maximum-likelihood tree constructed based on readout sequences from 21 leaf samples and 24 progeny samples of plant 1. A total of 6,688 somatic readout sequences were downsampled to 1,038 sequences to balance the readout number from different somatic samples. Together with 24 germline readouts, a total of 1,062 sequences were used to construct the cell lineage tree. The colors in the outer ring indicate the branch origin of somatic readouts, and those in the inner ring indicate the branch origin of germline readouts.

(B) Schematic of the Slatkin-Maddison test permutation procedure. The observed minimum number of state (branch origin) transitions (left) was compared with expected values with tip states randomly permuted, keeping the tree topology unchanged (right).

(legend continued on next page)

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

each readout from dozens of Oxford Nanopore Technologies (ONT) reads with an identical UMI, thereby minimizing sequencing and PCR errors (Figures S1E and S1F).

We mapped the readout sequences to the reference readout sequence (Figure 1D) and identified a total of 165,681 somatic mutations in plant 1. The majority were C>T/G>A (84%; Figure 1E) and located in the proximal downstream regions of I-SceI binding motifs (Figure 1F), confirming the application of SMALT in plants. This plant SMALT system yielded an average of 25 mutations per readout (Figure S2B), validating the efficacy of AI editing. That 84% of the readouts were unique (i.e., each was detected only once in the entire plant; Figure S2C) implies editing persisted until late developmental stages, enabling a detailed view of the developmental dynamics in *A. thaliana*. Note that recurrent mutations ("hotspots")—those likely arising independently multiple times—confound phylogenetic analysis by suggesting false relationships; therefore, we identified and excluded them from all analyses (STAR Methods).

In *Arabidopsis*, seeds contain an embryo (which develops from the zygote), endosperm, and seed coat. Only mutations that occur in the zygote are passed to the embryo and thus to the next generation, and these are termed germline mutations. To identify germline mutations, we gathered 24 self-fertilized seeds containing the readout from all three branches and extracted genomic DNA from their whole seedlings (Figure 1C). To account for ongoing AI editing events in progeny, we performed TA cloning of individual readouts followed by Sanger sequencing, obtaining an average of 8.9 readout sequences per seedling. The multiple readout sequences allowed identification of common mutations as their parental germline mutations (Figure S2D; STAR Methods). In total, we identified 947 germline mutations, and these mutations were also consistent with the pattern expected for the SMALT system (Figures 1G, 1H, and S2E). We were then able to deduce the original germline readout sequence (Figures S2D and S2E) for constructing the cell lineage tree.

Some progenies from different plant branches clustered in the cell lineage tree

To determine germline segregation timing for *A. thaliana*, we constructed a cell lineage tree based on both somatic and germline readouts, using an approximately maximum-likelihood method (FastTree[32]). Since the number of somatic readouts varied among the 21 somatic samples (n = 41 to 1,300), to balance the readout number from different samples for tree construction, we randomly downsampled 50 readout sequences for each of the 21 somatic samples (when the readout number in the sample exceeded 50). A cell lineage tree was constructed for plant 1, including a total of 1,062 readouts (Figures 2A and S2F).

We first focused on somatic cells in the cell lineage tree. To investigate if somatic cells from the same *A. thaliana* branch tend to cluster in the cell lineage tree, we employed the Slatkin-Maddison test,[33] which first calculates the minimum number of state transitions (i.e., branch switches) required on the cell lineage tree to account for the observed distribution of readouts across the three branches and then compares this value to the expected number estimated from 1,000 times permutation (Figure 2B). The observed tree required 230 transitions, while the expected number ranged between 380 and 440 (Figure 2C), indicating a statistically significant clustering of somatic cells from the same branch ($p < 0.001$).

We then computed mutation-frequency profiles for each cauline leaf and found that correlations were significantly higher within a branch than between branches (Figure 2D; $p = 1.2 \times 10^{-6}$, Mann-Whitney test). We also observed a positive relationship between the number of unshared mutations in a leaf pair and the number of internodes separating them (Figure 2E). These observations are expected under a local founder-cell growth pattern from the shoot tip, where axillary primordia are locally recruited and then expand clonally, producing higher similarity among leaves of the same branch and greater divergence across branches and with each additional internode.

We then moved our focus to the relationship between somatic and germline readouts. Fifteen germline readouts grouped together with the somatic readouts from their originating plant branch (Figures 2A and 2F, as an example), supporting late segregation of some germline cells in *Arabidopsis*. Notably, the cell lineage tree also revealed that eight germline readouts clustered together ($p < 0.001$, Slatkin-Maddison test; Figures 2A and S2G), and these germline readouts were from all three plant branches (Figure 2G). These observations support early segregation of at least some germline cells in *Arabidopsis*.

Mutations causing the clustering of progenies from different plant branches are rare in leaves

Having excluded recurrent hotspot mutations that could create false cell lineage signals, we further tested whether progenies from distinct branches shared more mutations than expected by chance. To do this, we used somatic readouts to establish a baseline expectation for random mutation sharing, which was then compared with the number of observed germline

(C) Distribution of state transition numbers (i.e., the minimum total number of branch switches required to explain the distribution of cells across branches in the cell lineage tree) from 1,000 permutations. The observed state transition number (indicated by the red dashed line) was significantly smaller than expected by chance ($p < 0.001$, the Slatkin-Maddison test).

(D) Boxplot comparison of Pearson's correlation coefficient for mutation frequencies between two samples within a plant branch versus between branches (the Mann-Whitney test). For each sample, the frequency of a given mutation was calculated as the proportion of readouts containing that mutation relative to total readouts.

(E) Relationship between the number of unshared mutations for a cauline leaf pair and the number of internodes separating them. Each dot represents a pair of leaves. For each leaf sample, the 100 mutations with the highest frequencies were selected, and mutations present in only one of the two samples were considered as unshared mutations. The linear fit (solid line) and 95% confidence interval (shaded area) are shown.

(F and G) Zoomed-in view of the cell lineage tree for the two regions boxed in red in (A), showing clustering of progeny and somatic cells from the same plant branch (F) or clustering of progenies from different plant branches (G). Bootstrap support values are provided for the cell lineage branches leading to the cluster. See also Figures S2, S4, and S5 and Tables S5 and S6.

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

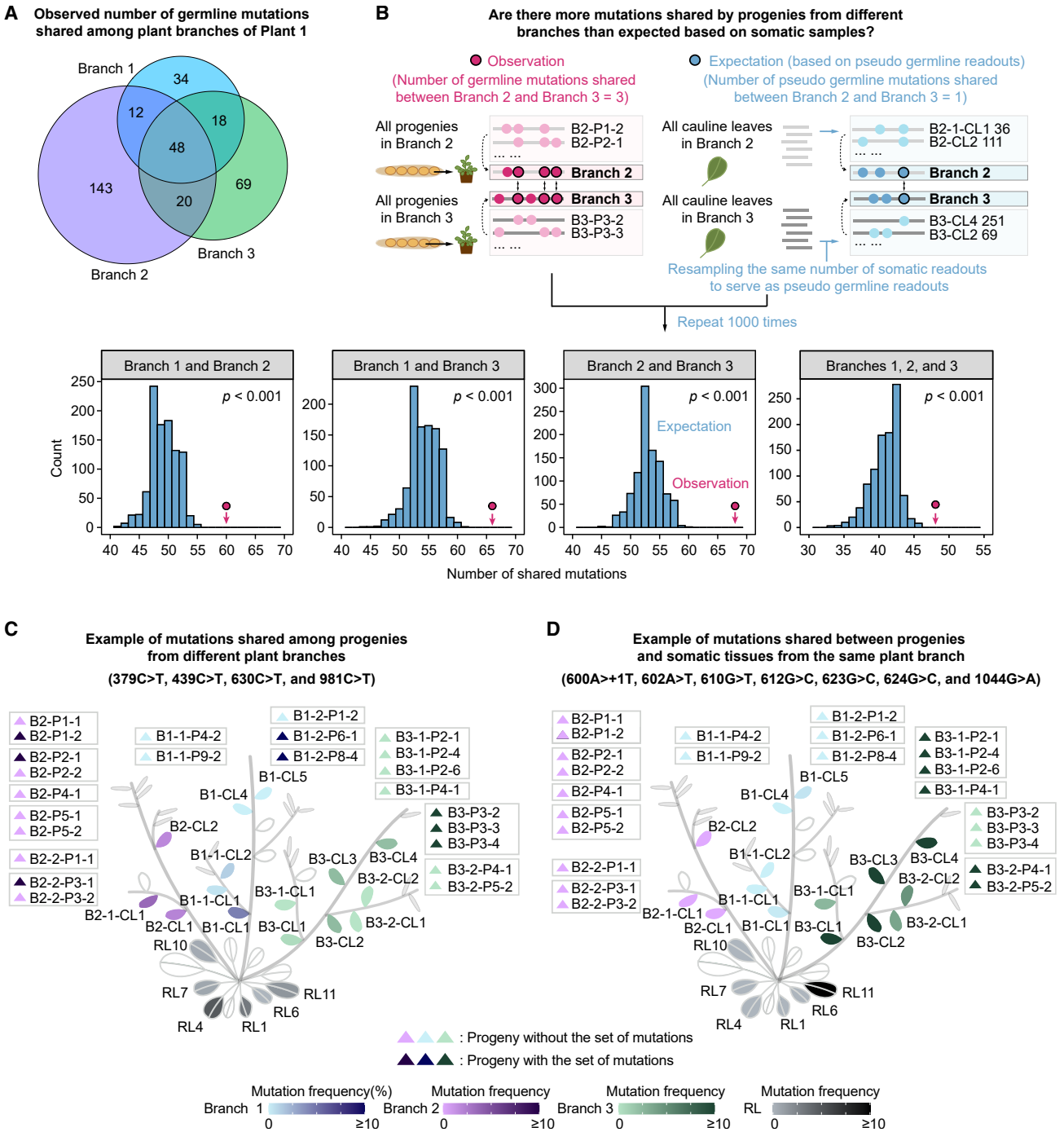


Figure 3. Mutations shared among progenies and leaves

(A) Venn diagram showing the observed number of germline mutations from three branches of plant 1.

(B) Observed and expected numbers of germline mutations shared among different branches and their expectation. Upper panels illustrate the sampling strategy used to estimate both observation and expectation. All germline mutations from progeny were combined to represent germline mutations in a branch (pink dots). Germline mutations from different branches were then compared, with shared mutations indicated by connected black circled dots. We sampled cauline leaves from the same plant branch, while controlling for mutation load, to serve as pseudo germline readouts. From these, pseudo germline mutations were identified and plotted (blue dots). Histograms in the lower panels show the distribution of the number of shared mutations from 1,000 times of resampling (expectation) for the indicated branch combinations, with the observed value indicated by a pink arrow. *p* values were estimated based on this resampling.

(legend continued on next page)

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

mutations (Figure 3A). We simulated pseudo germline readouts by randomly sampling somatic readouts from their respective branches while preserving the original mutation count distribution (Figure S2H). This sampling process was repeated 1,000 times, and for each repeat, we quantified the number of shared mutations across branches to generate an expectation. We then constructed a distribution of these expected shared mutation counts and compared it to the empirically observed number of shared mutations among actual progenies. The observed value significantly exceeded the expectation (resampling test, $p < 0.001$; Figure 3B), indicating that the clustering of germline readouts is not due to chance.

Indeed, a set of four C>T/G>A mutations was present in eight of the 24 progenies across all three branches yet appeared in a frequency less than 4.9% across all 21 leaf samples (Figure 3C). Meanwhile, mutations shared between progeny and somatic tissues within the same branch were also observed. For example, a group of seven mutations was detected in six of the nine branch 3 progenies and was also found in five somatic samples from the same branch, with frequencies ranging from 13.8% to 33.5% (Figure 3D). Together, these mutation profiles clearly distinguish the cellular origins of progenies from either early- or late-segregated germlines.

Similar cell lineage pattern from two additional *Arabidopsis* plants

To test whether the early segregation of some germline cells observed in plant 1 could be repeatedly detected, we repeated the experiment in two additional *A. thaliana* plants transformed with the SMALT system (plants 2 and 3; Figures S3A and S3B). We sequenced readouts from 37 leaves and 55 progenies across four branches of plant 2 and from 26 leaves and 38 progenies across three branches of plant 3 (Figures S3C–S3F). Among the mutations identified—2,184 somatic and 255 germline in plant 2 and 86,634 somatic and 336 germline in plant 3—the majority were C>T/G>A substitutions. We constructed cell lineage trees from both somatic and progeny readout sequences. In plant 2, 11 progenies from all four branches formed a strongly supported cluster ($p < 0.001$, Slatkin-Maddison test; Figures 4A–4C). In plant 3, six progenies from all three branches also clustered ($p < 0.001$, Slatkin-Maddison test; Figures 4D–4F). These results demonstrate that early-segregated germline lineages can be repeatedly observed.

Green progenies were observed from albino siliques in *A. thaliana msh1* mutant

To visualize early-segregated germline lineages more directly, we used an *A. thaliana* mutant deficient in the *MSH1* gene, which is required for chloroplast DNA mismatch repair.[34] Loss of *MSH1* function can lead to loss of chloroplast genomes, resulting in an albino phenotype, which can therefore serve as a visual lineage tracing marker.[35,36] We cultivated hundreds of *msh1* plants to identify individuals with albino sectors and then tested whether green progeny could arise from these sectors. Because the albino phenotype is irreversible—once chloroplast genomes are lost, they cannot be restored—the occurrence of green progeny within albino sectors would provide clear evidence for the existence of early-segregated germline lineages.

In one case, we observed an axillary branch where one half, including a flower and its subtending cauline leaf, was albino while the other half remained green (Figure 5A). This pattern suggests that the cauline leaf primordium and its associated axillary meristem—giving rise to the branch's somatic tissues—originated from two (mesophyll) progenitor cells, one albino and one green. After self-pollination, we obtained both an albino and a green silique from this branch. All 11 germinated seeds collected from the green silique grew into seedlings with green cotyledons. Remarkably, four out of 14 germinated seeds from the albino silique also showed a green phenotype, indicating that some seeds were derived from a germline lineage distinct from the stem cells in the inflorescence meristem that produced the albino somatic tissues (Figure 5B).

In another *msh1* plant, we found an albino phenotype restricted to siliques 3, 11, and 19 (an ~52.5° sector) of a branch containing 21 siliques (Figure 5C). This observation indicates a progenitor cell with the albino phenotype underwent successive divisions to produce spatially adjacent albino siliques within this sector. After self-pollination, seeds from siliques 3 and 11 were either non-germinating or albino, whereas 35 of 38 germinated seeds from silique 19 developed into green seedlings (Figure 5C). A similar pattern of green progeny arising from albino siliques was observed in a third *msh1* plant, which displayed an albino sector spanning ~180° (Figure 5D). In all cases, the contrast between the albino somatic tissue of siliques and the green progeny derived from them indicates that some germline cell lineages had segregated prior to the formation of the albino sector in the axillary primordium.

Limited transmission of late-arising mutations to progeny in *Arabidopsis*

A previous study reported that 2%–6% of mutations detected in somatic tissues are inherited by progeny in *Arabidopsis*, rice, or *Brachypodium*.[9] To determine if the transmission of mutations in our SMALT plants is comparable to that of sparsely distributed genomic mutations, we estimated the transmission rate—defined as the proportion of unique mutations (counted once per plant) passed to progeny. In plant 1, we detected 3,715 unique mutations in leaf samples, of which only 280 were detected in progeny samples (Figure 6A), giving a transmission rate of 7.5%. Similar rates were observed in plants 2 and 3 (8.7% and 7.2%, respectively; Figure 6A). These values are consistent with the previous report based on genomic mutations,[9] suggesting that the basic rules of mutation transmission to progeny are not substantially altered by the implementation of the SMALT system.

(C) Example of mutations shared among progenies from different branches of plant 1. A schematic plant shows the relative positions of leaves and progenies. Branch identity is indicated by color. The presence of the mutation combination (shown on the top) in progeny is indicated by dark triangles, and its frequencies in leaves are presented by shading intensity.
(D) Example of mutations shared between progeny and somatic samples from the same branch in plant 1. The figure is otherwise similar to (C).
See also Figure S2.

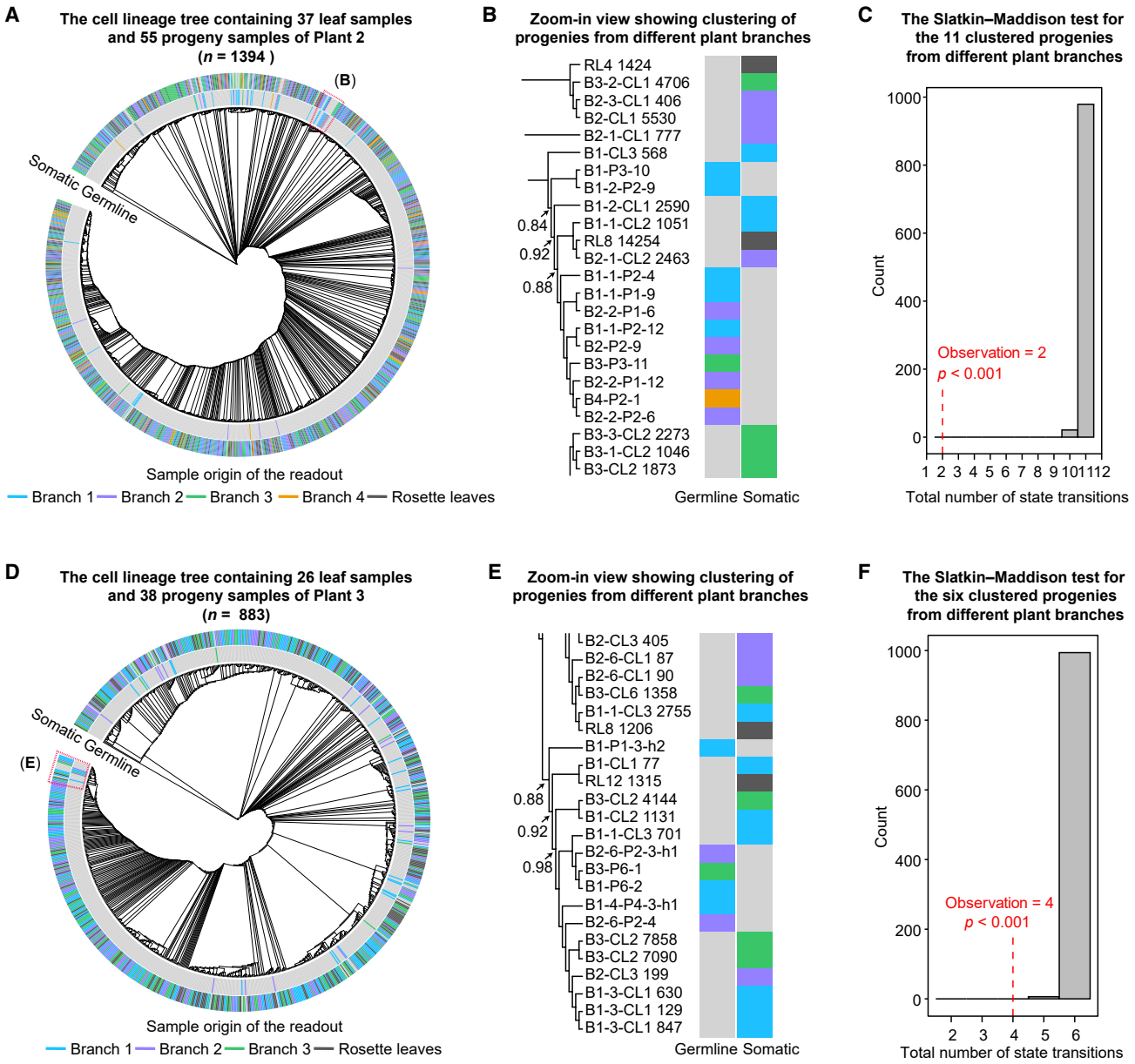


Figure 4. Cell lineage structure of somatic tissues and progeny in plants 2 and 3

(A) Cell lineage tree constructed based on readout sequences from 37 leaf samples and 55 progeny samples of plant 2. 1,339 sequences were sampled from a total of 1,721 readout sequences to balance the readout number from different somatic samples. The labeling scheme for the cell lineage tree is consistent with that used in Figure 2A.

(B) Zoomed-in view of the cell lineage tree from the region boxed in red in (A), showing clustering of progenies from different plant branches. Bootstrap support values are provided for the branches leading to the cell lineage clusters.

(C) Distribution of state transition numbers from 1,000 permutations for the clustered 11 progenies from different branches in the cell lineage tree of plant 2 ($p < 0.001$, the Slatkin-Maddison test). The observed transition number is indicated by the red dashed line.

(D–F) Similar to (A)–(C), for plant 3. Maximum-likelihood tree constructed based on readout sequences from 26 leaf samples and 38 progeny samples. 837 sequences were sampled from a total of 8,679 readout sequences to balance the readout number from different branches of plant 3.

See also Figures S3–S5 and Tables S5 and S6.

Examining this in more detail, among the unique 344 mutations identified in progeny of plant 1, the majority ($n = 248$) were also present in both rosette and cauline leaves (Figure 6A), and they were early-arising mutations broadly distributed across the apical meristem. To test whether transmission is biased toward such early-arising variants, we classified parental mutations in plant 1 into three groups: branch specific, shared by two branches, or shared by all three branches. Branch-specific mutations showed the lowest transmission rate (1.1%), whereas mutations shared by all three branches had the highest

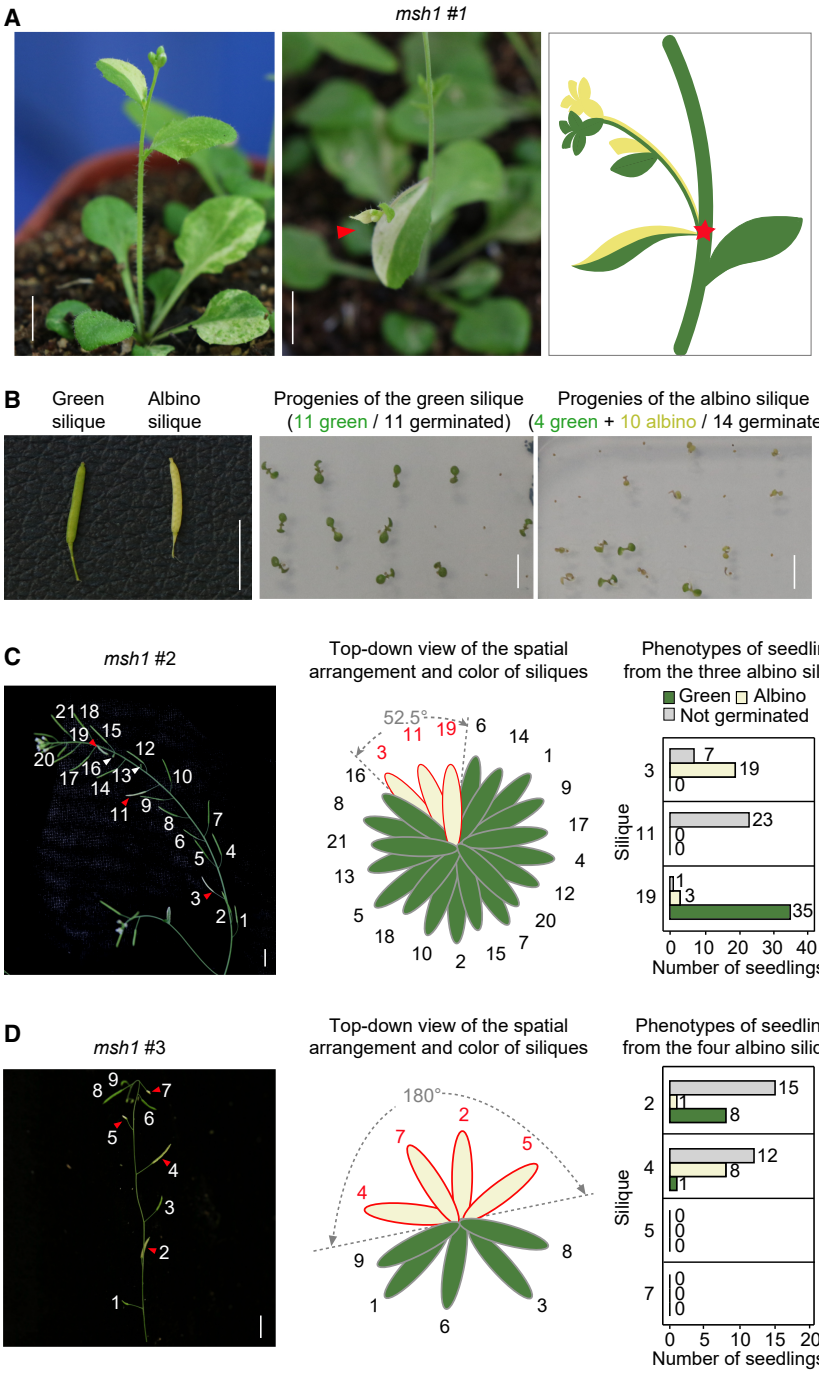


Figure 5. Albino phenotypic characterization of the *msh1* plants

(A) Albino phenotype of *msh1* #1 in a branch including cauline leaves, flowers, stems, and its subtending leaf. Albino sectors are indicated by red arrows and illustrated schematically in the rightmost panel. A red star marks the position of two progenitor cells exhibiting phenotypic divergence.

(B) Albino phenotype of *msh1* #1 in two siliques and all seedlings grown from their seeds (when germinated).

(C) Albino phenotypes of siliques and progenies on a single branch of *msh1* #2. The temporal order of silique formation is indicated by the numbers labeled in the left and middle panels. The middle panel represents a top-down view of silique distribution, showing the spatial arrangement of albino siliques forming an approximate 52.5° angle. Bar plots in the right panel show the phenotypes of seedlings from the three albino siliques.

(D) Albino phenotypes of siliques and progeny on a single branch of *msh1* #3. The figure is otherwise similar to (C).

All bars in this figure represent 1 cm.

See also Table S3.

transmission rate (29.3%; Figure 6B). A similar pattern was observed in plants 2 and 3 (Figure 6B). Together, these findings indicate that late-arising mutations contribute little to the genetic makeup of progeny in *Arabidopsis*.

DISCUSSION

The potential inheritance of somatic mutations poses a puzzle in plants: if every mutation accumulated during their lifetime could pass to the progeny, the genetic burden would soon become unsustainable.[37–39] Our study shows that *Arabidopsis* protects future generations with a dual strategy (Figure 7A). On one hand, some germline cells arise from slowly dividing meristematic stem cells—the functional germline. These cells inherit some of the mutations accumulated during growth, but the number is probably small due to fewer cell divisions in these cells.[13,40,41] On the other hand, we discovered a second source of germline: a set of cells that segregate early, before branch formation, and contribute primarily to germline tissues. Progeny derived from these cells largely escape somatic mutations accumulated during the parent's vegetative growth. This second source is strikingly similar to the early-segregated germline of animals, providing experimental support for a plant counterpart to Weismann's germ plasm theory (Figure 7B). Interestingly, the presence of an early-segregated germline has also recently been reported in the fungal kingdom.[42] Although multicellularity has independently evolved in animals and plants, they seemingly used a similar strategy for safeguarding the genomic integrity of the next generation.

This mixed strategy potentially offers plants evolutionary advantages. The functional germline allows different branches to produce progeny with distinct genetic profiles. In this way, a single plant can "test" new somatic mutations across its branches,

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

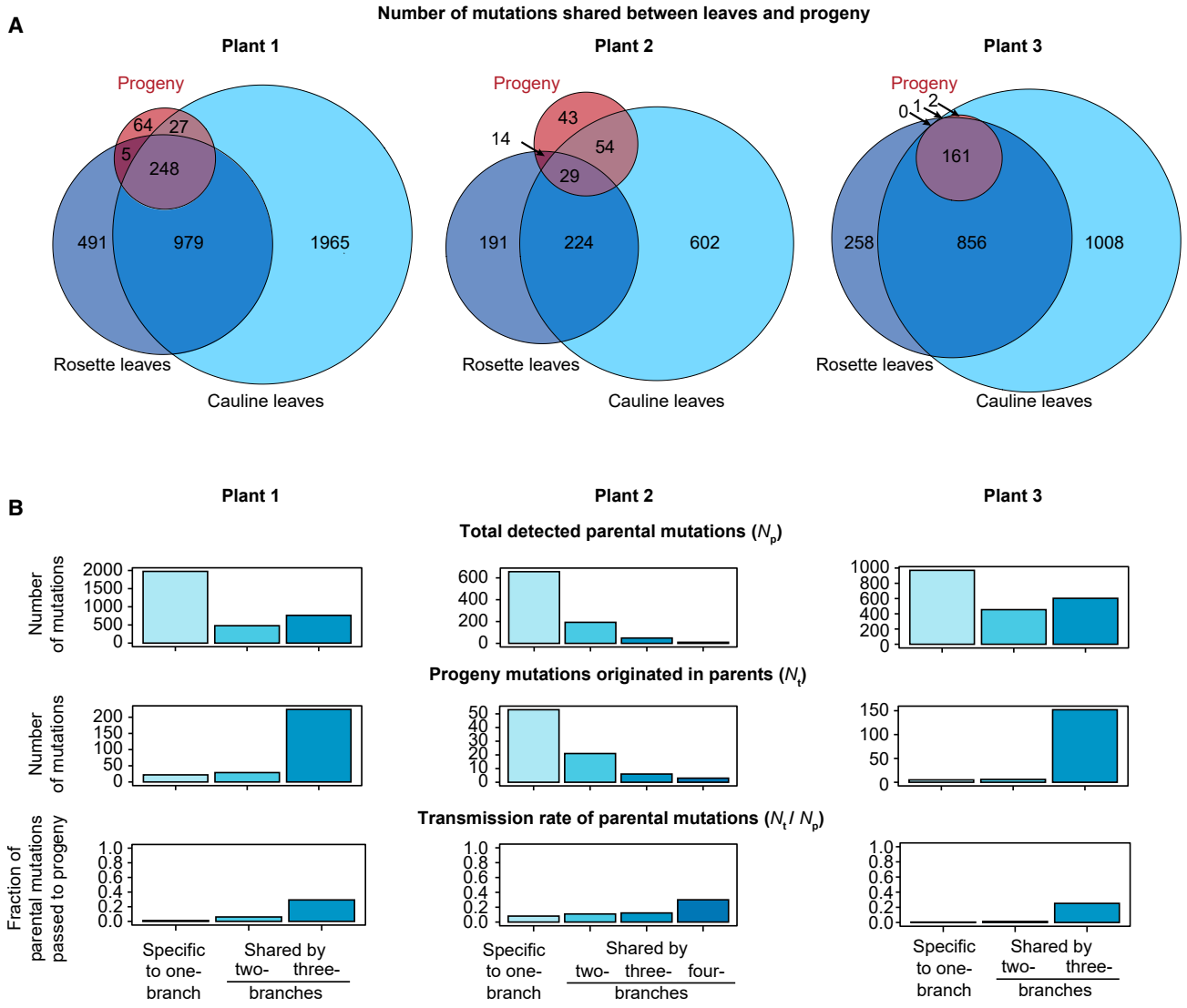


Figure 6. Transmission of parental mutations to progeny in *A. thaliana*

(A) Venn diagrams illustrating the overlap of unique mutations among the rosette leaves, cauline leaves, and progeny for each plant (plants 1, 2, or 3).

(B) Transmission rate of parental mutations to progeny. In each panel, bars from left to right represent parental mutations specific to a single branch and those shared across two, three, or all four branches for each plant.

and those carrying advantageous mutations contribute disproportionately to the next generation. This enables natural selection to act at a sub-individual level. Meanwhile, the early-segregated germline ensures that at least some progeny inherit a "clean" genome, free from most environmentally induced damage. Such a system may be especially important for long-lived species like trees.[18,43–45] By relying on ancient, slowly dividing cells, trees can still produce healthy progeny despite their age. This may also help explain the slower evolutionary rates observed in trees compared with herbaceous plants, a pattern that has been reported as a violation of the molecular clock in plants.[46,47] This finding also suggests that while each meristem could be considered an individual, making the plant analogous to a population,[48,49] the early-segregated germline we report provides a genetic linkage between meristems, indicating they are more closely related than members of a typical population.

What types of cells might constitute this early-segregated germline? We propose that they may reside in the central, slowly dividing region of the apical meristem. When a new axillary meristem forms, these cells may undergo a few divisions, leaving daughter cells in each newly established meristem, while vigorous division of other cells spatially separates the apical meristem and the axillary meristem.[14] These cells stay largely quiescent (and undifferentiated) until they are recruited into reproductive development, thereby minimizing mutation accumulation from both DNA replication errors and environmental damage.[13,50] It would be valuable for future studies to examine the proliferation dynamics, branching behavior, and fate specification of these cells.

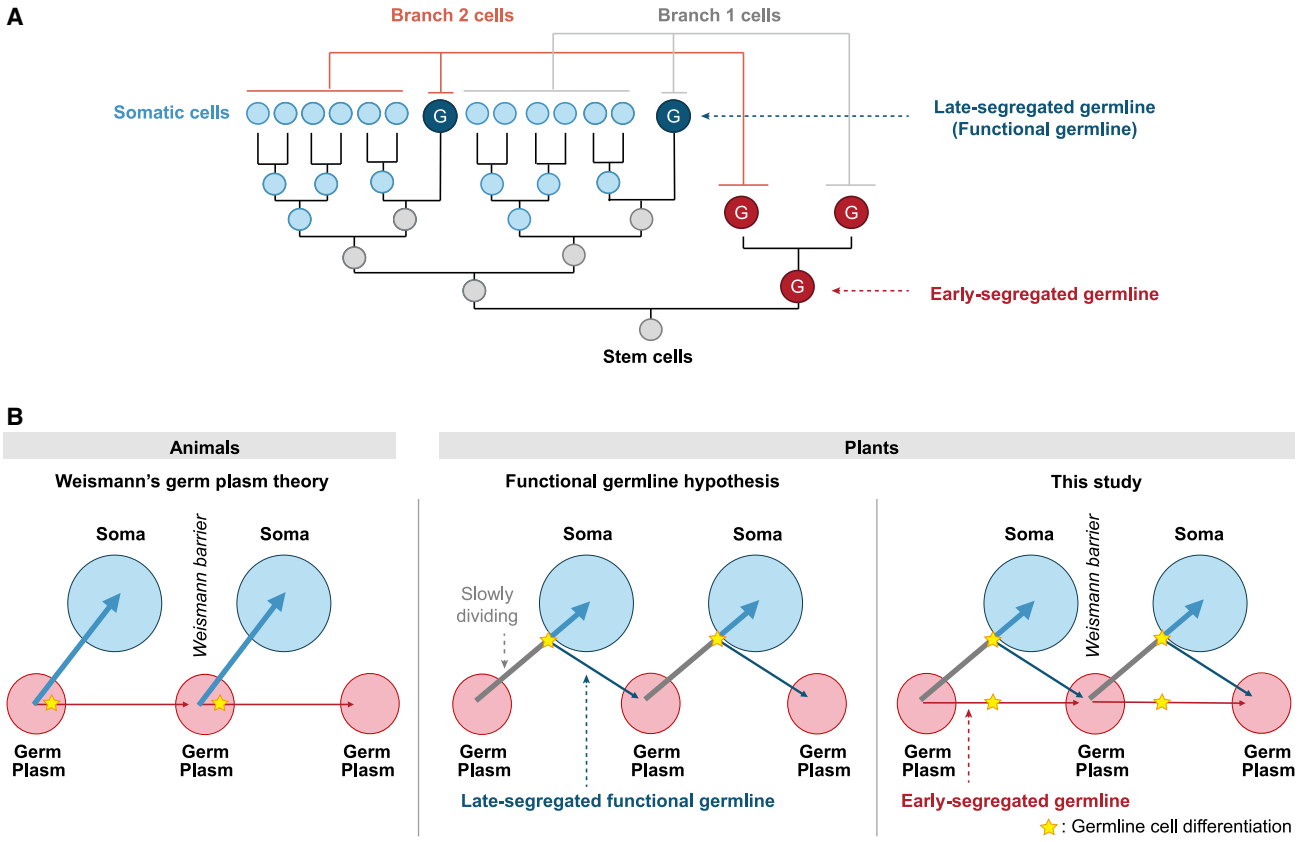


Figure 7. Dual origins of germline cells in *Arabidopsis thaliana*

(A) Lineage diagram showing two possible sources of germline in *A. thaliana*. Functional germlines proposed in a previous study and early-segregated germlines reported in this study are labeled.

(B) Models of germline segregation in animals and plants. In animals (left), the germline is segregated early from somatic lineages, limiting somatic mutation transmission to progeny (i.e., the Weismann barrier). Previous plant studies (middle) indicated the presence of late-segregated functional germline. This study (right) shows that plants possess an early-segregated germline, revealing a mechanistic parallel to the Weismann barrier in animals.

It has been proposed that although germline differentiation generally occurs late in plants, segregation from somatic lineages can take place much earlier.[12] Our data show precisely this for a subset of germline lineages. Unlike what is reported in many animals, where germline and somatic lineages are strictly early separated (and differentiated), *A. thaliana* appears to maintain a more permeable boundary between them. Consequently, early-segregated germline lineages can occasionally contribute to somatic tissues—the lineage cluster containing progenies from different branches also included somatic cells (Figure 2G). Nevertheless, these lineages are distinct from germline lineages that cluster within the somatic clade of their branch of origin (Figure 2F), which is more consistent with later segregation events.

It is also important to ask whether early- and late-segregated germline cells contribute differently to siliques positioned at various levels of the plant architecture. While our data do not allow a definitive conclusion, they may offer tentative insights into this question. We observed cases in which both early- and late-segregated germlines contributed to the same silique. Among siliques with readout sequences available in at least two seeds, mixed contributions occurred in 1/3 of siliques in plant 1, 9/14 in plant 2, and 5/11 in plant 3, indicating that siliques cannot be simply classified as "early" or "late" by germline origin. Moreover, we found no significant differences in the distribution between early- and late-segregated progenies among different branches (chi-square tests; Figures S4A and S4B), and the lack of statistical significance was also observed when we compared different branch levels—including primary branches, secondary branches arising from the axils of rosette leaves or cauline leaves on the primary branch, and tertiary branches originating from the axils of cauline leaves on secondary branches (Figures S4C and S4D). These results indicate that positional context alone is insufficient to predict germline origin, and plants integrate germline contributions from different origins through a complex mechanism.

Our study has several limitations that warrant consideration. First, it remains unclear whether the SMALT system remained active until the final cell divisions leading to seed formation. Although not emphasized in the results, we detected 1,724 post-germination somatic mutations in progeny seedlings (T_2) of plant 1 through Sanger sequencing, with a predominance of C>T/G>A substitutions (91%) in their molecular spectrum (Figure S2E). A similar pattern was also observed in plants 2

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

and 3 (Figure S3G), suggesting that SMALT activity persisted even during T_2 development. Second, endoreduplication—commonly observed in mesophyll cells of many plants, including *A. thaliana*[51,52]—could potentially affect mutation detection. However, we believe this phenomenon is unlikely to alter our central finding that some progenies from different branches cluster together, because endoreduplication increases DNA copy number within individual cells but does not create shared mutations across independent lineages. Third, plant 2 showed a relatively low mutation rate, potentially due to unidentified epigenetic factors or stochastic variation in mutation events. Nevertheless, clustering of progenies from different plant 2 branches was still observed (Figures 4A and 4B), supporting the conclusion that early-segregated germline cells exist in *A. thaliana*.

Another limitation of our study concerns the model used to construct the cell lineage tree. FastTree uses the Jukes-Cantor nucleotide substitution model, which assumes equal substitution rates for C>T (or G>A) and T>C (or A>G) mutations, i.e., a time-reversible process. This assumption may introduce errors because, in the SMALT system, C>T (or G>A) mutations occur far more frequently than their reverse. Although not highlighted in the results, we also used *Startle*,[53] a program designed for CRISPR-based lineage tracing that assumes irreversible mutation rates, to construct lineage trees including both somatic tissues and progeny. The clustering of progenies from different branches was preserved (Figure S5), indicating that the time-reversible assumption in FastTree did not affect our main conclusions. We chose to present FastTree results because it additionally provides confidence measures via bootstrap values. Nonetheless, future work should develop computational methods tailored to dynamic editing-based lineage tracing systems that inherently violate the time-reversibility assumptions underlying conventional phylogenetic approaches designed for natural species.

Like many plants, *A. thaliana* has three developmental layers—L1, L2, and L3—giving rise predominantly to epidermal, mesophyll, and vascular tissues, respectively, and the germline is generally thought to derive from L2.[54,55] Because we sequenced whole leaves, each sample pooled cells from all three layers; however, mutation calling was performed at the level of individual cells (UMIs), so the mixture should not affect mutation detection. Even so, cell lineage trees built from leaf readouts (Figures 2A, 4A, and 4D) did not show a simple three-layer partition. This is expected if some mutations occurred before layer specification (e.g., an octant-stage mutation would appear in a subset of cells of all three layers). Future work that couples cell lineage tracing with cell-type mapping—such as using single-cell transcriptomics on the same samples—should resolve layer-specific contributions to branch architecture and germline origin.

Our study opens several avenues for future research. While our work focused on *A. thaliana*, the findings may not be universally applicable given the wide developmental diversity among plant species. Future studies should extend dynamic cell lineage tracing to a broader range of plants to validate and expand these results. In addition, it will be important to identify, visualize, and trace the specific meristematic cells that segregate early and ultimately give rise to germ cells. Together, these efforts will deepen our understanding of plant development, environmental adaptation, and genomic evolution.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Wenfeng Qian (wfqian@genetics.ac.cn).

Materials availability

Materials generated in this study will be made available on request.

Data and code availability

- The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive[56,57] in the National Genomics Data Center, China National Center for Bioinformation, Chinese Academy of Sciences (CRA029430, publicly accessible at https://ngdc.cncb.ac.cn/gsa), and the plasmid sequence data reported in this paper have been deposited in GenBase (GB0003227 for *pRPS5A*-driven AI plasmid and GB0003228 for *pRPS5A*-driven Readout plasmid) and are publicly accessible at https://ngdc.cncb.ac.cn/genbase.
- All original code and resulting data have been deposited at Zenodo (10.5281/zenodo.17052105) and at GitHub (https://github.com/ZhangGeyu/Pipeline-of-lineage-tracing-in-Arabidopsis).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Prof. Xionglei He from Sun Yat-Sen University, Prof. Qikun Liu and Prof. Yuling Jiao from Peking University, Prof. Danhua Jiang and Prof. Bo Ren from the Institute of Genetics and Developmental Biology CAS for discussion, and Prof. Xiaofeng Cao from the Institute of Genetics and Developmental Biology CAS for providing plasmids XF675 and XF4363. We thank Dr. Kejia Shan for experiment design and Weiguang Wang, Chao Chen, Hongbin Ren, Shuo Zhang, Yawen Liu, and Bingke Jiao from the Institute of Genetics and Developmental Biology CAS for assisting with data analysis. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB1090202), the Biological Breeding-National Science and Technology Major Project (2023ZD0407303), and the Project for Young Scientists in Basic Research of the Chinese Academy of Sciences (YSBR-078).

AUTHOR CONTRIBUTIONS

W.Q., Q.H., H.G., and G.Z. designed the study. H.G., L.G., Y.L., Q.H., Z.L., C.G., and L.L. performed experiments. G.Z., H.G., Y.W., and C.Z. analyzed data. W.Q., Q.H., H.G., G.Z., and L.G. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plants
- METHOD DETAILS
  - Plasmid construction
  - Plant transformation and sampling
  - UMIC-seq and Oxford Nanopore sequencing
  - Acquisition of progeny readout
  - Genome sequencing of the SMALT T_1 plants
  - Live imaging
  - Albino phenotypes in the *msh1* mutants
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data analysis of the SMALT T_1 plant genomes

- Identification of consensus sequences and *bona fide* mutations from UMIC-seq data
- Mutation identification from progeny
- Identification of recurrent (hotspot) mutations
- Construction of cell lineage trees
- Slatkin–Maddison test
- Resampling test for mutations shared by multiple branches

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2026.03.003.

Received: October 30, 2025
Revised: January 31, 2026
Accepted: March 2, 2026

REFERENCES

1. Weismann, A. (1892). Das Keimplasma: eine Theorie der Vererbung (G. Fischer) (Gustav Fischer).
2. Wolf, N., Priess, J., and Hirsh, D. (1983). Segregation of germline granules in early embryos of *Caenorhabditis elegans*: an electron microscopic analysis. J. Embryol. Exp. Morphol. *73*, 297–306. https://doi.org/10.1242/dev.73.1.297.
3. Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A signaling principle for the specification of the germ cell lineage in mice. Cell *137*, 571–584. https://doi.org/10.1016/j.cell.2009.03.014.
4. Castillo-Venzor, A., Penfold, C.A., Morgan, M.D., Tang, W.W., Kobayashi, T., Wong, F.C., Bergmann, S., Slatery, E., Boroviak, T.E., Marioni, J.C., et al. (2023). Origin and segregation of the human germline. Life Sci. Alliance *6*, e202201706. https://doi.org/10.26508/lsa.202201706.
5. Niklas, K.J. (2014). The evolutionary-developmental origins of multicellularity. Am. J. Bot. *101*, 6–25. https://doi.org/10.3732/ajb.1300314.
6. de Jong, M., and Leyser, O. (2012). Developmental plasticity in plants. Cold Spring Harb. Symp. Quant. Biol. *77*, 63–73. https://doi.org/10.1101/sqb.2012.77.014720.
7. Reinhardt, D., and Kuhlemeier, C. (2002). Phyllotaxis in higher plants. In Meristematic Tissues in Plant Growth and Development (Sheffield Academic Press), pp. 172–212.
8. Hong, L., and Fletcher, J.C. (2023). Stem cells: engines of plant growth and development. Int. J. Mol. Sci. *24*, 14889. https://doi.org/10.3390/ijms241914889.
9. 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*, e3000191. https://doi.org/10.1371/journal.pbio.3000191.
10. Klekowski, E.J. (1988). Mutation, Developmental Selection, and Plant Evolution (Columbia University Press). https://doi.org/10.7312/klek92068.
11. Plomion, C., Aury, J.M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., et al. (2018). Oak genome reveals facets of long lifespan. Nat. Plants *4*, 440–452. https://doi.org/10.1038/s41477-018-0172-3.
12. Lanfear, R. (2018). Do plants have a segregated germline? PLoS Biol. *16*, e2005439. https://doi.org/10.1371/journal.pbio.2005439.
13. Watson, J.M., Platzer, A., Kazda, A., Akimcheva, S., Valuchova, S., Nizhynska, V., Nordborg, M., and Riha, K. (2016). Germline replications and somatic mutation accumulation are independent of vegetative life span in *Arabidopsis*. Proc. Natl. Acad. Sci. USA. *113*, 12226–12231. https://doi.org/10.1073/pnas.1609686113.
14. Burian, A., Barbier de Reuille, P., and Kuhlemeier, C. (2016). Patterns of stem cell divisions contribute to plant longevity. Curr. Biol. *26*, 1385–1394. https://doi.org/10.1016/j.cub.2016.03.067.
15. Burian, A. (2021). Does shoot apical meristem function as the germline in safeguarding against excess of mutations? Front. Plant Sci. *12*, 707740. https://doi.org/10.3389/fpls.2021.707740.
16. Orr, A.J., 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*, 20192364. https://doi.org/10.1098/rspb.2019.2364.
17. Ren, Y., He, Z., Liu, P., Traw, B., Sun, S., Tian, D., Yang, S., Jia, Y., and Wang, L. (2021). Somatic mutation analysis in *Salix suchowensis* reveals early-segregated cell lineages. Mol. Biol. Evol. *38*, 5292–5308. https://doi.org/10.1093/molbev/msab286.
18. 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*, 926–929. https://doi.org/10.1038/s41477-017-0066-9.
19. McKenna, A., and Gagnon, J.A. (2019). Recording development with single cell dynamic lineage tracing. Development *146*, dev169730. https://doi.org/10.1242/dev.169730.
20. Deng, S., and He, X. (2023). Tree of life at two levels: from species to cell. Sci. Bull. (Beijing) *68*, 2515–2518. https://doi.org/10.1016/j.scib.2023.09.018.
21. McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science *353*, aaf7907. https://doi.org/10.1126/science.aaf7907.
22. Raj, B., Gagnon, J.A., and Schier, A.F. (2018). Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT. Nat. Protoc. *13*, 2685–2713. https://doi.org/10.1038/s41596-018-0058-x.
23. Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E., et al. (2020). An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. Cell *181*, 1410–1422.e27. https://doi.org/10.1016/j.cell.2020.04.048.
24. Lu, X., Zhang, Q., Wang, Z., Cheng, X., Yan, H., Cai, S., Zhang, H., and Liu, Q. (2025). Development of an inducible DNA barcoding system to understand lineage changes in *Arabidopsis* regeneration. Dev. Cell *60*, 305–319.e5. https://doi.org/10.1016/j.devcel.2024.10.023.
25. 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. Preprint at Research Square. https://doi.org/10.21203/rs.3.rs-5203004/v1.
26. Liu, K., Deng, S., Ye, C., Yao, Z., Wang, J., Gong, H., Liu, L., and He, X. (2021). Mapping single-cell-resolution cell phylogeny reveals cell population dynamics during organ development. Nat. Methods *18*, 1506–1514. https://doi.org/10.1038/s41592-021-01325-x.
27. Zhang, T.Q., Chen, Y., and Wang, J.W. (2021). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. Dev. Cell *56*, 1056–1074.e8. https://doi.org/10.1016/j.devcel.2021.02.021.
28. Tian, C., Zhang, X., He, J., Yu, H., Wang, Y., Shi, B., Han, Y., Wang, G., Feng, X., Zhang, C., et al. (2014). An organ boundary-enriched gene regulatory network uncovers regulatory hierarchies underlying axillary meristem initiation. Mol. Syst. Biol. *10*, 755. https://doi.org/10.15252/msb.20145470.
29. Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., Cumming, M., Kelley, L.A., Sternberg, M.J., Krishnakumar, V., et al. (2017). ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. Plant Cell *29*, 1806–1821. https://doi.org/10.1105/tpc.17.00073.
30. Weijers, D., Franke-van Dijk, M., Vencken, R.J., Quint, A., Hooykaas, P., and Offringa, R. (2001). An *Arabidopsis* minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene. Development *128*, 4289–4299. https://doi.org/10.1242/dev.128.21.4289.

31. Zurek, P.J., Knyphausen, P., Neufeld, K., Pushpanath, A., and Hollfelder, F. (2020). UMI-linked consensus sequencing enables phylogenetic analysis of directed evolution. Nat. Commun. *11*, 6023. https://doi.org/10.1038/s41467-020-19687-9.

32. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.

33. Slatkin, M., and Maddison, W.P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics *123*, 603–613. https://doi.org/10.1093/genetics/123.3.603.

34. Wu, Z., Waneka, G., Broz, A.K., King, C.R., and Sloan, D.B. (2020). *MSH1* is required for maintenance of the low mutation rates in plant mitochondrial and plastid genomes. Proc. Natl. Acad. Sci. USA. *117*, 16448–16455. https://doi.org/10.1073/pnas.2001998117.

35. Shao, M.R., Kumar Kenchanmane Raju, S., Laurie, J.D., Sanchez, R., and Mackenzie, S.A. (2017). Stress-responsive pathways and small RNA changes distinguish variable developmental phenotypes caused by MSH1 loss. BMC Plant Biol. *17*, 47. https://doi.org/10.1186/s12870-017-0996-4.

36. Broz, A.K., Keene, A., Fernandes Gyorfy, M., Hodous, M., Johnston, I.G., and Sloan, D.B. (2022). Sorting of mitochondrial and plastid heteroplasmy in *Arabidopsis* is extremely rapid and depends on MSH1 activity. Proc. Natl. Acad. Sci. USA. *119*, e2206973119. https://doi.org/10.1073/pnas.2206973119.

37. Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618. https://doi.org/10.1038/nrg2146.

38. Shaw, F.H., Geyer, C.J., and Shaw, R.G. (2002). A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. Evolution *56*, 453–463. https://doi.org/10.1111/j.0014-3820.2002.tb01358.x.

39. Walbot, V. (1985). On the life strategies of plants and animals. Trends Genet. *1*, 165–169. https://doi.org/10.1016/0168-9525(85)90071-X.

40. Besnard, F., Vernoux, T., and Hamant, O. (2011). Organogenesis from stem cells in planta: multiple feedback loops integrating molecular and mechanical signals. Cell. Mol. Life Sci. *68*, 2885–2906. https://doi.org/10.1007/s00018-011-0732-4.

41. Meyerowitz, E.M. (1997). Genetic control of cell division patterns in developing plants. Cell *88*, 299–308. https://doi.org/10.1016/s0092-8674(00)81868-1.

42. Thorén, M.H., Olsson, B., Vonk, P.J., Siljestam, M., Reimegård, J., Ryberg, M., and Johannesson, H. (2025). Early germline sequestration in a basidiomycete fungus. Science *389*, 720–723. https://doi.org/10.1126/science.adu8580.

43. Schmitt, S., Heuret, P., Troispoux, V., Beraud, M., Cazal, J., Chancerel, É., Cravero, C., Guichoux, E., Lepais, O., Loureiro, J., et al. (2024). Low-frequency somatic mutations are heritable in tropical trees *Dicorynia guianensis* and *Sextonia rubra*. Proc. Natl. Acad. Sci. USA. *121*, e2313312121. https://doi.org/10.1073/pnas.2313312121.

44. Satake, A., Imai, R., Fujino, T., Tomimoto, S., Ohta, K., Na'iem, M., Indrioko, S., Widiyatno, W., Purnomo, S., Morales, A.M., et al. (2024). Somatic mutation rates scale with time not growth rate in long-lived tropical trees. eLife *12*, RP88456. https://doi.org/10.7554/eLife.88456.

45. 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*, 2040–2049. https://doi.org/10.1093/treephys/tpac058.

46. Smith, S.A., and Donoghue, M.J. (2008). Rates of molecular evolution are linked to life history in flowering plants. Science *322*, 86–89. https://doi.org/10.1126/science.1163197.

47. Lanfear, R., Ho, S.Y., 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*, 1879. https://doi.org/10.1038/ncomms2836.

48. Fagerström, T., and Fagerstrom, T. (1992). The meristem-meristem cycle as a basis for defining fitness in clonal plants. Oikos *63*, 449–453. https://doi.org/10.2307/3544971.

49. Tuomi, J., and Vuorisalo, T. (1989). What are the units of selection in modular organisms? Oikos *54*, 227–233. https://doi.org/10.2307/3565271.

50. Hu, Z., Cools, T., and De Veylder, L. (2016). Mechanisms used by plants to cope with DNA damage. Annu. Rev. Plant Biol. *67*, 439–462. https://doi.org/10.1146/annurev-arplant-043015-111902.

51. Inzé, D., and De Veylder, L. (2006). Cell cycle regulation in plant development. Annu. Rev. Genet. *40*, 77–105. https://doi.org/10.1146/annurev.genet.40.110405.090431.

52. Tsukaya, H. (2019). Re-examination of the role of endoreduplication on cell-size control in leaves. J. Plant Res. *132*, 571–580. https://doi.org/10.1007/s10265-019-01125-7.

53. Sashittal, P., Schmidt, H., Chan, M., and Raphael, B.J. (2023). *Startle*: A star homoplasy approach for CRISPR-Cas9 lineage tracing. Cell Syst. *14*, 1113–1121.e9. https://doi.org/10.1016/j.cels.2023.11.005.

54. Satina, S., Blakeslee, A.F., and Avery, A.G. (1940). Demonstration of the three germ layers in the shoot apex of Datura by means of induced polyploidy in periclinal chimeras. Am. J. Bot. *27*, 895–905. https://doi.org/10.1002/j.1537-2197.1940.tb13952.x.

55. Furner, I.J., and Pumfrey, J.E. (1992). Cell fate in the shoot apical meristem of *Arabidopsis thaliana*. Development *115*, 755–764. https://doi.org/10.1242/dev.115.3.755.

56. Zhang, S., Chen, X., Jin, E., Wang, A., Chen, T., Zhang, X., Zhu, J., Dong, L., Sun, Y., Yu, C., et al. (2025). The GSA family in 2025: a broadened sharing platform for multi-omics and multimodal data. Genomics Proteomics Bioinformatics *23*, qzaf072. https://doi.org/10.1093/gpbjnl/qzaf072.

57. CNCB-NGDC Members and Partners (2025). Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2025. Nucleic Acids Res. *53*, D30–D44. https://doi.org/10.1093/nar/gkae978.

58. Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. *40*, D1202–D1210. https://doi.org/10.1093/nar/gkr1090.

59. Letunic, I., and Bork, P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. *52*, W78–W82. https://doi.org/10.1093/nar/gkae268.

60. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

61. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

62. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

63. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

64. Schliep, K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics *27*, 592–593. https://doi.org/10.1093/bioinformatics/btq706.

65. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Agrobacterium tumefaciens GV3101	Tsingke	Cat# TSC-A01
Escherichia coli DH5α	Tsingke	Cat# TSC-C14
Biological samples		
Arabidopsis thaliana: *msh1*	AraShare	https://www.arashare.cn/
Chemicals, peptides, and recombinant proteins		
Hygromycin B	Invitrogen	Cat#10687010
Kanamycin	Solarbio	Cat# K1030
Rifampicin	JSENB	Cat# JS9004
FM4-64	Thermo Fisher	Cat# T3166
Phanta Flash Super-Fidelity DNA Polymerase	Vazyme	Cat# P510-01
KpnI	NEB	Cat# R3142S
HindIII	NEB	Cat# R3104S
PstI	NEB	Cat# R3140S
XbaI	NEB	Cat# R0145S
Critical commercial assays		
FastPure Plasmid Mini Kit	Vazyme	Cat# DC201
Clean & Concentrator-5 Kit	Zymo	Cat# D4004
pEASY-Blunt Cloning Kit	Transgene	Cat# CB501
Deposited data		
Raw whole genome sequencing data for Wild-type, Plants 1, 2 and 3	This paper	GSA: CRA029430
Raw whole genome Oxford Nanopore sequencing data for Plants 2 and 3	This paper	GSA: CRA029430
Raw UMIC Oxford Nanopore sequencing data for Plants 1, 2 and 3	This paper	GSA: CRA029430
Raw UMIC Oxford Nanopore sequencing data of progeny for Plants 2 and 3	This paper	GSA: CRA029430
Sanger sequencing data for progeny readouts of Plant 1	This paper	GSA: CRA029430
Arabidopsis thaliana genome TAIR10	TAIR[58]	http://www.arabidopsis.org
Experimental models: Organisms/strains		
Arabidopsis thaliana: Columbia-0	N/A	N/A
Arabidopsis thaliana: SMALT system edited *Arabidopsis* Plant 1	This paper	N/A
Arabidopsis thaliana: SMALT system edited *Arabidopsis* Plant 2	This paper	N/A

(*Continued on next page*)

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Arabidopsis thaliana: SMALT system edited *Arabidopsis* Plant 3	This paper	N/A
Arabidopsis thaliana: Progeny of SMALT-system-edited *Arabidopsis* Plant 1	This paper	N/A
Arabidopsis thaliana: Progeny of SMALT-system-edited *Arabidopsis* Plant 2	This paper	N/A
Arabidopsis thaliana: Progeny of SMALT-system-edited *Arabidopsis* Plant 3	This paper	N/A
Oligonucleotides		
Oligonucleotides used in this study are listed in Table S1	This paper	N/A
Recombinant DNA		
pRPS5A-driven AI plasmid	This paper	GenBase: GB0003227
pRPS5A-driven Readout plasmid	This paper	GenBase: GB0003228
Software and algorithms		
UMIC-seq	Zurek et al.[31]	https://github.com/fhlab/UMIC-seq
FastTree v2.1.10	Price et al.[32]	https://github.com/morgannprice/fasttree
iTOL v7.2.1	Letunic and Bork[59]	https://github.com/TongZhou2017/itol.toolkit
Startle v1.0	Sashittal et al.[53]	https://github.com/raphael-group/startle
BWA v0.7.17	Li and Durbin[60]	https://github.com/lh3/bwa
minimap2 v2.17-r941	Li[61]	https://github.com/lh3/minimap2
GATK v4.0.5.1	McKenna et al.[62]	https://github.com/broadinstitute/gatk
Samtools v1.15	Li et al.[63]	https://github.com/samtools/samtools
phangorn v2.12.1	Schliep[64]	https://github.com/KlausVigo/phangorn
ImageJ	Schindelin et al.[65]	https://imagej.net
NIS-Elements AR software v5.42.01	Nikon	https://www.microscope.healthcare.nikon.com/en_AOM/products/software/nis-elements/nis-elements-advanced-research
Analysis pipeline	This paper	https://github.com/ZhangGeyu/Pipeline-of-lineage-tracing-in-Arabidopsis

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plants

A. thaliana lines utilized in this study were either in the wild-type Columbia (Col-0) genetic background or the *msh1* mutant (SALK_041951C, T-DNA insertion line) obtained from Arashare (https://www.arashare.cn/index/). The primers used for genotyping the T-DNA insertion in the *msh1* mutant were listed in Table S1.

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

METHOD DETAILS

Plasmid construction

To perform lineage tracing in *Arabidopsis*, the AI plasmid was constructed by inserting the *pRPS5A:AI* fragment into the plasmid XF4363, which harbors kanamycin-resistance as a selection marker. The readout plasmid was created through the insertion of the *pRPS5A:eGFP* (with 1.2-kb readout serving as its 3′-untranslated region) fragment into the plasmid XF675, which harbors hygromycin-resistance as a selection marker. The *pRPS5A* fragment was amplified from *A. thaliana* genomic DNA and the *AI* fragment was obtained from a previous study.[26] The *pRPS5A* and *AI* fragments were subsequently assembled into the PstI-digested plasmid XF4363 via Gibson assembly, while the *pRPS5A* and *eGFP* fragments were assembled into the KpnI-digested plasmid XF675.

Plant transformation and sampling

We performed *Agrobacterium*-mediated transformation of *A. thaliana* in this study using the floral dip method. To prevent premature editing of readout sequence within *Agrobacterium tumefaciens*, the AI plasmid and the readout plasmid were separately transformed into *A. tumefaciens* strain GV3101. Individual *A. tumefaciens* cultures were prepared and resuspended in a 5% sucrose solution to achieve an OD600 of 1.8∼2.0. Two cultures were then mixed equally, and Silwet-77 was added just before the floral dipping transformation. Transgenic seeds (T_0) were harvested.

Dry T_0 seeds were sterilized with 75% ethanol for one minute, followed by a 15-minute treatment with 10% bleach and subsequent rinsing with sterile water until clean. These seeds were selected on 1 × Murashige and Skoog (MS) medium supplemented with 1% sucrose (w/v), 50 μg ml^{-1} kanamycin, and 25 μg ml^{-1} hygromycin. Antibiotic-resistant T_1 seedlings were selected and transplanted into the soil for further growth. The plants were grown in a greenhouse at 22 °C under a 16-hour light/8-hour dark photoperiod. After the maturation of siliques on individual branches, rosette and cauline leaves were collected, and genomic DNA was extracted from individual leaves for UMIC-seq and DNA-seq library construction.

UMIC-seq and Oxford Nanopore sequencing

To append both the UMI and the sample-specific barcode to each readout fragment, a two-step PCR approach was employed to construct a UMIC-seq library[31] (Figure S1E). To uniquely label each readout molecule, genomic DNA from each sample was used as a template for a brief five-cycle amplification, utilizing readout-specific forward and reverse primers. The reverse primer includes a 40-nucleotide UMI sequence, which is composed of two 15-nucleotide random sequences followed by an anchor sequence "GTATC" (for improving accuracy of UMI mapping). The PCR products were purified using the Zymo Clean & Concentrator-5 Kit following the manufacturer's guidelines. Then, the sample-specific barcodes were subsequently introduced into the purified products, through thirty cycles of PCR amplification, using a forward primer containing a 10-nucleotide sample-specific barcode. The primers and sample barcode sequences used in this study can be found in Table S1.

The tagged variants and the linearized pUC19 plasmid underwent ligation through Gibson assembly. The ligation product was then transformed into *E. coli* to reduce molecule diversity, with approximately 4000 colonies collected for each sample. Next, plasmids were isolated and subjected to digestion with restriction enzyme (XbaI and KpnI for Plant 1, and HindIII and KpnI for Plants 2 and 3). Finally, the purified tagged variants from each sample were pooled together.

To construct Oxford Nanopore sequencing libraries, the tagged DNA pool then underwent end repair and dA-tailing, followed by ligation to a unique dT-tailed library barcode adapter. The resulting high-quality library was loaded onto chips for sequencing using the Oxford Nanopore PromethION platform.

Acquisition of progeny readout

Self-pollinated T_2 seeds were gathered and screened on 1 × MS medium supplemented with 1% sucrose (w/v) and 25 ug ml^{-1} hygromycin to confirm inherited readout sequences. Hygromycin-positive seedlings (whole plants) were harvested at seven days old for DNA extraction.

For Plant 1, the readout sequences were PCR-amplified using the primers barF2 and barR5 (Table S1), producing 1.3-kb PCR products, which were then gel-purified (Vazyme, DC301) and ligated into the pEASY-blunt cloning plasmid (Transgene, CB501). Following transformation into *E. coli* strain DH5α, an average of 8.9 clones were collected for Sanger sequencing, using the general primers M13F and M13R, resulting in complete readout sequences for each clone. For Plants 2 and 3, the readout sequence of each T_2 seedling was amplified and constructed for UMIC-seq libraries, following the same experimental procedure used for their parental plants.

Genome sequencing of the SMALT T_1 plants

Genomic DNA extracted from a rosette leaf of either a wild-type *A. thaliana* or from each of the three SMALT T_1 plants (Plant 1–3) was used to construct individual DNA-seq libraries. Sequencing libraries were generated using the NEB Next Ultra DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's protocol. We performed sonication to shear genomic DNA to an approximately 350-bp fragment size. Subsequently, the DNA fragments were end-polished, A-tailed, and ligated to full-length adapters for Illumina sequencing, followed by PCR amplification. These four DNA-seq libraries were sequenced on the Illumina NovaSeq 6000 PE150 platform, with ten gigabases obtained for each sample (Table S2). This represents an approximately 50-fold coverage on the *A. thaliana* genome.

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

We also constructed the Oxford Nanopore sequencing libraries using the genomic DNA from a rosette leaf of Plant 2 or Plant 3. The tagged DNA pool then underwent end repair and dA-tailing, followed by ligation to a unique dT-tailed library barcode adapter. The resulting high-quality library was loaded onto chips for sequencing using the Oxford Nanopore PromethION platform, yielding ten gigabases of long-read data (Table S2).

Live imaging

The expression of readout and *AI* are driven by *pRPS5A* in our SMALT system, a promoter known to be activated in proliferating cells.[30] To characterize the expression of the SMALT system, we used live imaging to detect GFP fluorescence in root tips, flowers, shoot apices, and inflorescence apices. These tissues were dissected from fresh plants. The GFP fluorescence was captured using a fluorescent stereomicroscope (Leica, M165FC) with a MshOt MC25-M camera for all these tissues. Images were processed and analyzed using ImageJ.[65]

We then used confocal microscopy to detect the expression of the SMALT system in detail in the meristem. Considering that the shoot apex is relatively soft and difficult to dissect, we observed only in the inflorescence meristem. We dissected inflorescence apex with a short stem remaining, placed it into dissecting medium (3% [w/v] agarose), and then carefully removed the remaining floral primordia (older than needed) using a fine needle tip under a stereomicroscope (Nikon, SMZ18). After dissection, FM4-64 (Thermo Fisher, 10 mg/mL) was applied to the apex for 10 min. The inflorescence apex was then mounted in imaging medium (half-strength MS medium topped with 1% [w/v] agarose) and submerged in water for imaging. These images were captured using a Nikon A1R MP confocal laser scanning microscope equipped with 60 × water dipping lenses. GFP was excited with the 488 nm laser and collected using a 500–550 nm band-pass filter. To detect the signal of FM4-64 staining, a 561 nm laser was used for excitation and a 575–616 nm band-pass filter was used for detection. Image acquisition and microscope control were performed using NIS-Elements AR software (version 5.42.01).

Albino phenotypes in the *msh1* mutants

We cultivated a total of 461 *msh1 A. thaliana* plants. Among them we observed 11 branches from 10 plants containing both albino and green siliques, clearly divided into sectors and containing seeds from at least one albino silique. These 11 branches contained a total of 74 green siliques and 33 albino siliques. We grew all seeds from these 107 siliques in MS medium agar plates. Among the 33 albino siliques, 15 siliques grew no seed, 7 siliques with no seeds germinated, and the remaining 11 siliques had an average germination rate of 57% (116 germinated seeds / 203 total seeds). We recorded the cotyledon color phenotypes of seedlings from all germinated seeds from these 11 branches after 14-days growth (Table S3). It is worth noting that due to its reliance on stochastic chloroplast genome loss in individual cells, this lineage-tracing method has low labeling efficiency, necessitating screening of a large mutant population.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis of the SMALT T_1 plant genomes

Genomic DNA sequencing reads were aligned to the *A. thaliana* reference genome (TAIR10, https://www.arabidopsis.org/)[58] using BWA (version 0.7.17)[60] with default parameters. Base substitution identification was performed with GATK (version 4.0.5.1)[62] HaplotypeCaller. A mutation in the *A. thaliana* genome (out of the readout sequence) was classified as somatic if it met the following criteria: (1) coverage at the site was between 30–100×; (2) at least three reads supported the mutation; (3) it was not shared by the wild-type and the three independent SMALT T_1 plants (to filter out pre-existing variation from the founder line); (4) QD (QUAL by Depth) > 2, QUAL (Base Quality) > 30, SOR (Strand Odds Ratio) < 3, and FS (Fisher Strand) < 60. The molecular spectrum of these somatic mutations in the T_1 plants was similar to the wild-type background but distinct from the somatic mutations identified in the readouts (Figures S6A and 1E).

To determine the insertion site and copy number of the AI and readout plasmids, we searched for reads that simultaneously covered both the chromosome and the T-DNA borders. For Plant 1, reads spanning the T-DNA left border were detected at positions 24,610,921 (carrying the T-DNA for *AI*) and 24,742,560 (carrying T-DNA for the readout) on chromosome 1. Additionally, reads covering a segment of the right border of AI and a reverse right border of the readout were also identified. Furthermore, Samtools (version 1.15, https://www.htslib.org/)[63] depth analysis revealed a ∼50% reduction in read coverage between positions 24,610,921 and 24,742,560. And the coverage within this range was similar to that of *AI* and readout. These observations together indicated that *AI* and readout each inserted one copy, head-to-head into the same location, leading to a heterozygous deletion between positions 24,610,921 and 24,742,560 on chromosome 1 (Figures S6B and S6C). The consensus readout sequence from Plant 1 was identical to the reference readout sequence, suggesting no mutation occurred in *Agrobacterium* before transformation to *A. thaliana*.

For Plant 2, the read coverage of *eGFP* fragment is approximately 2.1-fold of the genomic average, suggesting that this plant contains multiple readout copies integration. To further identify these copies, the Oxford Nanopore sequencing reads were aligned to the *A. thaliana* reference genome (TAIR10) using minimap2 (version 2.29-r1283)[61] with parameters: -ax map-ont. Mutation identification was performed with Samtools mpileup using parameters: –max-depth 0 –output-BP. We identified three chromosomal insertion sites (two of which were on the same chromosome) and six distinct readout copies by searching for reads that spanned both genomic and plasmid sequences (Table S4). Note that all six readout copies can be amplified from the genomic DNA during UMIC-seq library

preparation and sequencing, so mutations must be analyzed copy-specifically to avoid conflating sequence differences among distinct copies with true mutations within a single copy. Among the six copies, one readout integration containing a signature 20-nucleotide deletion, likely occurring within *Agrobacterium*, making it easily identified from the UMIC-seq results. Other copies were sometimes not distinguishable between each other due to lack of signature mutations. This readout copy was used for the construction of the cell lineage tree in this study (Plant 2-Copy 1, 7 mutations).

Plant 3 was also sequenced with Oxford Nanopore Technology. The alignment of sequencing reads, mutation identification, and copy identification were performed following the same protocol as for Plant 2. We identified four chromosomal insertion sites and four distinct readout copies (Table S4). Among them, we selected the copy with the largest number of signature mutations, for subsequent analysis (Plant 3-Copy 1, 11 mutations).

Identification of consensus sequences and *bona fide* mutations from UMIC-seq data

We extracted sample barcodes and UMIs, and clustered reads, using the UMI-linked consensus sequencing tool[31] (Figures S7A–S7D; Table S5). Specifically, sample barcodes were extracted from reads using the UMIextract command in UMIC-seq.py under default parameters, upstream of the readout sequence ("sample-barcode-probe"), "TGCCCGACAACCACTACCTGAGCACC CAGTCCGCCCTGAGCAAAGACCCC" for Plant 1 and "GACTCACTATAGGGGATCCCGCCAAAGGTGCACGCGGATTCGTCGG CAGC" for Plants 2 and 3. Individual reads were demultiplexed into separate files based on the extracted sample barcodes, using the demultiplex command in UMIC-seq-helper.py, allowing up to two mismatches in the sample barcode. Subsequently, the UMIextract command was applied to extract UMIs from sequencing reads in each file, downstream of the readout sequence ("UMI-probe") "CGCGATATAAGCGCGGCGTGCCAAGCGCGGCCGCTTGCTAGCCGTATTAC" for Plant 1 and "CAAGGACGATGATGACAAGGG CGACTATAAGGATGACGACGATAAGCTGC" for Plants 2 and 3, allowing 50% mismatch under the default parameters. The clusterfull command in UMIC-seq.py was then applied to cluster UMIs, with –aln_thresh set to 50 and –size_thresh set to 10, and –stop_thresh set to 0 (Figure S7B). Samtools mpileup was then utilized to identify mutations in comparison to the original readout sequences, with parameters –max-depth 0 –output-BP (Figure S7C). Mutations present in over 60% of readouts within the same UMI cluster were considered as *bona fide* mutations (Figure S7D). We retained only readouts containing at least one mutation, as they provided cell lineage information.

Mutation identification from progeny

Mutation identification for each readout sequence was performed using Samtools mpileup with parameters –max-depth 0 –output-BP. Presumably, a seed could inherit a single readout from either the sperm or egg cell. Alternatively, a seed could inherit two readouts from both (leading to heterozygous readout in a progeny sample if additional mutations were accumulated in either sperm or egg lineage). For simplicity, and assuming a recent divergence of ancestral cells that produce spermatocytes and oocytes, we focused on recovering just one readout copy for seedling genotyping for Plant 1. Considering the ongoing editing in the readout in T_2 plants, we identified mutations with a frequency $\geq$50% among all readouts from a progeny sample as the original mutations present in the zygote (Table S6). We also identified post-germination mutations that occurred in T_2 plants, defined as those with a frequency <50% and appeared in at least two readouts.

For Plants 2 and 3, the rich genotype information from UMIC-seq enabled us to determine whether the progeny samples were heterozygous or homozygous. To this end, we plotted the distributions of mutation frequencies in the progeny samples of Plants 2 and 3. The density plots showed a single peak at ~90% for Plant 2, whereas two distinct peaks at ~50% and ~90% were observed for Plant 3 (Figure S3D). These observations suggested that all progenies of Plant 2 were homozygous, while some progenies from Plant 3 were heterozygous. To identify the two haplotypes in the heterozygous progenies from Plant 3, for each progeny sample we identified mutations with frequencies $\geq$ 30% and extracted readouts containing these mutations. The two most frequently occurring mutation haplotypes were identified for each sample (two examples shown in Figure S3E). If a progeny contained two haplotypes both present in a $\geq$ 30% read frequencies, the progeny was considered as heterozygous; otherwise, it is considered as homozygous. Among all 38 progenies of Plant 3, 30 were classified as homozygous and 8 as heterozygous (Table S6).

Identification of recurrent (hotspot) mutations

We identified and excluded recurrent mutations ("hotspots") that likely arose independently multiple times across the sampled leaves of a T_1 plant. Such mutations would confound phylogenetic analysis by suggesting false phylogenetic relationships. To identify hotspot mutations, we first established the empirical relationship between a mutation's overall frequency in the plant (calculated from all sequenced samples combined) and the number of individual leaf samples in which it was detected (box plots shown in Figure S7E). Mutations that were statistical outliers above the expected trend in the box plot were classified as putative hotspots. As a secondary filter, we required that a putative hotspot be detectable in all major branches of the plant, including the collective rosette leaves, which were treated as a single branch for this analysis. This conservative criterion helped ensure the mutation was widespread and likely recurrent, rather than a clonal event in a single lineage. A total of 10 (53C>−2GA, 77G>A, 445C>T, 513C>G, 795G>T, 965G>A, 1092T>−1C, 1114A>G, 1226C>T, and 1335C>T), 0, and 9 (77C>T, 306T>C, 850C>T, 941C>T, 1007C>T, 1013C>T, 1100G>A, 1186A>G, and 1242G>A) hotspot mutations were identified and subsequently excluded from Plants 1, 2, and 3, respectively. All downstream analyses, including the construction of cell lineage trees, were performed after removing these hotspots.

Please cite this article in press as: Guo et al., Testing Weismann's germ plasm theory in *Arabidopsis*, Current Biology (2026), https://doi.org/10.1016/j.cub.2026.03.003

Construction of cell lineage trees

Identical mutations shared among different samples can be used to construct the cell lineage tree. Maximum-likelihood trees were constructed using the software FastTree (version 2.1.10),[32] under default parameters. The distribution of mutation number per readout showed a long tail for three plants (i.e., Plants 1, 2 and 3) (Figure S7F). To reduce the impact of long-branch attraction in cell lineage tree construction, we used only the readouts with number of mutations less than or equal to mean + standard deviation for each plant for tree construction. *Startle* (version 1.0)[53] was also used to construct the cell lineage tree under the assumption of irreversibility. We used only the upstream 600 base substitution positions in the readout for this analysis, due to the limited number of mutations that can be included in the *Startle* software. The cell lineage tree was visualized using iTOL (version 7.2.1, https://itol.embl.de/)[59] and re-rooted based on the reference readout sequence.

Slatkin–Maddison test

We performed Slatkin–Maddison test to determine if somatic readouts from a plant branch tend to cluster in the cell lineage tree. To this end, we retrieved all somatic readouts from the three branches and treated them as three categories in the analysis. The number of state transitions among the three categories along the cell lineage tree was calculated using the "parsimony" function in R package "phangorn" (version 2.12.1).[64] To assess statistical significance, the tip labels (belonging to which category) of the tree were randomly permuted 1000 times and the number of state transitions was recalculated for each permutation. By comparing the observed transition number with the distribution from the random permutations, *p-value*s were estimated. To test whether germline readouts tend to cluster in the cell lineage tree, we designated the germline readouts of interest as one category and treated the remaining readouts as the other category in the analysis.

Resampling test for mutations shared by multiple branches

To determine if progenies from different plant branches shared more mutations than the random expectation, we estimated the number of expected shared mutations, using somatic readouts as pseudo germline readouts. For each branch, we counted number of germline readouts from this branch and randomly sampled an equal number of somatic readouts from all somatic samples from this branch. To control the differences in mutation load among readouts, we sampled somatic readouts to match the mutation number distribution of all germline readouts of a branch. To approximate the heritable germline signal, we restricted our analysis to somatic mutations that were detectable in the progeny of a branch. This sampling was performed 1000 times, and in each replicate, the number of shared mutations between a pair of branches was recorded. The observed number of shared mutations in progeny between a pair of branches was then evaluated against the distribution obtained from 1000 times somatic resampling, based on which a *p-value* was estimated. The expected number of mutations shared by all three branches was similarly estimated.

Current Biology, Volume *36*

Supplemental Information

Testing Weismann's germ plasm theory in *Arabidopsis*

Haotian Guo, Geyu Zhang, Lu Gao, Yang Liu, Yu Wang, Zhan Liu, Chunmei Guan, Chun Zhang, Li Liu, Qing Huan, and Wenfeng Qian

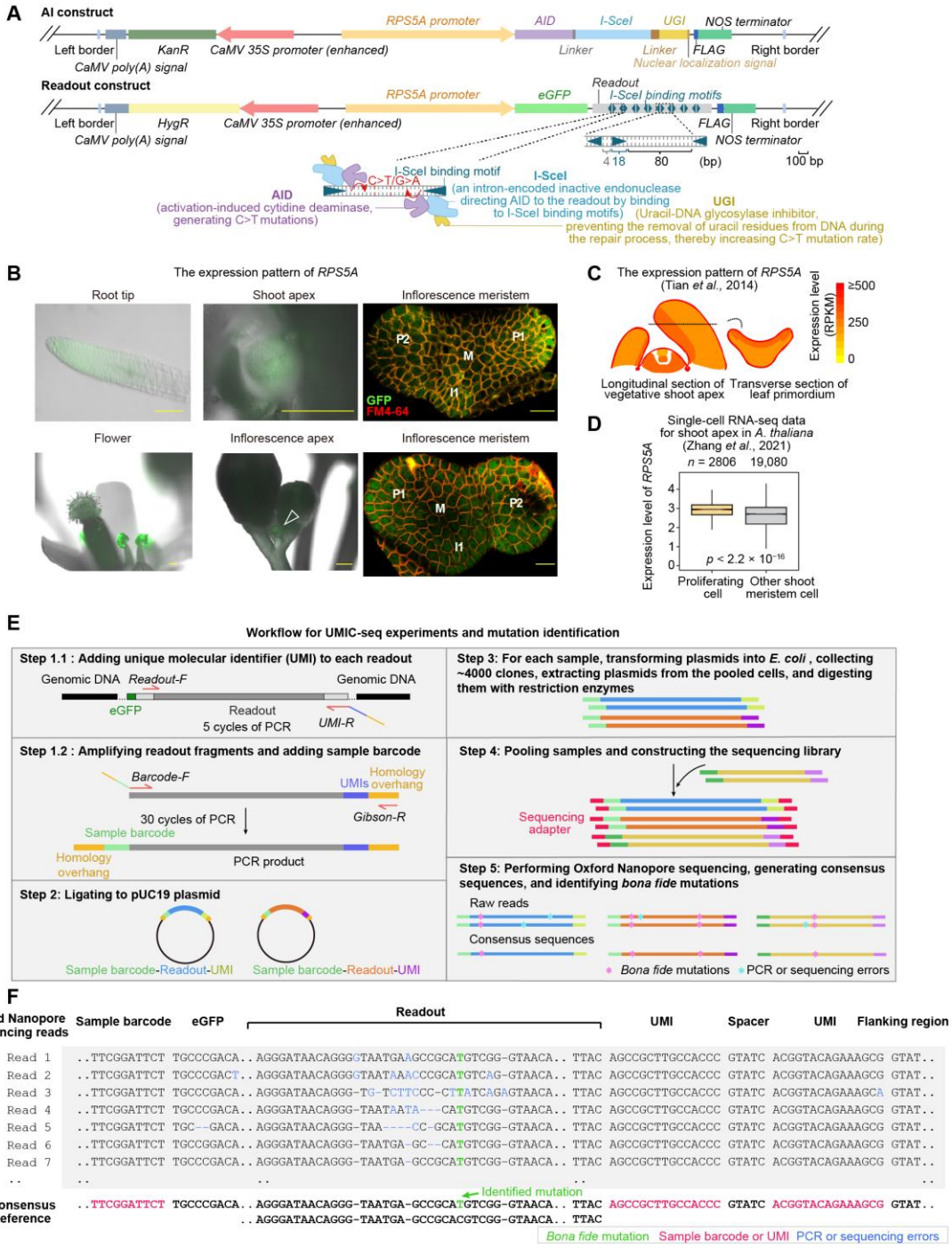


Figure S1. The SMALT system in *A. thaliana* and detection of its induced mutations. Related to Figure 1 and STAR Methods.

(A) Schematic representation of the AI and the Readout constructs for the SMALT system applied to *A. thaliana*. The AI fusion protein contains AID, I-SceI, and UGI, from N-terminus to C-terminus. The readout involves 16 I-SceI binding motifs, with each back-to-back motif pairs positioned at intervals of 4 base pairs and 80 base pairs apart from head-to-head motif pairs, facilitating the recording of mutations.

(B) Microscopic images of *pRPS5A*-driven *eGFP* expression in SMALT T_1 plants. Green fluorescence

in root tips, flowers, shoot apices, and inflorescence apices was captured using a Leica fluorescence stereomicroscope (scale bar: 100 μm). Higher-resolution images of expression patterns in the inflorescence meristem were acquired with a Nikon confocal laser scanning microscope (scale bar: 10 μm). M, inflorescence meristem; I, incipient floral primordia; P, floral primordia.

(C) Visualization of the *RPS5A* expression profile in the *A. thaliana* shoot apex. This image was generated based on RNA sequencing data from a previous study (Tian *et al*., 2014), using the Shoot Apex eFP browser (https://bar.utoronto.ca/eplant/). Different regions in the shoot apex are colored according to the mean expression level in reads per kilobase per million (RPKM), derived from three biological replicates.

(D) Box plot depicting the expression level of *RPS5A* in proliferating cells versus other shoot meristem cells of *A. thaliana*, based on single-cell RNA-seq data from 21,886 shoot apex cells (replicate 1) reported in a previous study (Zhang *et al.*, 2021). These data were subjected to principal component analysis using "RunPCA" (Seurat *R* package, https://satijalab.org/seurat/) on the 3,000 most variable genes among cells. Cells were clustered with the Louvain algorithm based on the top 20 principal components using "FindNeighbors" (dims = 1:20, k.param = 20) and "FindClusters" (resolution = 0.75), resulting in the identification of 21 clusters. Among them, three clusters showed >3-fold higher than average expression among all cells for each of the three marker genes used to identify proliferating cells (*3xHMG-box2*, *CDKB2;1*, and *MAD2*). These same clusters also showed >3-fold higher proportions of cells expressing relative to all cells, so they were classified as proliferating cells in this figure. The remaining 18 clusters were classified as other shoot meristem cells. *RPS5A* expression was compared between these two classes of cells. Expression levels are presented as log-normalized transcript counts per 10,000 transcripts per cell, with outliers excluded. *N* indicates the number of cells used for calculating the expression level. *P*-value was calculated using the two-tailed *t*-test.

(E) Workflow for UMIC-seq experiments and mutation identification. Step 1: UMIC-seq libraries were constructed using a two-step PCR, enabling pooled readouts to be labeled with both a UMI and sample-specific barcode. Step 2: Tagged variants and linearized pUC19 plasmids underwent ligation through Gibson assembly. Step 3: Ligation products were subsequently transformed into *E. coli* to restrict molecule diversity, and plasmids were then subjected to digestion. Step 4: Purified tagged variants from each sample were pooled, and their DNA ends were ligated to a library adapter to create the Oxford Nanopore library for sequencing. Step 5: Oxford Nanopore sequencing was performed, consensus sequence was generated for each UMI group (*i.e.*, reads sharing the same sample barcode and UMI), and those with a mutated read proportion of at least 60% were identified as *bona fide* mutations The primers utilized in these steps are detailed in Table S1.

(F) Example illustration of deriving consensus sequences from Oxford Nanopore sequencing reads.

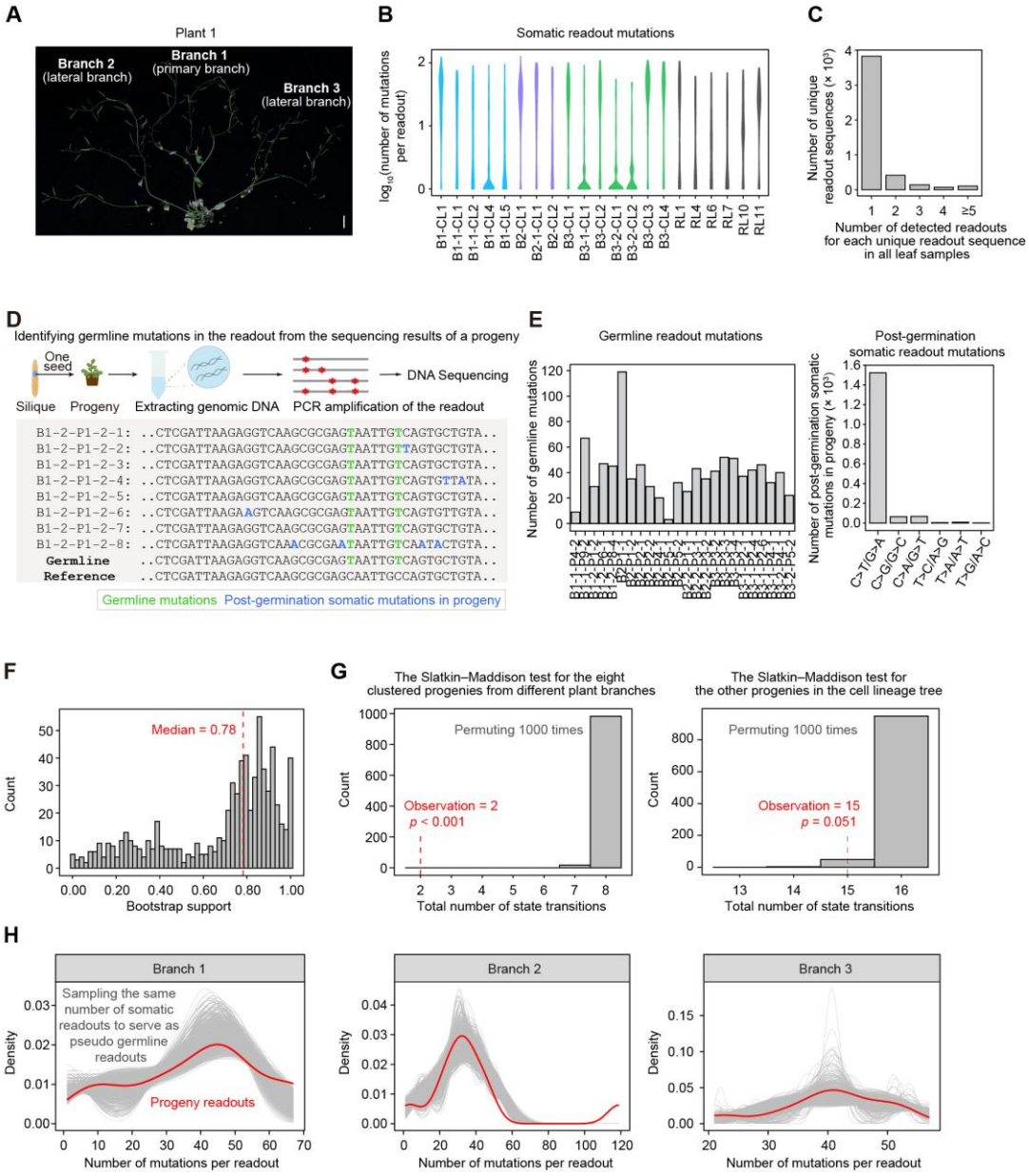


Figure S2. Mutation profiles in the readout across samples from Plant 1. Related to Figures 1–3.

(**A**) Photograph of Plant 1 with branch ID labeled. Bar represents 1 cm.

(B) Violin plot presenting the distribution of mutation numbers per readout in leaf samples.

(C) Histogram displaying the number of detected readouts for each unique readout sequence in all leaf samples.

(D) A schematic depicting the workflow for identifying parental germline mutations from the readouts derived from progeny sequencing. Taking one progeny sample (B1-2-P1-2) as an example, mutations detected in at least 50% of sequencing reads were regarded as parental germline mutations. The number following the last hyphen indicates the readout ID.

(E) Bar plot showing the number of parental germline mutations identified from each progeny (*i.e.*, seedling) and histogram showing the molecular spectrum of post-germination somatic mutations in T_2 seedlings.

(F) Distribution of bootstrap support values in the cell lineage tree of Plant 1 (presented in Figure 2A).

(G) Distribution of transition numbers from 1000 permutations of the Slatkin–Maddison test for the eight clustered progenies from different plant branches (left panel, $p < 0.001$) and for the other progenies in the cell lineage tree (right panel, $p = 0.051$). The observed transition number was indicated by the red dashed line.

(H) The distributions of mutation numbers per germline readout (red lines) and per pseudo germline readout (gray lines), used in the permutation test in Figure 3B.

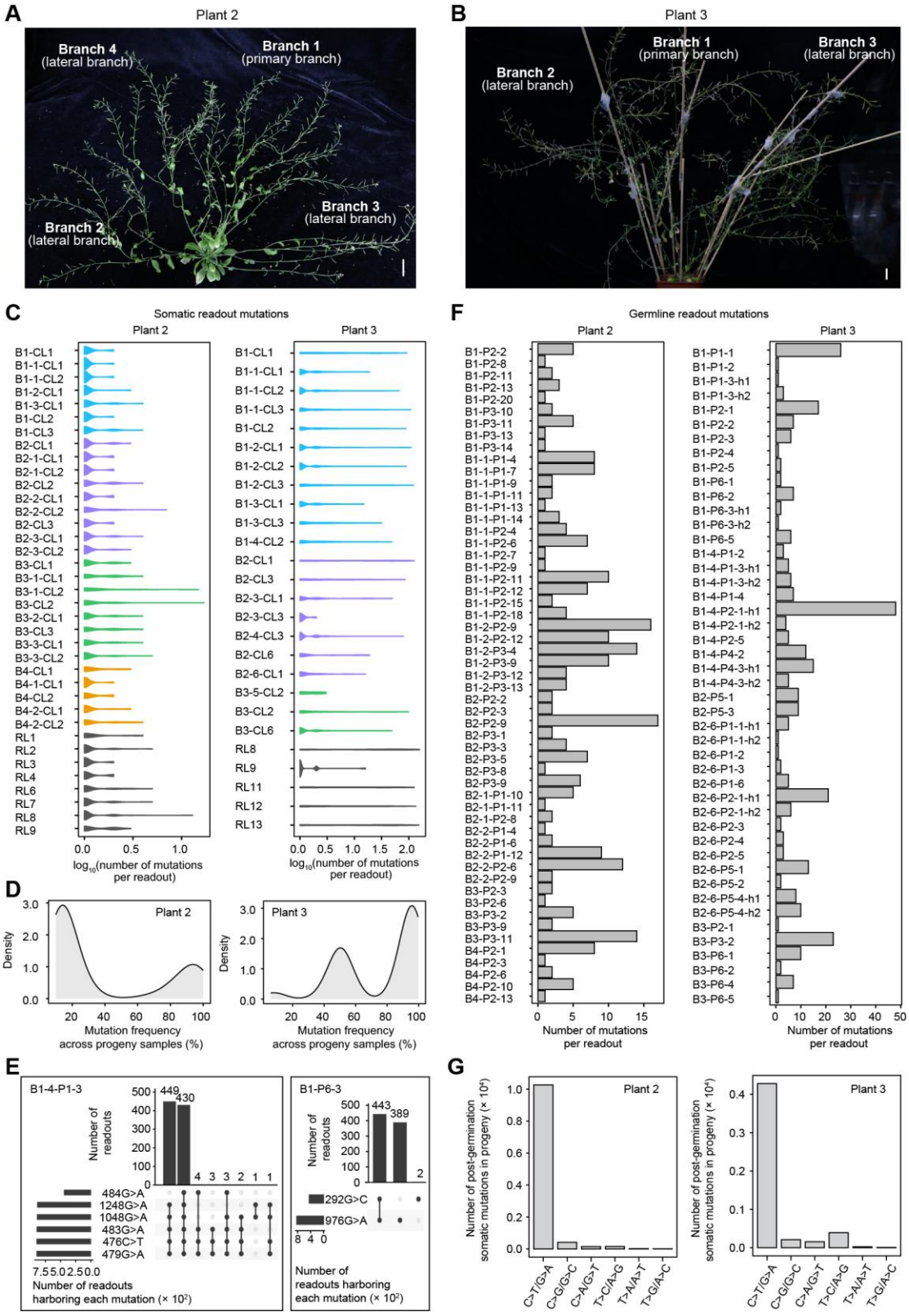


Figure S3. Mutation profiles in the readout across samples from Plants 2 and 3. Related to Figure 4.

(A and B) Photographs of the Plants 2 and 3 with branch ID labeled. Bar represents 1 cm.

(C) Violin plots representing the distribution of mutation numbers per readout in leaf samples in Plants 2 and 3.

(D) Density curves showing the distribution of mutation frequencies across progeny samples for Plants 2 and 3. The *x*-axis represents the frequency of each mutation within a sample (progeny), while the *y*-axis represents the probability density among all progeny samples of a T_1 plant, estimated using kernel density smoothing.

(E) UpSet visualization of the two major haplotypes identified within two representative progeny samples derived from Plant 3. The two major haplotype sequences for all eight heterozygous progenies are provided in Table S6.

(F) Bar plots showing the number of parental germline mutations identified from each progeny of Plants 2 and 3. Readouts from eight heterozygous progenies from Plant 3 were separated into two haplotypes, which were denoted by the suffixes "-h1" and "-h2".

(G) Histograms showing the molecular spectrum of post-germination somatic mutations in T_2 seedlings.

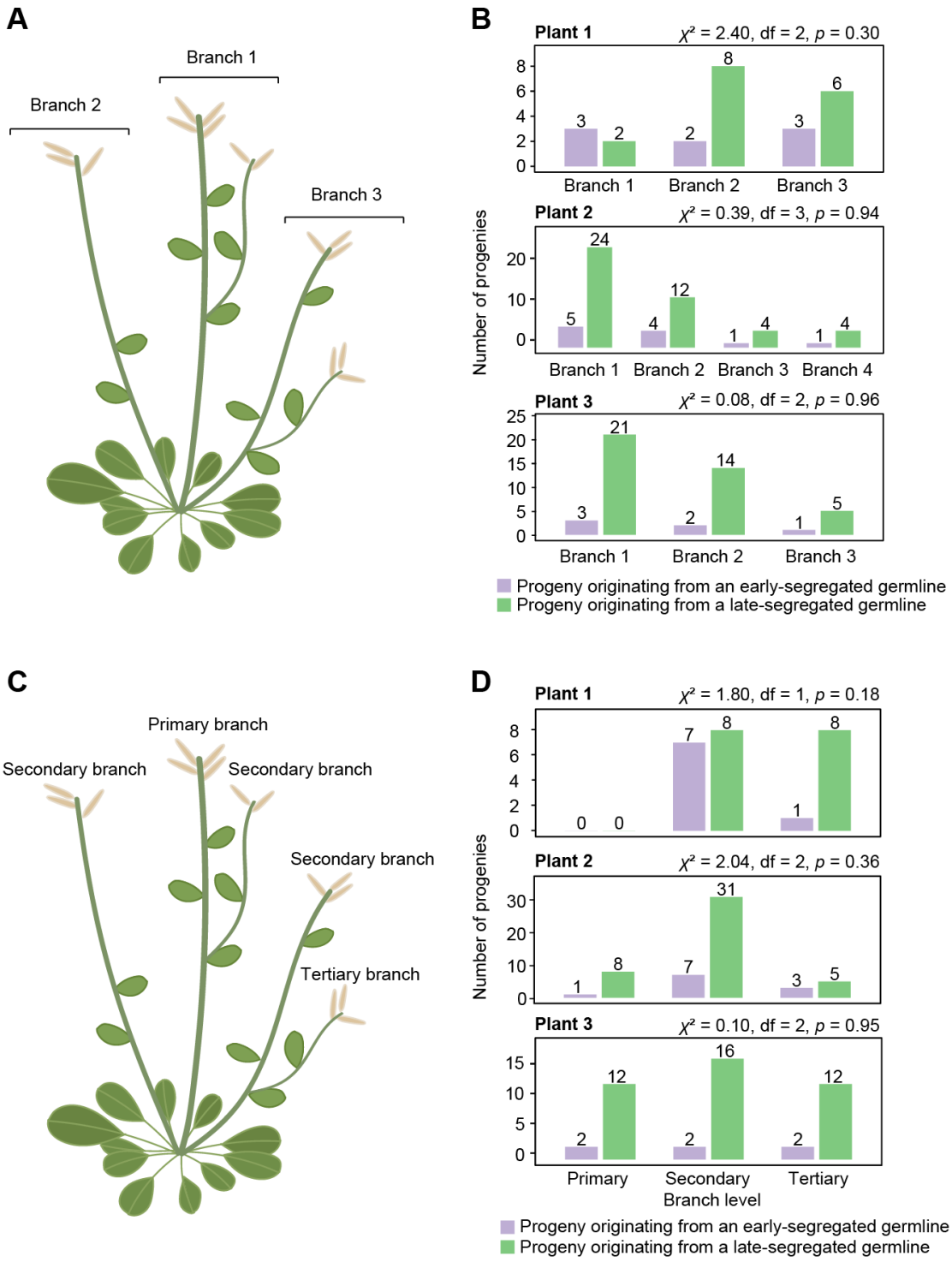


Figure S4. Distribution of progenies of different germline origins (early vs. late) in plant architecture. Related to Figures 2 and 4.

(A) Schematic diagram of an *A. thaliana* plant showing three plant branches as in Figure 1C.

(B) Comparison of progeny numbers derived from early- and late-segregated germlines across different plant branches. A Chi-square test for independence was performed to test the association between branch developmental order and the number of progenies derived from early- versus late-segregated

germlines.

(C) Similar to (A), but labeling different branch levels. The primary branch emerges directly from the shoot apical meristem. Secondary branches originate from the axils of rosette leaves (basal) or the axils of cauline leaves on the primary branch. Tertiary branches develop from the axils of cauline leaves on secondary branches.

(D) similar to (B), but to compare number of progenies derived from early- vs. late-segregated germlines across different branch levels, using a Chi-square test for independence.

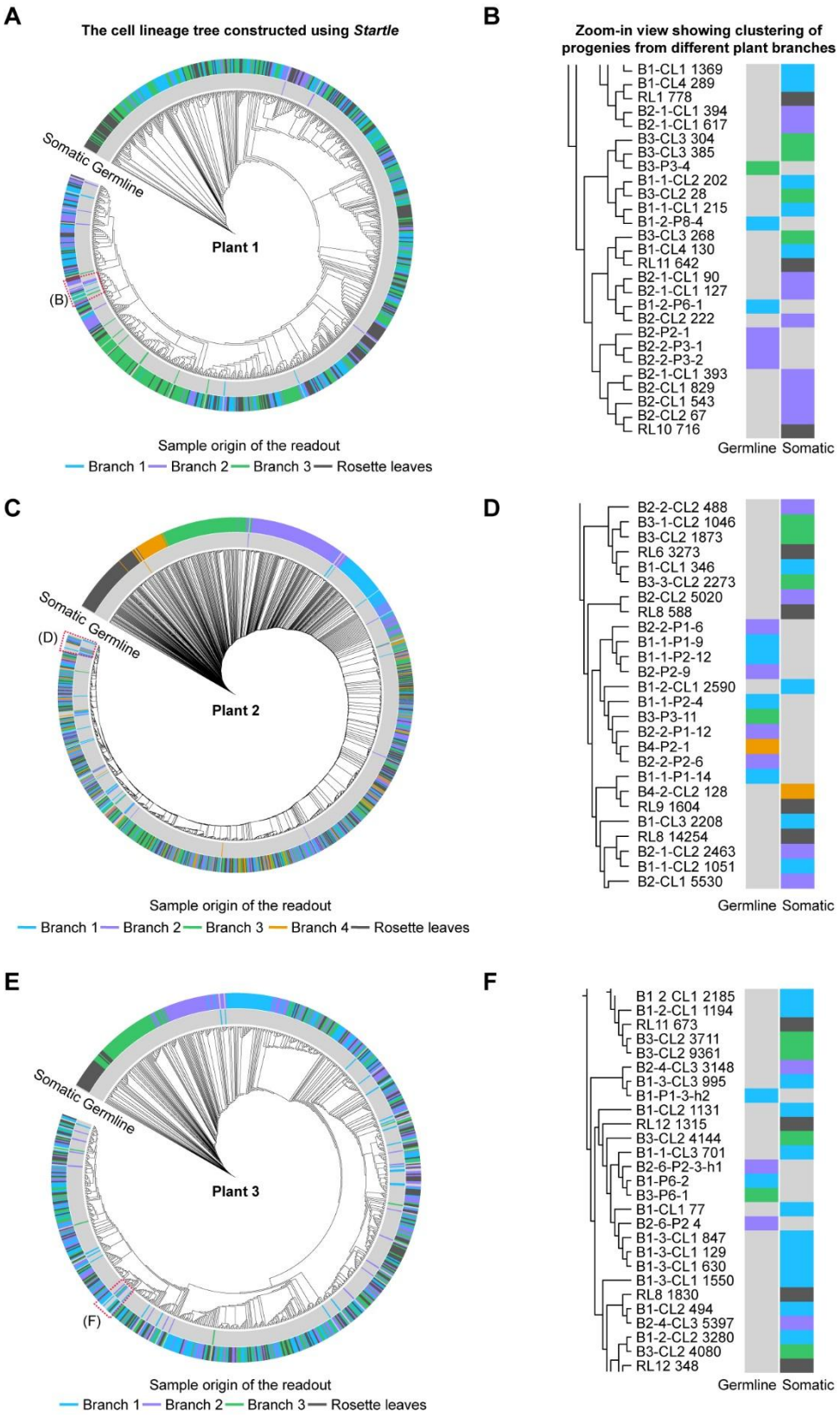


Figure S5. Cell lineage trees constructed using *Startle*. Related to Figures 2 and 4.

(A) Cell lineage tree of Plant 1 constructed using *Startle* under the assumption of irreversible mutation rates. The readout sequences used for tree construction were the same as those used in Figure 2A.

(B) Zoomed-in view of the cell lineage tree from the region boxed in red in panel (A).

(C–F) Similar to (A and B), for Plants 2 and 3. The readout sequences used for tree construction were the same as those used in Figures 4A and 4D.

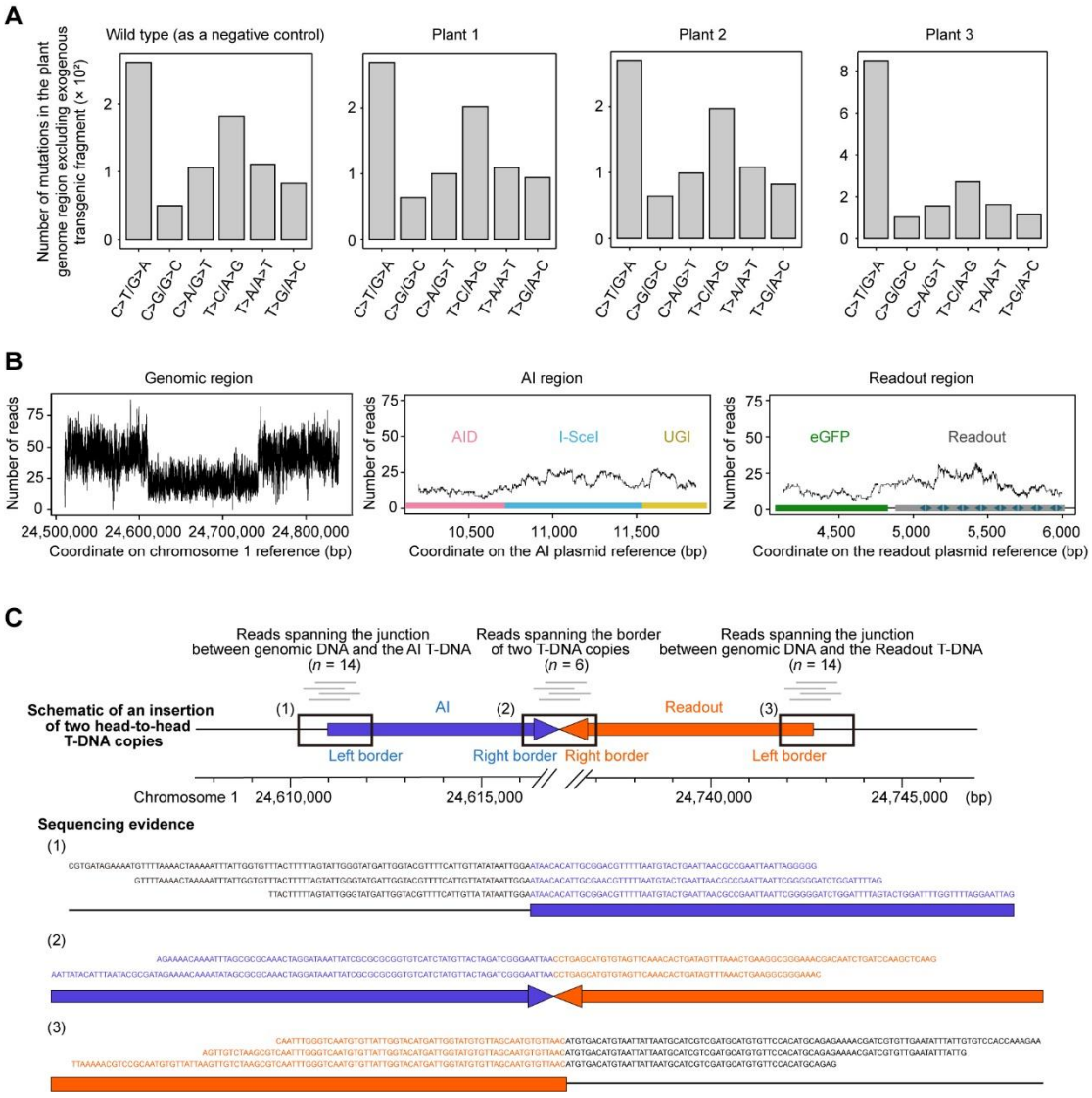


Figure S6. Analysis of T-DNA insertions and genomic somatic mutations in SMALT T_1 plants. Related to STAR Methods.

(A) Bar plots showing the number of mutations by type detected in the genomic regions (i.e., out of the readout), for the wild-type and the three SMALT T_1 plants.

(B) Graphs displaying the read coverage across the genomic region (left), the AI region (middle), and the readout region (right) for Plant 1.

(C) Diagram showing heterozygous T-DNA integration of the AI and the readout constructs into the *Arabidopsis* genome. Read counts spanning the three junction regions are displayed above, with example read sequences shown below.

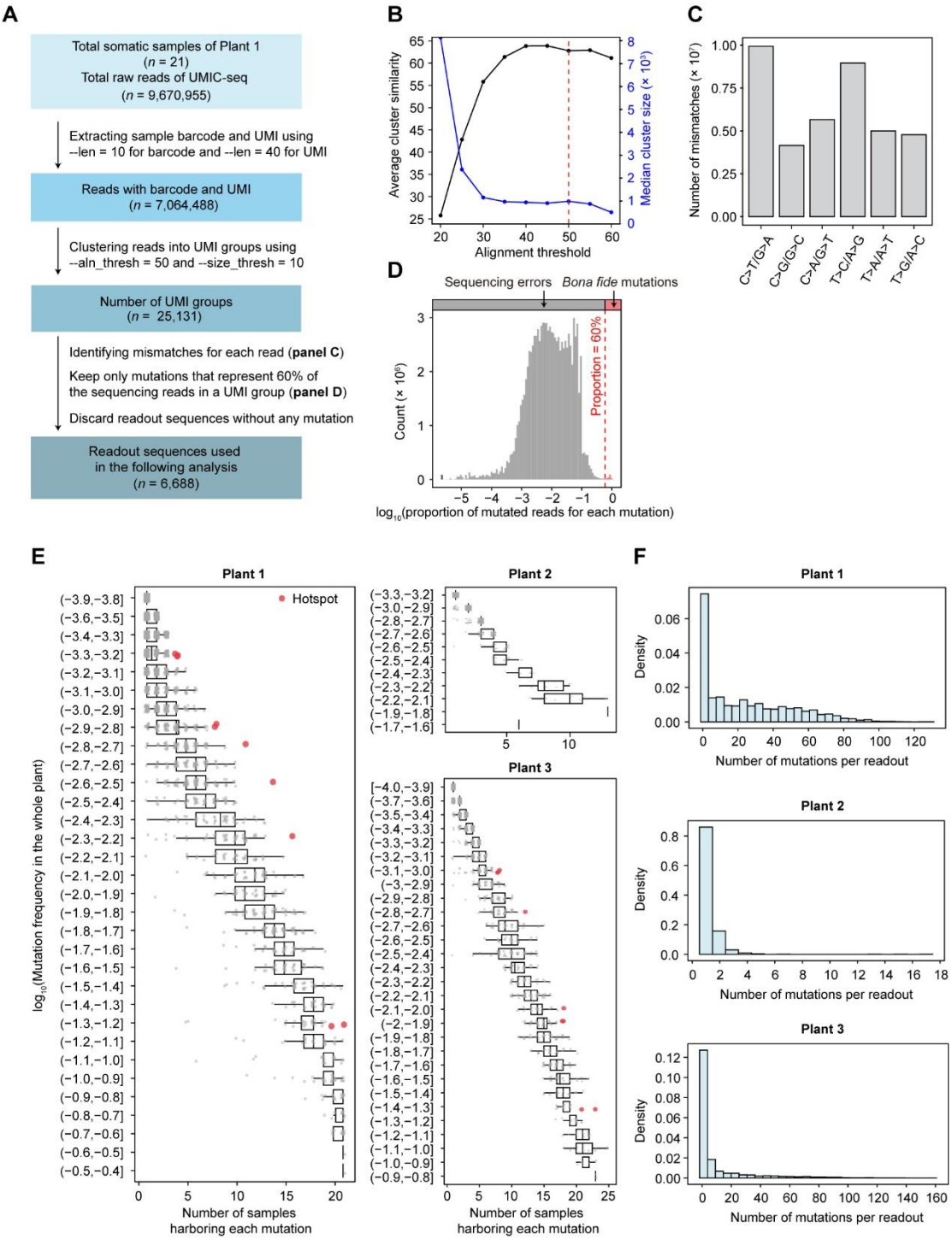


Figure S7. Identification of mutations from UMIC-seq data. Related to STAR Methods.

(A) Flowchart illustrating the steps involved in processing the sequencing data, from raw reads to the final consensus sequences.

(B) The line chart depicts the sequence similarity among reads within corresponding clusters. It also includes the median number of reads at each alignment threshold.

(C) Bar plot illustrating the molecular spectrum of raw mutations identified from UMIC-seq data,

without filtering out PCR and sequencing errors.

(D) Histogram representing the distribution of the proportion of mutated reads for each mutation. Plant 1 was used as an example throughout (A)–(D).

(E) This boxplot identifies hotspot mutations, defined as upper outliers detected in all branches of a T_1 plant, by plotting the number of leaf samples harboring each mutation against its $\log_{10}$-transformed frequency in the whole plant.

(F) Histograms showing the distribution of the mutation number across all somatic readouts for each plant. Mutation number per readout is shown on the *x*-axis; probability densities are shown on the *y*-axis.