PNAS RESEARCH ARTICLE | PLANT BIOLOGY

Quantitative and sensitive sequencing of somatic mutations induced by a maize transposon

Justin Scherer[a,b], Michael Hinczewski[c], and Brad Nelms[a,b,d,1]

Edited by James Birchler, University of Missouri, Columbia, MO; received January 22, 2025; accepted June 20, 2025

Cells accumulate mutations throughout development, contributing to cancer, aging, and evolution. Quantitative data on the abundance of de novo mutations within plants or animals are limited, as new mutations are often rare within a tissue and fall below the limits of current sequencing depths and error rates. Here, we show that mutations induced by the maize Mutator (Mu) transposon can be reliably quantified down to a detection limit of 1 part in 16,000. We measured the abundance of millions of de novo Mu insertions across four tissue types. Within a tissue, the distribution of de novo Mu allele frequencies was highly reproducible between plants, showing that, despite the stochastic nature of mutation, repeated statistical patterns of mutation abundance emerge. In contrast, there were significant differences in the allele frequency distribution between tissues. At the extremes, root was dominated by a small number of highly abundant de novo insertions, while endosperm was characterized by thousands of insertions at low allele frequencies. Finally, we used the measured pollen allele frequencies to reinterpret a classic genetic experiment, showing that evidence for late Mu activity in pollen is better explained by cell division statistics. These results provide insight into the complexity of mutation accumulation in multicellular organisms and a system to interrogate the factors that shape mutation abundance.

somatic mutation | transposon | tissue development | maize

Multicellular organisms accumulate mutations throughout development, producing genetic heterogeneity within and between tissues. With the increased sensitivity to detect de novo mutations through sequencing, it has become clear that genetic mosaicism is ubiquitous even in healthy individuals (1–8): Over 1,000 single-base substitutions are present per adult human fibroblast (1) and megabase-sized structural variants can be observed in 30% of healthy human neurons (2). In plants, low-frequency mutations can be transmitted to the next generation (3), and preexisting (somatic) mutations contribute to variation between plants regenerated in tissue culture (4).

To interpret and predict the effect of de novo mutations, it is critical to understand what influences their abundance and spread within the organism. This is challenging for both biological and technical reasons. Biologically, mutation accumulation is complex and depends on processes that impact the initial mutation rate (e.g., mutagenic exposure, DNA repair) as well as the spread of mutations once they arise (e.g., cell division, selection, cell death). While there have been theoretical advances in understanding how these factors interact to shape mutation abundance (9–15), there is a need for quantitative, empirical data to constrain and inform the theory.

This is where the technical challenge comes in: New mutations can span many orders of magnitude in abundance, down to 1 per cell, pushing the limits of current sequencing depths and error rates. To date, genome-wide studies on de novo mutation in plants and animals have reported detection limits around 1 to 5% (2–8) (*SI Appendix*, Fig. S1*A*), which cover only the most abundant mutations. Targeted sequencing has helped bridge this gap (16–20), with an inverse relationship between genomic coverage and sensitivity to detect rare mutations (*SI Appendix*, Fig. S1*A*). However, targeted sequencing still suffers from a limited dynamic range, as more abundant mutations are unlikely to occur within a narrow genomic region.

Mutations caused by transposable elements (TEs) play important roles in evolution, contributing to genome-size evolution (21), alleles selected during crop domestication (22), and the origin of new genes through exon shuffling (23). Unlike other classes of mutation, TE insertions introduce defined sequences into the genome that can be targeted by PCR (24). The potential of this is significant: By selectively amplifying only genome sequences containing the TE, de novo insertions can be identified without sequencing through an overwhelming number of wild-type copies at the same location (*SI Appendix*, Fig. S1*B*).

Significance

New mutations provide the raw material for evolution and contribute to cancer, aging, and genetic diseases. It has been challenging to follow the origin and spread of new mutations because they can be exceptionally rare and difficult to detect. By focusing on a class of mutation that can be detected more readily—Mu transposon insertions—we followed the abundance of new mutations in multiple maize tissues. We find that the Mu has broad activity across tissues, but with significant tissue-specific differences in how abundant individual new mutations become. Most mutations were below the detection limit available for other classes of mutation. These results provide a glimpse into the complexity of mutation within multicellular organisms.

Author affiliations: [a]Department of Genetics, University of Georgia, Athens, GA 30602; [b]The Plant Center, Office of Research, University of Georgia, Athens, GA 30602; [c]Department of Physics, Case Western Reserve University, Cleveland, OH 44106; and [d]Department of Plant Biology, University of Georgia, Athens, GA 30602

Author contributions: B.N. designed research; J.S. performed research; J.S., M.H., and B.N. analyzed data; and B.N. wrote the paper with input from the other authors.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: nelms@uga.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2426650122/-/DCSupplemental.

Published August 6, 2025.

This shares advantages of targeted mutation sequencing without needing to focus on a predefined genomic region.

Here, we evaluate the maize Mutator (Mu) transposon as a quantitative model of de novo mutation accumulation in multicellular tissues. Mu has long been a valuable model in maize genetics (25–28) because of its high forward mutation rate, availability of Mu-active and inactive genetic stocks, and ease of identifying Mu insertion locations by sequencing (24, 29). Mu is a class I (DNA) transposon that predominately transposes duplicatively, i.e., transposing to new locations without loss of the donor element (30); this apparent "copy-and-paste" behavior is thought to be caused by DNA repair pathways that restore the original sequence after transposition (25, 26). Mu transposes into unlinked sites, with no preference to insert near its site of origin (30).

We find that Mu sequencing can accurately measure the absolute allele frequencies of de novo Mu insertions within complex tissues, with a sensitivity, dynamic range, and error rate that are orders of magnitude better than currently possible for single-base substitutions. We then measured the allele frequency distribution for de novo Mu insertions in leaf, root, pollen, and endosperm. Mu had broad activity in all four tissues, with no evidence for a preference of late insertion in pollen. These results provide a rich dataset with which to test and refine theoretical models of mutation accumulation in multicellular organisms and highlight the importance of tissue type in shaping the abundance of de novo mutations during development.

Results

Sensitive Identification of De Novo TE Insertion Sites. To identify Mu TE insertions, we established a sequencing assay based on MuSeq (24), which has been widely used to map Mu insertions in maize genetic stocks (24, 31, 32). MuSeq applies nested PCR to specifically amplify and sequence DNA fragments that span the TE-genome boundary (*SI Appendix*, Fig. S2). We optimized MuSeq for quantifying the abundance of rare, de novo TE insertions that are heterogeneous within a tissue sample. Two key changes were implemented in "MuSeq2"; first, we introduced molecule counting by incorporating unique molecular identifiers (33) (UMIs) during an initial adapter ligation step (*SI Appendix*, Fig. S2). This makes it possible to identify and remove PCR duplicates, improving quantitative accuracy. Second, we limited the amplification of non-Mu products using suppression PCR, providing more specific TE amplification with fewer PCR cycles (*SI Appendix*, Fig. S3).

To test MuSeq2, we first applied it to seedling leaves from Mu-active and inactive maize lines. Samples were sequenced to a mean of 1.7 and 2.8 million TE-spanning molecules per Mu-active and inactive plant, respectively. For the inactive plants, all Mu elements are expected to map to a fixed set of genomic locations, representing historical TE insertions. Indeed, 99.8% of molecules from Mu-inactive samples mapped to only 29 locations (*SI Appendix*, Table S1). In contrast, Mu-active plants had Mu elements mapping to a wide range of new genomic sites (*SI Appendix*, Fig. S4 *A* and *B*), with a mean of 184,432 insertion sites detected per leaf. These can be confirmed as bona fide Mu insertions because i) the TE border was consistently sequenced along with the genomic region (*SI Appendix*, Fig. S4*C*) and ii) 123,312 sites were sequenced out of both directions of the TE, including the 9 bp target site duplication that is characteristic of Mu insertions (26). To estimate the error rate of MuSeq2, we leveraged the fact that Mu-inactive lines have a negligible rate of new Mu transposition, providing a genetic control for no transposon activity. Assuming that all molecules mapping outside the 29 historical locations were false positives, the error rate of MuSeq2 is 0.11 ± 0.04 falsely identified insertions per diploid cell (mean ± SE; N = 3 Mu-inactive plants). This corresponds to 2.7×10^{-11} false positive insertions per bp, two orders of magnitude lower than the most accurate duplex methods to measure single-base substitutions (34).

De Novo and Inherited Mu Insertions across Matched Tissues. We next applied MuSeq2 to leaf, pollen, endosperm, and root from Mu-active plants. Using sequential tissue isolations and controlled genetic crosses (Fig. 1*A*), we were able to separate de novo insertions from inherited ones and further divide the inherited insertions by parent-of-origin. First, the plants were generated from a cross between a Mu-inactive female and Mu-active male; the female parent contributes a defined set of historical insertions (Fig. 1*B* and *SI Appendix*, Table S1), and so all other insertion sites in the offspring were either de novo or paternally inherited. To distinguish de novo from paternally inherited insertions, we used matched tissues with early and well-defined divergence times. The endosperm, which comprises the bulk of maize seed mass, inherits its paternal DNA from a sister sperm cell during double fertilization (Fig. 1 *A*, *Left*). Thus, insertions present at high abundance in both endosperm and embryo-derived tissues must be paternally inherited (hereafter: "paternal insertions"). Indeed, paternal insertions were well separated from de novo insertions based on their abundance in both endosperm and other tissues from the same plant (Fig. 1 *B* and *C*).

The initial output of MuSeq2 is the relative abundance of different Mu insertion sites within a sample. To convert relative abundances (UMI counts) to absolute allele frequencies (variant allele frequency; VAF), we normalized the data using the paternal insertions (*SI Appendix*, Fig. S4*A*), which have a known allele frequency within the sample: 0.5 (heterozygous) in leaf, root, and pollen and 0.33 in triploid endosperm (this normalization is insensitive to realistic Mu excision rates; *SI Appendix*, Fig. S5). The measured allele frequencies were reproducible between technical replicates (independent libraries prepared from the same DNA; Fig. 1*D*), with strong quantitative agreement across 5 orders of magnitude ($R^2 = 0.997$ for all insertions; $R^2 = 0.955$ for de novo insertions). In contrast, there was no correlation in the abundance of de novo insertions between matched tissues from the same plant (Fig. 1*C* and *SI Appendix*, Fig. S6), reflecting their independent and recent origin. In total, MuSeq2 measured TE allele frequencies down to a detection limit of 6.0×10^{-5} for the median sample (1 part in 16,569; *SI Appendix*).

Allele Frequencies Measured in Pollen Match Paternal Inheritance Patterns in the Offspring. To benchmark the accuracy of the measured allele frequencies, we compared allele frequencies in bulk pollen to paternal transmission patterns in the offspring (Fig. 2*A*). The number of de novo Mu insertions per pollen cell can be calculated as the sum of allele frequencies in pollen (Σvaf_i). We estimate 22.5 ± 3.8 de novo insertions per pollen cell in this Mu-active line (median ± SE). We then genotyped Mu insertions in the parents and offspring of four families generated from a Mu-active male parent crossed with a Mu-inactive female (N = 30 offspring total, 7 to 8 per family), making it possible to directly follow the paternal transmission of Mu insertions to the offspring (*SI Appendix*, Fig. S7). There were 27.8 ± 6.7 new Mu insertions per offspring, in agreement with the estimate from bulk pollen (Fig. 2*B*). Pollen allele frequencies also agreed with offspring transmission patterns for higher-order statistics, such as the number of de novo Mu insertions shared by any two siblings

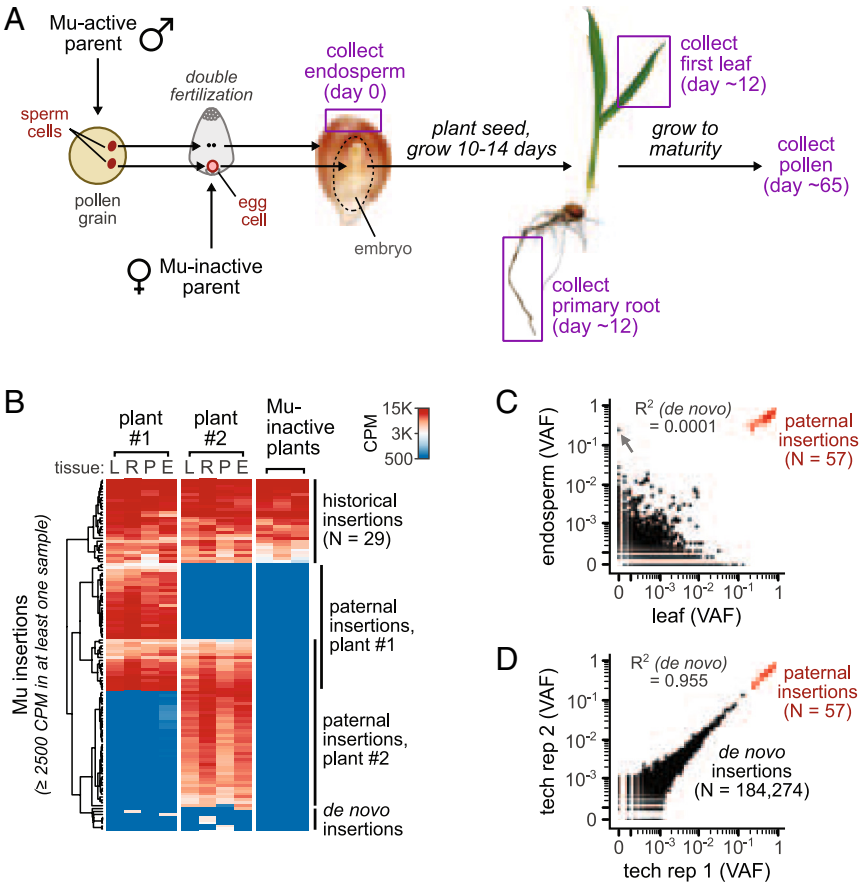

Fig. 1. Sensitive and quantitative assessment of de novo mutation abundance for an active maize transposon. (*A*) Cartoon of experimental design and tissue collection. Sequential, matched isolations from endosperm, and other tissues make it possible to distinguish inherited from de novo insertions, because endosperm is derived from a sister sperm cell during double fertilization. (*B*) Heatmap showing the abundance of Mu insertions in matched tissue samples from two siblings as well as control Mu-inactive plants. The Mu-inactive samples were from the family used as the female parent and represent historical insertions that were maternally inherited. All insertion sites with >= 2,500 CPM in at least one of the samples are shown. CPM, counts per million (number of TE-spanning molecules at a given genomic site). (*C*) Allele frequencies of Mu insertions for matched endosperm and leaf from a single plant. Paternal insertions were abundant in both samples, while de novo insertions were only abundant in one (e.g., gray arrow). Black dot, de novo insertion; red dot, paternal insertion; VAF, variant allele frequency. (*D*) Technical replicates for a representative leaf sample.

(Fig. 2*C* and *SI Appendix*, Fig. S7*D*; estimated from pollen as the sum of allele frequencies squared, Σvaf_i^2).

To put the Mu transposon activity in context, there are ~67 single-base substitutions per generation in maize (35), and similar per base substitution rates have been reported for *Arabidopsis* (36) and human (37). Thus, the number of TE insertions in this Mu-active line (20 to 30 per generation through pollen) is comparable to the background rate of single-base substitutions.

De Novo Mu Insertions Occur at a Wide Range of Allele Frequencies. A histogram of Mu allele frequencies for a representative leaf sample is shown in Fig. 3*A*. There were 211,097 de novo insertions detected in this single leaf, with allele frequencies ranging from 0.28 down to $<10^{-4}$ (the detection limit of the assay). These data suggest that Mu is active throughout development, including insertions that likely arose in the meristem [based on their high abundance (38); Fig. 3*A*] down to lower frequency insertions that more likely arose in the leaf itself (39). Mu has a strong preference to insert into and immediately upstream of genes (29), targeting a reduced portion of the genome. Despite this, we did not observe saturation of the available Mu target sites (*SI Appendix*, Fig. S8). The most abundant de novo Mu insertions occurred at uncorrelated, independent sites between samples (Fig. 1*C*). Thus, the specific mutations induced by Mutator were stochastic and infrequently repeated between plants. In contrast, the allele frequency distribution was reproducible across its entire range (Fig. 3*B*). Essentially, while the specific set of TE insertions varied widely, a predictable number of insertions were present at any given abundance.

Mu Insertion Activity Is Much Broader than for Mu Excisions. Most prior data on somatic Mu activity is based on the excision of Mu elements in the endosperm (25, 26, 40), which can be observed by the appearance of revertant purple sectors after Mu excises from an anthocyanin reporter gene (*SI Appendix*, Fig. S9). Endosperm excisions produce almost entirely small sectors (40), suggesting that Mu excision activity is highest later in development (25, 26). The excision rate also varies 1,000-fold between tissues, ranging from ~10% excisions per element in endosperm (40) down to $<10^{-4}$ excisions per element transmitted through pollen (26). Compared to excisions, de novo Mu insertions were much more broadly distributed across space and time (Fig. 4*A*). There was substantial new insertion activity in every tissue type, despite large divergence in the developmental origins and biology of the selected tissues. Furthermore, de novo Mu insertions were observed at a wide range of allele frequencies. While Mu excisions almost never occur above an allele frequency of 0.002 (40), Mu insertions were often observed beyond this limit, even within endosperm (Fig. 4*A* and *SI Appendix*, Fig. S9). Thus, Mu insertions and excisions behave differently, with new insertions occurring throughout development.

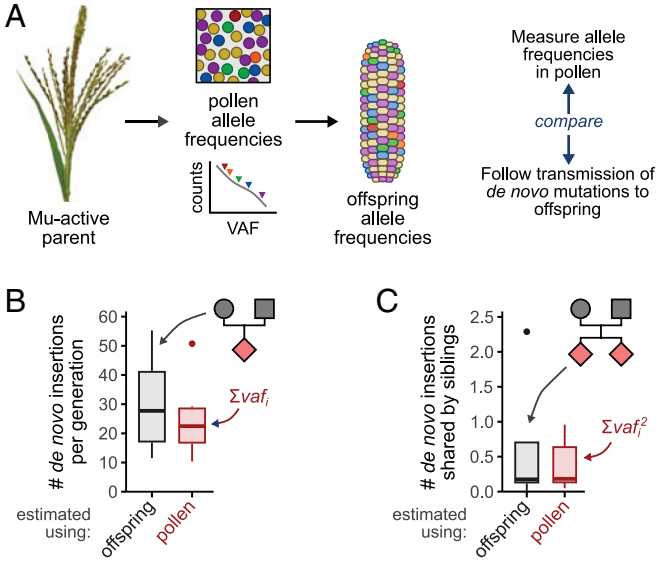

Fig. 2. Pollen allele frequencies match paternal inheritance patterns. (*A*) Allele frequencies in pollen should predict allele frequencies in the offspring. (*B*) Number of de novo insertions per generation, measured by genotyping parents and offspring (gray; see *SI Appendix*, Fig. S7*A*), or estimated from the pollen allele frequency data (red). The difference between estimates was not significant ($P = 0.71$, Mann–Whitney *U* test). (*C*) Number of paternal insertions shared by two siblings, measured by genotyping parents and offspring (gray; see *SI Appendix*, Fig. S7*D*) or estimated from the pollen allele frequency data (red). The difference between estimates was not significant ($P = 0.94$, Mann–Whitney *U* test). For panels *B* and *C*, N = 4 families with 7 to 8 offspring each; N = 9 pollen samples.

Tissues Show Distinct Allele Frequency Distributions for De Novo Mu Insertions. While not as dramatic as the divergence between excisions and insertions, there were significant differences in the behavior of de novo Mu insertions between tissues (Fig. 4). The total number of de novo insertions per cell varied by up to fourfold (Fig. 4*B* and *SI Appendix*, Fig. S10), ranging from 8.2 in root to 32.8 in leaf. Moreover, each tissue had a reproducible but distinct allele frequency distribution (Fig. 4*A*). Root was most dominated by insertions at high allele frequencies (Fig. 4 *B* and *C*), suggesting that new Mu insertions often formed relatively large sectors in this tissue. At the other extreme, endosperm had a much higher proportion of rare (low VAF) insertions. Thus, there was variation not only in the total number of Mu insertions but also in how widespread the individual insertions were throughout the tissue.

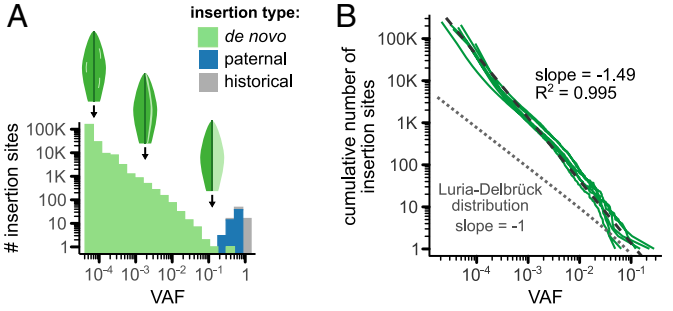

Fig. 3. De novo Mu insertions occur across a wide range of allele frequencies. (*A*) Histogram of Mu allele frequencies in a representative leaf sample. Colors indicate whether the insertion sites were historical, paternally inherited, or de novo. Leaf cartoons show the potential spatial distribution of mutations at selected VAFs, estimated from sector sizes in ref. 39. Insertions above a VAF of 2×10^{-3} likely originated in the meristem, based on estimates that 250 meristematic cells form a leaf primordia in maize (38). VAF, variant allele frequency. (*B*) Cumulative number of Mu insertion sites in individual leaf samples (N = 6). Dashed line, best linear fit to the log–log transformed data; gray dotted line, theoretical expectation for random mutation in an exponentially dividing cell population (Luria-Delbrück distribution).

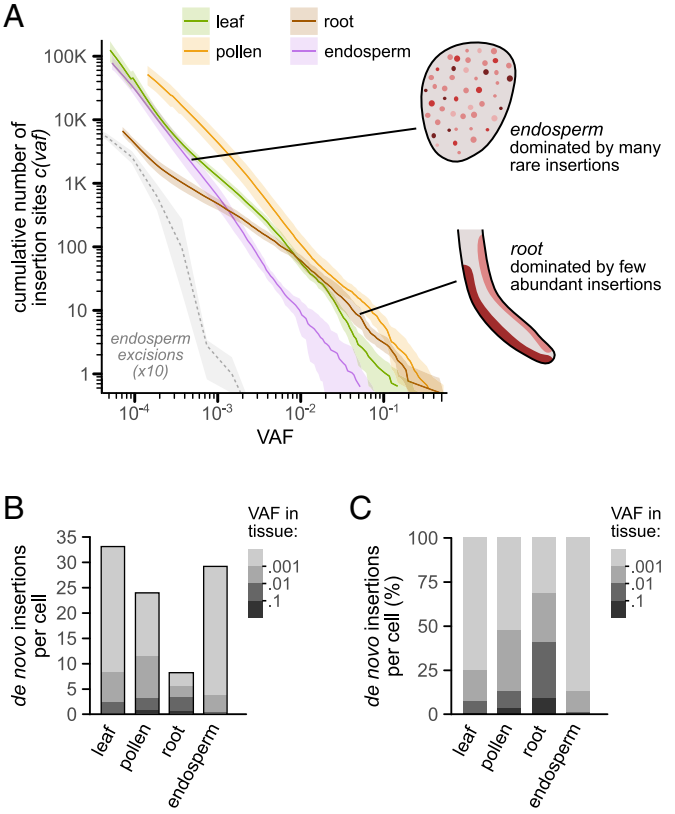

Fig. 4. Allele frequency distribution of de novo Mu insertions across maize tissue types. (*A*) Cumulative number of de novo Mu insertion sites at different allele frequencies. For insertion data (solid lines), curves show the mean and 95% CI (bootstrap test). For endosperm excision data (dotted line), the curve shows the reported values; shaded area, 95% CI assuming Poisson counting error. Endosperm excision data are from ref. 40; the reported number of excision events was multiplied by 10 to make the insertion and excision distributions easier to compare. VAF, variant allele frequency. (*B* and *C*) Number and % of de novo Mu insertions per cell, calculated from the sum of allele frequencies (Σvaf_i). These estimates are robust to sequencing depth (*SI Appendix*, Fig. S10) and are calculated for cells at the baseline DNA content in each tissue (2C for leaf/root, 3C for endosperm); thus, these are values for a cell that has not endoreduplicated and is in the G1 phase of the cell cycle. Colors indicate the contribution of insertions at different allele frequencies.

What might contribute to the observed allele frequencies? Since early studies on bacterial mutation by Luria and Delbrück, many theoretical models of mutation accumulation have been developed. We compared the empirical allele frequency distributions to established theory. Leaf, pollen, and endosperm all closely followed a linear relationship on a log–log plot (a power law distribution; Fig. 3*B* and *SI Appendix*, Fig. S11). Power-law relationships are well known in mutation accumulation (14), as this distribution occurs in an exponentially dividing cell population subjected to a constant rate of neutral mutations (a Luria–Delbrück process). However, the empirical data were a bad fit to the Luria–Delbrück model, because the slopes were far steeper than the theoretical expectation of −1 (14) (Fig. 3*B* and *SI Appendix*, Figs. S11 and S12). In animals, a common model for mutation accumulation is based on an exponential growth phase early in development, followed by a later, stable-population phase (9, 41–44); however, this model predicts a strong deviation from power law behavior and a shallow slope for much of the range, again a poor fit to the data (*SI Appendix*, Fig. S13). Other models of mutation accumulation, including boundary-driven growth (13, 14) (where cells divide preferentially at the edge of an expanding population),

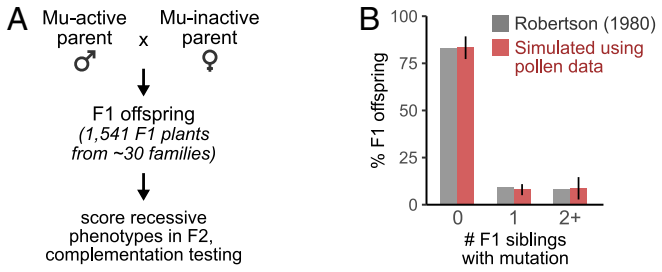

Fig. 5. Pollen allele frequencies are consistent with outcross data from classical genetics. (*A*) Experimental design from Robertson (27). F1 offspring were generated by outcrossing a Mu-active male parent, then the offspring were assessed for appearance of visible mutant phenotypes after self-fertilization. F1 siblings segregating similar mutant phenotypes were subjected to complementation testing to determine whether they shared the same (allelic) mutation. (*B*) The experimental design in A was simulated using mutant alleles randomly drawn with probabilities matching the measured pollen allele frequencies. The % of F1 offspring that share mutations with 0, 1, or 2+ siblings were then calculated and compared to Robertson (27). Error bars, SEM.

linear growth (9) (such as occurs during asymmetric stem-cell divisions), and the glandular fission model (12) (developed for solid cancer tumors) also predict sharp deviations from power law behavior. Given the complexity of multicellular development and transposon regulation, it is perhaps unsurprising that established theory cannot explain the data. The availability of quantitative allele frequency data across several orders of magnitude can inform and constrain future theoretical developments to understand mutation accumulation in plants.

Mu Outcross Experiments Can Be Explained by Cell Division Statistics. The classic view has been that Mu is most active late in germinal development (25–28), with activity peaking around the time of meiosis or during pollen maturation. In contrast, our data suggests that Mu insertions occurred throughout development (Fig. 4*A*), with >90% of insertions occurring prior to meiosis (*SI Appendix*). The most direct evidence for Mu activity late in germinal development comes from a study by Robertson (27), in which he outcrossed Mu-active plants and characterized new mutations in the F1 offspring (Fig. 5*A*). He identified 177 mutant F1 plants that segregated recessive seedling phenotypes; then, through extensive complementation testing, determined that 82.8% of the F1 offspring had unique mutations. The frequent occurrence of unique mutations among the offspring led to the idea that Mu must be most active late in development (25–28).

To reconcile these results, we directly compared our data to Robertson (27). We previously showed that pollen allele frequency data can predict inheritance patterns in the offspring (Fig. 2); this approach can also be used to predict more complex experimental designs, such as Robertson's. We simulated Robertson's experiment 1,000 times, randomly drawing new mutations at probabilities defined by the bulk pollen data (*SI Appendix*). On average, the simulations predict that 83.3% of F1 offspring would have unique mutations (Fig. 5*B*), in close agreement with the reported value ($P = 0.83$, two-tailed bootstrap test). An advantage of the simulated experiments is that it is possible to computationally go "back in time" and see how abundant any given mutation was in the Mu-active parent (*SI Appendix*, Fig. S14). For offspring that shared mutations with 2+ siblings, the source mutations had an average allele frequency of 0.13 in pollen [consistent with a mutation at the time the seed was planted (45)]; thus, the fact that Robertson observed any such offspring (8.2% of the total) suggests that early Mu activity occurred at an appreciable rate in this experiment.

Here, we can provide an alternative explanation for Robertson's data: In a dividing cell population, most mutations will be rare simply because there are more cells later in development and therefore more opportunities for a mutation to occur. While there is one chance for a mutation in the zygote, there are two in the following division, then four, and so forth. This will lead to an exponentially increasing number of mutations at decreasing allele frequencies, as was observed in all tissues for Mu insertions (Fig. 3*A*) and is even predicted by the Luria-Delbrück distribution (Fig. 2*B* and *SI Appendix*, Fig. S15). Rather than evidence for tissue-specific activity, the preponderance of unique mutations in Mu outcross families is better explained by the statistics of cell division.

Discussion

De novo mutations are difficult to identify because they can be extremely rare within a tissue. This has led to an acute depth-vs.-breath trade-off (46), where mutations can either be sequenced to lower depth across the genome or at higher depth for targeted loci. Here, we overcame this technical barrier with a strategic model system—the maize Mu transposon. We show that Mu sequencing can accurately measure the absolute allele frequency of de novo TE insertions genome-wide, while achieving a detection limit rivaling the most sensitive of targeted mutation studies (18).

As a model of mutation, a limitation of our approach is that it is only applicable to TEs; however, several findings are likely to be representative of other mutation classes. First, there were a large number (>100,000) of de novo Mu insertions per sample. If single-base substitutions could be sequenced to the same depth, a similar number of events might be expected. It has been estimated that every gene is mutated multiple times in an organism the size of maize or humans (9), a prediction consistent with data from deep sequencing single-genes (18). The number of de novo Mu insertions per cell was similar to the germline single-base substitution rate in maize (35) and far below the number of single-base substitutions per somatic cell in animals (34). Thus, Mu simply provides a glimpse into the scope of genetic mosaicism for an organism with a cell population measured in the trillions.

Second, most de novo mutations were present at low allele frequencies. A strong trend toward low frequency mutation is expected from the statistics of cell division, as there are exponentially more cells later in development and thus more chances for mutations to occur. While individually rare, these mutations can collectively add up to important effects and may contribute to aging, cancer, and evolution. Finally, tissues varied not only in the number of mutations per cell, but also in how widespread the mutations were. When considering the rise and spread of de novo mutations, it will be important to recognize that multicellular organisms are large, complex populations with extensive genetic heterogeneity.

Our results provide greater resolution into Mu activity across maize tissues. Mu insertions have been observed in somatic tissues such as leaf (29), but quantitative data on their number and abundance were not available. We found that Mu insertions occurred continuously throughout development in both somatic and germinal tissue. This is in contrast to Mu excisions, as there is a clear bias against early excision activity (40) and a >1,000-fold range in excision rates between endosperm (40) and pollen (26) (vs. fourfold maximum range for de novo insertions). This provides further evidence that Mu insertions can be decoupled from excision outcomes, perhaps due to tissue-specific differences in the use of DNA repair to restore a Mu element after a transient excision event (26).

What might drive the tissue-specific variation in Mu allele frequencies? Differences in transposon activity may contribute but

are not the only explanation. For instance, spatial biases in cell division rates have a profound impact on mutant allele frequencies (9, 11–13), and so differences in tissue development may contribute to the patterns we observed. Selection for and against specific mutations has been observed in healthy human tissues (16) and might similarly impact the persistence or spread of de novo Mu insertions. Future work can dissect the relative contribution of tissue-specific transposon activity, cell division patterns, selection, and other processes on the ultimate abundance of de novo mutations within and across the plant.

Materials and Methods

Sample Preparation and MuSeq2 Libraries. Mu-active and inactive lines were descended from Maize Co-op stocks 919J and 910I (respectively). DNA was extracted using a modified CTAB protocol (endosperm, pollen) or the Qiagen Dneasy Plant Mini kit (leaf, root) and then sheared with a Covaris E220 Sonicator. A MuSeq2-specific adapter (*SI Appendix*, Table S2) was ligated onto the DNA with the NEBNext Ultra II DNA Library Preparation Kit, and Mu-containing fragments were then selectively amplified with nested PCR using primers that target a conserved region at the edge of Mu elements (*SI Appendix*, Table S3). More detailed methods are available in *SI Appendix*.

Data Mapping and Analysis. Reads were mapped to the W22 genome (47) and converted to molecular counts using a UMI added during adapter ligation. Paternal insertions were then identified as insertion sites detected at >1,000 counts per million in both endosperm and another matched tissue for a given plant. To convert to allele frequencies, the number of UMIs at each insertion site was divided by the mean UMIs for the paternal insertions and then multiplied by 0.33 (for endosperm) and 0.5 (other tissues). Detailed analysis methods are described in *SI Appendix*.

Data, Materials, and Software Availability. Sequencing data are available at NCBI GEO under accessions GSE279993 (48) and GSE296286 (49). Code for mapping and analysis can be found at 10.5281/zenodo.15635051 (50).

ACKNOWLEDGMENTS. We thank Jonathan Gent, Robert Erdmann, Bob Schmitz, and Chris McFarland for invaluable discussions and critical reading of the manuscript. We thank Grant Freeman for initial testing of the MuSeq2 protocol. We thank the Duke University Sequencing and Genomics Technologies Core for sequencing services. Funding was provided by NIH grant R35GM151237 to B.N.

1. A. Abyzov *et al.*, One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. *Genome Res.* **27**, 512–523 (2017), 10.1101/gr.215517.116.
2. M. Schmid *et al.*, Mosaic copy number variation in human neurons. *Science* **342**, 632–638 (2013).
3. S. Schmitt *et al.*, Low-frequency somatic mutations are heritable in tropical trees *Dicorynia guianensis* and *Sextonia rubra*. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2313312121 (2024), 10.1073/pnas.2313312121.
4. K. R. Amundson *et al.*, Differential mutation accumulation in plant meristematic layers. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.09.25.559363 (Accessed 6 January 2025).
5. K. Grimes *et al.*, Cell type-specific consequences of mosaic structural variants in hematopoietic stem and progenitor cells. *Nat. Genet.* **56**, 1134–1146 (2024), 10.1038/s41588-024-01754-2.
6. K. Siudeja *et al.*, Unraveling the features of somatic transposition in the Drosophila intestine. *EMBO J.* **40**, 1–19 (2021), 10.15252/embj.2020106388.
7. L. Moore *et al.*, The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021), 10.1038/s41586-021-03822-7.
8. T. Bae *et al.*, Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2018), 10.1126/science.aan8690.
9. S. A. Frank, Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1725–1730 (2010), 10.1073/pnas.0909343106.
10. L. Shahriyari, N. L. Komarova, The role of the bi-compartmental stem cell niche in delaying cancer. *Phys. Biol.* **12**, 055001 (2015), 10.1088/1478-3975/12/5/055001.
11. K. Chkhaidze *et al.*, Spatially constrained tumour growth affects the patterns of clonal selection and neutral drift in cancer genomic data. *PLoS Comput. Biol.* **15**, e1007243 (2019), 10.1371/journal.pcbi.1007243.
12. R. Noble *et al.*, Spatial structure governs the mode of tumour evolution. *Nat. Ecol. Evol.* **6**, 207–217 (2022), 10.1038/s41559-021-01615-9.
13. M. A. Lewinsohn, T. Bedford, N. F. Müller, A. F. Feder, State-dependent evolutionary models reveal modes of solid tumour growth. *Nat. Ecol. Evol.* **7**, 581–596 (2023), 10.1038/s41559-023-02000-4.
14. D. Fusco, M. Gralka, J. Kayser, A. Anderson, O. Hallatschek, Excess of mutational jackpot events in expanding populations revealed by spatial Luria-Delbrück experiments. *Nat. Commun.* **7**, 12760 (2016), 10.1038/ncomms12760.
15. W. Postek, K. Staskiewicz, E. Lilja, B. Waclaw, Substrate geometry affects population dynamics in a bacterial biofilm. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2315361121 (2024), 10.1073/pnas.2315361121.
16. I. Martincorena *et al.*, Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018), 10.1126/science.aau3879.
17. A. L. Young, R. Spencer Tong, B. M. Birmann, T. E. Druley, Clonal hematopoiesis and risk of acute myeloid leukemia. *Haematologica* **104**, 2410–2417 (2019), 10.3324/haematol.2018.215269.
18. R. Salazar *et al.*, Discovery of an unusually high number of de novo mutations in sperm of older men using duplex sequencing. *Genome Res.* **32**, 499–511 (2022), 10.1101/gr.275695.121.
19. J. H. Bae *et al.*, Single duplex DNA sequencing with codec detects mutations with high sensitivity. *Nat. Genet.* **55**, 871–879 (2023), 10.1038/s41588-023-01376-0.
20. G. Waneka, B. Pate, J. G. Monroe, D. B. Sloan, Exploring the relationship between gene expression and low-frequency somatic mutations in *Arabidopsis* with duplex sequencing. *Genome Biol. Evol.* **16**, evae213 (2024), 10.1093/gbe/evae213.
21. N. V. Fedoroff, Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758–767 (2012), 10.1126/science.338.6108.758.
22. A. Studer, Q. Zhao, J. Ross-Ibarra, J. Doebley, Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* **43**, 1160–1163 (2011), 10.1038/ng.942.
23. R. L. Cosby *et al.*, Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* **371**, eabc6405 (2021), 10.1126/science.abc6405.
24. D. R. McCarty *et al.*, Mu-seq: Sequence-based mapping and identification of transposon induced mutations. *PLoS One* **8**, e77172 (2013), 10.1371/journal.pone.0077172.
25. D. Lisch, Mutator and MULE transposons. *Microbiol Spectra.* **3** (2015), 10.1128/microbiolspec.mdna3-0032-2014.
26. D. Lisch, Mutator transposons. *Trends Plant Sci.* **7**, 498–504 (2002), 10.1016/S1360-1385(02)02347-6.
27. D. S. Robertson, The timing of Mu activity in maize. *Genetics* **94**, 969–978 (1980).
28. D. S. Robertson, Mutator activity in maize: Timing of its activation in ontogeny. *Science* **213**, 1515–1517 (1981).
29. X. Zhang, M. Zhao, D. R. McCarty, D. Lisch, Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes. *Nucleic Acids Res.* **48**, 6685–6698 (2020), 10.1093/nar/gkaa370.
30. D. Lisch, P. Chomet, M. Freeling, Genetic characterization of the mutator system in maize: Behavior and regulation of Mu transposons in a minimal line. *Genetics* **139**, 1777–1796 (1995).
31. P. Liu, D. R. McCarty, K. E. Koch, Transposon mutagenesis and analysis of mutants in UniformMu maize (*Zea mays*). *Curr. Protoc. Plant Biol.* **1**, 451–465 (2016), 10.1002/cppb.20029.
32. C. Marcon *et al.*, BonnMu: A sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiol.* **184**, 620–631 (2020), 10.1104/pp.20.00478.
33. T. Kivioja *et al.*, Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011), 10.1038/nmeth.1778.
34. F. Abascal *et al.*, Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021), 10.1038/s41586-021-03477-4.
35. Y. Jiao *et al.*, Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **44**, 812–815 (2012), 10.1038/ng.2312.
36. S. Ossowski *et al.*, The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94 (2010), 10.1126/science.1180677.
37. H. Jónsson *et al.*, Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017), 10.1038/nature24018.
38. S. Poethig, "Cellular parameters of leaf morphogenesis in maize and tobacco" in *Contemporary Problems in Plant Anatomy*, R. A. White, W. C. Dickison, Eds. (Academic Pres, New York, 1984), pp. 235–259.
39. J. A. Langdale, B. Lane, M. Freeling, T. Nelson, Cell lineage analysis of maize bundle sheath and mesophyll cells. *Dev. Biol.* **133**, 128–139 (1989).
40. A. Levy, V. Walbot, Regulation of the timing of transposable element excision during maize development. *Science* **248**, 1534–1537 (1990).
41. S. A. Frank, M. A. Nowakt, Developmental predisposition to cancer. *Nature* **422**, 494 (2003), 10.1038/422494a.
42. R. Meza, J. Jeon, S. H. Moolgavkar, E. Georg Luebeck, Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16284–16289 (2008), 10.1073/pnas.0801151105.
43. M. E. Moeller, N. V. Mon Père, B. Werner, W. Huang, Measures of genetic diversification in somatic tissues at bulk and single-cell resolution. *Elife* **12**, RP89780 (2024), 10.7554/eLife.89780.
44. E. B. Gunnarsson, K. Leder, J. Foo, Exact site frequency spectra of neutrally evolving tumors: A transition between power laws reveals a signature of cell viability. *Theor. Popul. Biol.* **142**, 67–90 (2021), 10.1016/j.tpb.2021.09.004.
45. R. S. Poethig, E. H. Coe, M. M. Johri, Cell lineage patterns in maize embryogenesis: A clonal analysis. *Dev. Biol.* **117**, 392–404 (1986), 10.1016/0012-1606(86)90308-8.
46. G. Gydush *et al.*, Massively parallel enrichment of low-frequency alleles enables duplex sequencing at low depth. *Nat. Biomed. Eng.* **6**, 257–266 (2022), 10.1038/s41551-022-00855-9.
47. N. M. Springer *et al.*, The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* **50**, 1282–1288 (2018), 10.1038/s41588-018-0158-0.
48. J. Scherer, B. Nelms, Sequencing de novo Mu transposon insertions across maize tissues. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE279993. Deposited 21 October 2024.
49. J. Scherer, B. Nelms, Mu transposon insertion sites in matched leaf and endosperm samples from Mu-active W22 maize. Gene Expression Omnibus. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE296286. Deposited 5 May 2025.
50. J. Scherer, B. Nelms, Quantitative-and-sensitive-sequencing-of-somatic-mutations-induced-by-a-maize-transposon. Zenodo. https://doi.org/10.5281/zenodo.15635051. Deposited 10 June 2025.

SI APPENDIX

Literature survey on VAF detection limit vs. genomic coverage

For **Fig. S1A**, only papers that reported VAFs in the main text or figures were considered. The selected studies were not meant to be exhaustive, but rather representative of a range of techniques and mutation types. VAF detection limits were as given in each paper; when multiple limits were provided for different sample types (e.g. tissues), the lowest reported value was used. For targeted sequencing studies, genomic coverage was calculated from the reported target size divided by the mappable genome size[1]: 2,864,785,220 bp for human and 119,482,012 bp for *Arabidopsis*. For the purposes of the figure, whole-genome sequencing papers were considered to have 100% genomic coverage.

Two classes of study were not included: First, studies that identified *de novo* mutations from transcript data were excluded (e.g. ref 2). RNA-seq produces very uneven read depths across the genome, and so there is not a well-defined relationship between VAF detection limit and genome coverage – highly expressed genes, which represent a small portion of the transcriptome, dominate the minimum observed VAFs. Single-cell mutation studies were also not included[3]. This was not due to active exclusion, but rather because available single-cell mutation papers largely do not report VAF detection limits. While single-cell methods provide additional cell-type resolution and information about cell-to-cell variation, they do not fundamentally overcome the limitations on sequencing depth and error rates that are also present in bulk experiments.

Plant growth and tissue collection

Mu-active plants were maintained by continual outcrossing of Mu-active pollen onto Mu inactive ears, using the bz1-Mum9 anthocyanin reporter to confirm Mu activity. The Mu-inactive maintainer (female) parents were descended from maize Co-op stock 910I, and carry the sh1-bb1981 and bz1-m4::Ds alleles. Mu-active seeds were descended from maize Co-op stock 919J, which carries a mutable bz1-Mum9 allele. Both stocks were originally ordered from the Stock Center in January 2010 by Jonathan Gent. Continual outcrossing of Mu active lines onto Mu inactive ears was required to maintain Mu activity. Mu active kernels were phenotypically identified by the speckling pattern that occurs when Mu somatically excises from the bz1-Mum9 allele.

For tissue collection, kernels were chipped using a razor blade and the resulting endosperm samples were stored in 2 mL tubes and frozen. To minimize sample cross-contamination, the surface that kernels were chipped on was wiped with a 10% bleach solution and razor blades were only used once. Chipped kernels were then planted in vermiculite (Therm-O-Rock Vermiculite 3A-HORT Medium). After the second seedling leaf was fully emerged (V2 stage; 10-13 days after planting), plants were removed from vermiculite. Roots were rinsed thoroughly in water to remove any vermiculite, and the bottommost ~1 inch of primary root (up to and including the root tip) was collected into 2 mL tubes and frozen. The topmost half of the first leaf (~3/4 inch) was also harvested and collected into 2 mL tubes and frozen. Seedlings were then transplanted to soil in the Botany Greenhouses in Athens, Georgia, where they were grown in sunlight supplemented with LED fluorescent lights (Medic Grow 550W Slim Power 2) until maturity. At maturity, pollen was collected into 2 mL tubes in the morning from the plants at first pollen shed (9-10 am) and frozen.

DNA isolation

Leaf DNA isolation: Leaf tissue was disrupted in a 2 mL tube using liquid nitrogen and a pestle (Agilent cat. no. PES-15-B-SI). Once disrupted, DNA was extracted with the Qiagen DNeasy Plant Mini Kit (Qiagen cat. no. 69104) and eluted in two steps, first with 30 ul of elution buffer followed by 25 uL elution buffer. DNA size distributions were evaluated using 5 ul of sample on a 0.8% agarose gel.

Root DNA isolation: Root tissue was disrupted with a Qiagen Tissue Lyser II and three to six 3 mm glass beads per sample. Prior to disruption, the sample box was chilled overnight at -80 °C with root samples and 3 mm glass beads inside. Pre-chilled root tissue was then shaken in the Tissue Lyser at max frequency for 5 minutes. Samples were removed from the shaker and agitated using a pestle to dislodge root debris from the tube walls. Shaking was repeated at max frequency for 5 minutes. After this process, root DNA was extracted as described for leaf isolation using the Qiagen DNeasy Plant Mini Kit. DNA size distributions were evaluated on a 0.8% agarose gel.

Endosperm DNA isolation: Genomic DNA was isolated from endosperm using a modified CTAB DNA extraction protocol. To prepare CTAB buffer, CTAB stock was made with 3% Cetyltrimethyl ammonium bromide, 1.4 M NaCl, 20 mM EDTA (pH 8.0), 100 mM Tris-Cl (pH 8.0). The day of DNA extractions, 2% w/v polyvinylpyrrolidone (PVP, MW 40 kDa) was dissolved into CTAB stock by heating the solution to 65 °C, and then 800 ul preheated lysis buffer was aliquoted into tubes (one tube per sample to be processed). Then 8 ul proteinase k (ThermoFisher cat. no. EO0491) and 1 ul beta-mercaptoethanol were added to each tube.

Endosperm tissue was disrupted using liquid nitrogen, mortar, and pestle. Disrupted tissue was then incubated at 65º C for 1 h in the preheated lysis buffer. Samples were inverted to mix every 10 minutes during incubation. Following this, samples were spun down at 5000 rcf for 8 mins to pellet tissue debris. Lysate was transferred to a new tube using a metal spatula and combined with 1 volume of a 24:1 chloroform isoamyl alcohol solution. Samples were mixed by inversion for 5 minutes and then centrifuged at 8,000 rcf for 10 minutes. The upper aqueous phase was carefully transferred to a new tube following centrifugation. To precipitate DNA, 0.7X volumes of cold isopropanol was added to each sample and inverted to mix. Samples were incubated at -20 °C for 1 hour. Samples were then centrifuged at 10,000 rcf for 15 minutes. The supernatant was removed and the DNA pellet was washed using 1000 ul of freshly prepared 70% ethanol. Samples were inverted to mix and incubated at room temperature for 5 minutes. Samples were then centrifuged for 5 minutes at 10,000 rcf. The ethanol wash was repeated one more time and DNA pellets were dried until the pellet became translucent. The DNA pellet was resuspended using 55 ul of ultra-pure H2O (ThermoFisher cat. no. 10977015) and incubated overnight at 4 °C. Size distributions were visualized using 5 ul of purified DNA on a 0.8% agarose gel. The first batch of endosperm samples showed signs of cross-contamination; these data were used to identify paternal insertions but excluded from all other analyses (see ‘Sample assessment and quality control’). Prior to processing subsequent endosperm samples, the mortar, pestle, and metal spatula were incubated for 5 min in 10% bleach solution and then thoroughly rinsed with water; this additional washing step removed the cross-contamination.

Pollen DNA isolation: Pollen was disrupted using a Qiagen Tissue Lyser II as described for root. During disruption, pollen debris would stick to the lid of the tubes and so a pestle was used to scrape off the debris into the tube. After disruption, 800 ul of preheated CTAB lysis buffer (prepared as described for endosperm) was added to each sample. A pestle was again used to scrape off any material from the tube lid back into the tube as well as break up any pellet that had formed in the bottom of the tube. This step ensured a homogenous mixture during lysis, which greatly increased DNA quality and quantity. DNA extractions were then performed as described for endosperm, with the additional of third ethanol wash after DNA precipitation. The first batch of pollen samples had much lower sequencing depth compared to the other tissues (lower UMIs / sample). For subsequent samples, pollen DNA was purified an additional time with a Monarch DNA and PCR cleanup kit (New England Biolabs cat. no. T1030S); the DNA cleanup kit was performed after CTAB extraction and DNA shearing, prior to the end repair step in MuSeq2.

MuSeq2 adapter preparation

MuSeq2 adapter oligos (**Table S2**) were ordered from Integrated DNA Technologies, suspended to 100 uM in TE. The general adapter structure is as follows:

5'-[phos]rrrrrrrrrrUGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNbbbbbbbbbbT-3'

The 10 nt sequences labeled as strings of 'r' and 'b' are reverse complements of each other, allowing the adapter to form a hairpin with a 3' T overhang. These sequences vary by adapter (**Table S2**), providing a sample-specific barcode during adapter ligation. The series of 'N's is the 8 nt Unique Molecular Identifier (UMI). A uracil (U) near the 5' end makes it possible to cut the hairpin after adapter ligation. Read 2 begins at the UMI and continues through the sample barcode and subsequent genome sequence.

To prepare MuSeq2 adapters and anneal the hairpin, 7.5 ul of adapter oligo (100 uM) was diluted with 25 ul Duplex Buffer (Integrated DNA Technologies cat. no. 11-05-01-03) and 17.5 ul H2O. Diluted adapters were then placed in a thermocycler and incubated at 95 ºC for 2 minutes followed by a 0.1 ºC ramp down in 1-second intervals for 700 cycles, reaching a final temperature of 10ºC. Adapters were then stored at -80 ºC.

MuSeq2 library preparation

DNA samples were sheared using a Covaris E220 Evolution instrument in 50 uL of water. Shearing settings were optimized for each tissue to shear to a mean of 1000 bp. All tissues used settings of 2% Duty Factor with 200 cycles per burst. For pollen, the peak incident power was 140 and time was 50 seconds; Endosperm: 100 Peak Incident Power and 30 seconds; Leaf: 70 Peak Incident Power and 30 seconds. Root: 100 Peak Incident Power and 20 seconds. Concentrations and size distributions for sheared DNA was measured using an Agilent 4200 TapeStation with a D5000 screentape (Agilent cat. no. 5067-5589).

Sheared DNA (200-1000 ng / sample) was end-repaired using the NEBNext Ultra II DNA Library Preparation Kit (New England Biolabs cat. no. E7370L) according to manufacturer instructions, except that all reaction volumes were cut in half. MuSeq2 adapters were then ligated to the DNA using the same kit (NEBNext Ultra II) with half reaction volumes; a separate adapter was used for each sample, providing up to 48 sample-specific barcodes during the initial ligation step. After ligation, 1.5 uL USER enzyme (New England Biolabs cat. no. M5505S) and 2 uL Exonuclease 1 (New England Biolabs cat. no. M0293S) were added to each sample and the reaction was incubated at 37 for 15 min then 80 °C for 15 min. This step linearizes the hairpin adapters by cleaving at a uracil base, and the addition of Exonuclease 1 degrades residual unligated adapter to minimize carryover in subsequent PCR. Samples were then purified with Ampure XP Beads (Beckman Coulter cat. No A63880) using a bead:sample ratio of 0.8X. After bead purification, libraries were resuspended in 5 ul ultra-pure H2O.

Adpter ligated libraries were processed through 3 rounds of PCR to selectively amplify Mu-containing fragments and complete the Illumina adapter sequences (PCR primer sequences in **Table S3**). For the first PCR, 5 ul sample was mixed with 6 ul NEBNext Ultra II Q5 Master Mix (New England Biolabs cat. no. M0544S), 0.5 ul TIR6 primer (4.8 uM stock concentration; 0.2 uM final), and 0.5 ul UDz_i7 primer (4.8 uM stock concentration; 0.2 uM final). Reactions were pipetted to mix and incubated at 98 °C for 30 s, then 14 cycles of 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 30 s, followed by 72 °C for 2 min. To remove excess primers, 0.5 ul Exonuclease I was added and the tube was incubated at 37 °C for 15 min then 80 °C for 15 min.

For PCR2, an additional 4 ul of Q5 master mix was added along with 0.4 ul of Museq2_NestedTIR primer (10 uM stock), 0.4 ul of P7 primer (10 uM stock), and 2.7 ul of ultra-pure H2O. Reactions were pipetted to mix and incubated at 98 °C for 30 s, then 6 cycles of 98 °C for 10 s, 59 °C for 30 s, and 72°C for 30 s, followed by 72

°C for 2 min. To remove excess primers, 0.5 ul Exonuclease I was added and the tube was incubated at 37 °C for 15 min then 80 °C for 15 min.

For PCR3, 5 ul of PCR2 product was mixed with 25 ul Q5 master mix, 19.5 ul ultra-pure H2O, and 5.5 ul xGen indexed primer pairs (Integrated DNA Technologies xGen UDI Primers Plate 1, cat. no. 10005922). A distinct primer pair was added to each sample to allow for multiplexing. Reactions were pipetted to mix and incubated at 98 °C for 30 s, then 5-15 cycles of 98 °C for 10 s, 59 °C for 30 s, and 72°C for 30 s, followed by 72 °C for 2 min. The number of PCR cycles varied by sample and was determined using qPCR as follows: prior to PCR3, 5 ul of the prepared PCR reaction was withdrawn and mixed with 0.5 ul of a 1:1000 dilution of SYBR Green I DNA Gel Stain (Thermo Fisher cat. no. S7563) in water. The reaction aliquot with SYBR was then run on a BioRad CFX96 Real Time PCR Thermal Cycler for 25 cycles using the reaction conditions listed above. The cycle number for each PCR reaction was chosen to be within 30-80% of the plateau height, and the remaining PCR mix was run using the selected cycle number.

After PCR, libraries were cleaned up and size selected using magnetic beads. For cleanup, 50 ul Ampure XP beads were added to each PCR sample (1X ratio) and then the DNA was purified according to manufacturer instructions and eluted in 40 ul ultra-pure H2O. Size selection was then performed using SPRIselect beads (Beckman Coultier cat. no. B23317) with 0.6/0.8 bead ratios according to manufacturer instructions. DNA was eluted in a final volume of 20 ul and the size distribution and concentration was measured using an Agilent 4200 TapeStation with a D5000 screentape. Libraries were pooled to 15 nM such that each individual library was equally represented. Paired-end 150 bp sequencing was performed at the Duke University Sequencing and Genomics Technologies Core on an Illumina NovaSeq X Plus instrument with 20% PhiX spike-in.

Mu insertion mapping and quantification

For MuSeq2 libraries, read 1 contains the last 29 bp of the Mu transposon followed by genomic DNA sequence. Read 1 was first pre-processed by removing the first 23 bp, which contain transposon sequence matching the PCR primer, and moving bp 24-29 (the 'validation sequence') to the read header using Fastp v0.23.4[4] with all filters disabled. Read 2 contains the adapter ligated fragment with an 8 bp UMI, 11 bp sample-specific barcode, and genomic sequence. Read 2 was pre-processed using Fastp to move both the 8 nt UMI and 11 nt sample-specific barcode to the read header. During this step, adapter sequences were also trimmed and fragments with under 40 bp in read 2 were removed. Paired-end reads were then mapped to the W22 V2 genome[5] using Bowtie2 v2.5.4[6], with both mixed and discordant mapping disabled (--no-mixed --no-discordant). The UMI-tools v1.1.6[7] 'group' function was used to identify reads sharing the same UMI; to account for sequencing errors in the UMIs, the default 'directional-adjacency' method from the UMI-tools package was used (described in ref. 7).

Next, fragments were filtered using a custom R script to remove low quality mapping and PCR duplicates. Fragments were excluded if they had a mapping quality score <10 or a validation sequence that did not match 'TRTCTC' (the sequence at the edge of the Mu transposon). They were further excluded if they did not have an exact match to the sample-specific barcode added during adapter ligation. To remove PCR duplicates, fragments mapping to the same position (both read 1 and 2) with the same UMI were merged. The libraries in this study were intentionally over-sequenced to increase the amount of error-correcting from molecular counting, with a mean of 5.2 sequenced fragments per molecule (UMI). Each molecule was required to have a minimum support of at least 1/5 the average number of reads for a given sample; for the median library, this means that each molecule (UMI) was sequenced with a minimum of two reads.

Most Mu elements contain an intact Terminal Inverted Repeat (TIR) at both ends of the transposon, and so can be sequenced out of each direction. To connect molecules mapping to the left or right border of the same Mu element, the following steps were taken: First, for Mu elements present in the reference genome (N = 20; all

historical insertions), the left and right borders were defined based on the genome sequence. For other Mu elements, the left and right borders were connected by expecting a 9 bp target site duplication (TSD) to be generated during Mu insertion; this would result in both ends of the transposon mapping 8 bp apart in reverse orientation (for a 9 bp TSD, there is an 8 bp distance between positions 1 and 9). To allow for discrepancies in TSD length, we searched for cases where a left and right border were within 15 bp of the expected TSD, but where both borders had over 50-fold more molecule counts than the corresponding border 8 bp away. In such cases, the two borders with higher counts were considered to come from the same element. Deviations from the 9 bp TSD were rare, with only 273 such instances identified compared to more than 3 million with the 'ideal' 9 bp TSD. For Mu insertion sites supported by at least 10 transposon-spanning molecules, 92% were sequenced out of both directions and 8% were only supported by one TIR.

The total number of molecule counts mapping to either the left or right transposon border were then added to provide a single estimate for each element. A subset of elements were not sequenced effectively out of both directions, which could result in under-counting as there was only one border available for sequencing instead of the usual two. To adjust for this effect, we identified any elements where there was a greater than 2-fold difference in molecule counts between the left and right border after adding a pseudocount of 500. For these elements, the number of molecules was estimated as 2 times the greater of the left or right border counts. This process affected 422 elements (0.00013%). Finally, 18 elements were 'blacklisted' and removed from analysis (**Table S4**) because they were identified at moderate abundance (between 10-1000 counts per million) in over half of all samples or half of the Mu-inactive controls; many of the blacklisted sites were ancestral Mu elements with diverged sequences and would not be expected to amplify efficiently during MuSeq.

Estimating variant allele frequencies from Mu count data

To convert Mu insertion counts to variant allele frequencies (VAF), the data for each sample was first scaled to counts per million (CPM). Paternal insertions were identified as insertion sites with ≥1000 CPM in both endosperm and at least one matched sporophytic tissue (leaf, root, or pollen), excluding the 29 historical insertions (**Table S1**). CPM data were then normalized to VAF by dividing each sample by the mean CPM of the paternal insertions and multiplying by 1/2 (for leaf, root, and pollen) or 1/3 (for endosperm); the difference in normalization factor for endosperm is because the endosperm is triploid with a 2:1 maternal:paternal ratio, and so paternal DNA makes up 1/3 of the DNA in this tissue. The random error for normalization is estimated to be 6.2% (standard error of the mean for the paternal insertion sites).

Identifying transmitted *de novo* insertions from two-generation Mu families

Four families were created by crossing a Mu-inactive female parent with a Mu-active male. MuSeq2 libraries were prepared from endosperm of the male parent and both endosperm and leaf of 7-8 offspring as described above. After mapping and quantification, inherited insertion sites were identified as insertion sites with ≥500 CPM in leaf/endosperm of an offspring and ≥200 CPM in the matched leaf/endosperm (e.g. an insertion with 500 CPM in leaf and 200 CPM in the matched endosperm would qualify). Paternal genotype calls were made based on having ≥200 CPM in the paternal endosperm and inheritance in at least one offspring. The reason that a lower CPM threshold was used to genotype insertions in the two-generation families (compared to the other samples, as described in the "Estimating variant allele frequencies" section, above) was that these samples were sequenced to lower depth and the reduced thresholds avoided a few false negatives in the samples with the least sequencing depth. A matrix of genotype calls is available at GEO accession GSE296286. The strategy used to call de novo insertions transmitted from the Mu-active male parent is shown in **Fig. S7**.

Evaluating the sensitivity of MuSeq2

The detection limit for each sample was calculated as the VAF for a Mu insertion supported by a single transposon-spanning molecule, and ranged between 1.5×10^{-5} to 3.5×10^{-4} (median = 6.0×10^{-5}).

To estimate the % of Mu insertions captured during library prep, we compared the amount of input DNA to the number of transposon-spanning molecules sequence per paternal insertion (**Table S5**). The input DNA amount was converted into the number of genome equivalents using a conversion factor of 407.37 genome equivalents per ng total DNA; this factor was calculated based on a haploid genome size of 2.4×10^9 bp, an average molecular weight per bp of 651.98, and assumes all DNA is nuclear (e.g. the contribution of plastid and mitochondrial DNA is negligible). The number of genome equivalents was then converted into the expected number of times a paternally inherited transposon insertion would be recovered, if each transposon insertion contributed a single final molecule. On average, we estimate that 14.0%, 8.3%, 16.5%, and 16.7% of transposon insertions were sequenced in leaf, pollen, root, and endosperm (respectively), not accounting for loss during DNA purification. The lower % yield for pollen is likely explained because an extra column purification was performed after measuring the DNA concentration, while for every other tissue the DNA was directly input into library prep.

Yield per pollen grain: Other than yield during library preparation itself, loss can occur during DNA purification or because of incomplete sample loading (most libraries were prepared from a portion of the total isolated DNA). To estimate the total yield for pollen: each pollen sample was collected into a 2 mL tube and filled to the 100 uL mark. By weighing the mass of pollen in several samples collected this way, we estimate average mass of pollen per sample was 120 mg. To determine the number of pollen grains per mg, we diluted pollen samples in defined volumes of water and then counted the number of pollen grains in a 5 uL aliquot. From this, we estimate there are 1750 pollen grains / mg, or ~210,000 pollen grains per sample (1750 pollen grains / mg x 120 mg / sample). If every Mu insertion were sequenced once, then we would expect a pollen sample to result in 315,000 transposon-spanning molecules per heterozygous paternal insertion (210,000 pollen grains x 3 copies of the genome / pollen grain x 0.5 heterozygous insertions / genome copy); we observed an average of 4052 molecules per paternal insertion in pollen (**Table S5**), and so ~1 in 78 insertions were sequenced.

Estimated yield at each step of MuSeq2: With 100% yield after DNA purification, one pollen sample would thus produce 1550 ng total DNA (210,000 pollen grains x 3 genome equivalents / pollen grain / 407.37 genome equivalents / ng). After CTAB purification, we obtained an average of 630 ng DNA, or 41% yield. Libraries were then prepared from 45% of the total DNA; if higher total % yield were desired, then the entire DNA fraction could be used by simply scaling up volumes and using more enzyme. A second column purification was then performed (only for pollen), and we estimate a yield of 53% for the column purification by assuming that the reduced number of molecules sequenced for pollen compared to all other samples is explained by the extra column purification step (see above). Finally, there was a ~7.5% yield for converting DNA to sequencable insertions (15% yield per Mu insertion divided by 2 ends per Mu element = 7.5% yield per transposon border). One major source of loss during library preparation is the presence of sheared fragments where the transposon end is not the correct distance from the fragment border. DNA was sheared to 1000 bp and the final libraries were size selected to 250-500 bp (100-350 bp genomic DNA insert + 150 bp adapter sequence). If the transposon end were randomly distributed relative the the sheared DNA end, we would expect it to fall within 100-350 bp of the end 25% of the time, and so a 4-fold loss in yield is explainable simply by random shearing. As a result, the 7.5% yield during library prep is 25% yield from shearing size and 30% yield from incomplete conversion during adapter ligation, cleanup, and PCR. In summary:

~210,000 pollen grains → 41% yield, CTAB DNA purification → 45% of DNA used for library prep →

53% yield, column cleanup → 15% yield per insertion (7.5% yield per transposon end) during library prep

Contribution of pre-meiotic insertions to mutations transmitted through pollen

A transposon insertion during or after meiosis could occur in a maximum of 6 molecules (if the insertion were before meiotic S-phase). Given that only 1 in 78 insertions were sequenced, the probability of sequencing a meiotic or post-meiotic insertion more than once is nearly zero ($p = 0.003$; Poisson distribution with lambda = 6/78). Therefore, any insertions sequenced more than once occurred prior to meiosis. Out of ~22.5 insertions per pollen grain (Fig. 2, 4), an average of 19 (85%) were supported by insertions sequenced more than once (Σvaf_i for all insertions supported by ≥2 molecules). The remaining 15% can be attributed to a combination of meiotic, post-meiotic, and pre-meiotic insertions. At the detection threshold (78 molecules), over half of the remaining divisions are pre-meiotic (3-4 mitotic divisions + meiosis + 2 pollen divisions = ~78 molecules after replication). If the Mu transposition rate was equal throughout these late divisions, this would imply that slightly over half of the remaining 15% of insertions are pre-meiotic and the other half are post-meiotic. There are several uncertainties in estimating how to divide up this final 15% of insertions, including the potential for changing Mu activity over time and the fact that not all pollen grains were collected (the 210,000 pollen grain samples were maybe 25% of the total pollen shed in a given day). Conservatively, we estimate a minimum of 1/3 of the remaining insertions occur during the late pre-meiotic divisions and so the contribution of pre-meiotic insertions is >90% (best estimate, ~95% of insertions are pre-meiotic).

Sample assessment and quality control

In total, 46 Mu-active samples were collected and sequenced for this study. All samples were used when identifying inherited insertions (paternal and historical insertion sites), but 17 were excluded from further analysis because of concerns with library quality: 6 samples were excluded because they did not meet a minimum VAF detection threshold of 10^{-3}. This set included the first 5 pollen samples, which had consistently low molecule counts, and one endosperm sample. Subsequent pollen libraries incorporated an additional round of DNA purification (see ‘DNA isolation’ section, above) and resulted in much better sequencing depth. Second, the first 11 endosperm libraries were excluded because they showed evidence of cross-contamination; in these libraries, paternal insertions from one library consistently showed unusually high abundance in the others. For subsequent endosperm libraries, all non-disposable items used for tissue disruption (mortar, pestle, metal spatula) were subjected to a more stringent washing protocol that included soaking in 10% bleach for 5 min (see ‘DNA isolation’ section, above); this additional cleaning step resolved the cross-contamination issue. These sample exclusion criteria were set prior to analyzing the data further.

Interpreting the allele frequency distribution of *de novo* Mu insertions

To estimate the mean allele frequency distribution for each tissue (e.g. **Fig. 4A**), the cumulative number of *de novo* Mu insertion sites at or above a given VAF was first calculated for the individual samples. The single-sample allele frequency distributions were then log-transformed and interpolated at 200 evenly spaced points between $\log_{10}(10^{-5})$ and $\log_{10}(1)$ using the R function *approx* (R version 4.3.0). The mean and 95% confidence interval (CI95) for each tissue was then calculated by bootstrapping with 2000 bootstrap replicates. As the sequencing depth varied between samples, the mean was reported down the minimum VAF covered by at least 75% of samples in a tissue. Power-law fits to the allele frequency distributions were performed using the R *lm* function after $\log_{10}$ transformation.

Simulating the Robertson (1980) experiment from pollen allele frequency data

Robertson (1980) performed a series of outcrosses between Mu-active plants (F0) and Mu-inactive donors. The F1 progeny were then evaluated to determine if they segregated new mutations and whether any of the mutations were shared with siblings. In total, 1541 F1 offspring were tested, of which there were 171 mutant plants (11.1%) carrying an estimated 154 distinct mutations.

To simulate the results of one Mu outcross from Robertson (1980), a MuSeq pollen sample was first randomly selected. This sample represents a single Mu-active F0 plant from Robertson's study. Mutations (Mu insertion sites) were then randomly drawn based on the measured pollen allele frequencies. For instance, a mutation with a VAF of 0.1 was drawn with a 10% chance of occurring in each F1 offspring. After simulating 50 such F1 offspring (roughly the average number of offspring evaluated per outcross in ref. 8), the number of times a mutation occurred 1, 2, or >3 times among the offspring was recorded. Robertson's entire study had ~30 such outcrosses, for a total of 1541 F1 plants. Thus, to simulate a full iteration of Robertson's study, 30 simulated outcross experiments were performed using 30 different pollen samples (randomly sampled with replacement) and the totals were added together.

To estimate confidence intervals, it is important that the simulated study reflects the variation expected under the conditions of Robertson (1980). From the pollen data, an average of 32,038 mutations were recovered for each simulation, far more than the 154 mutations recovered by Robertson (1980). This discrepancy is explained because Robertson tracked mutations with visible seedling phenotypes, which would represent only a small portion of the total. To better match the counting noise during Robertson (1980), the simulated mutations were downsampled so that an average of 154 were recovered per simulation. This downsampling makes the simulation-to-simulation variation better matched to Robertson (1980), but does not affect the mean estimates: 83.1% of mutations were found to be unique prior to downsampling, compared to 83.3% after downsampling.

Simulating the Robertson (1980) experiment assuming a Luria-Delbrück process

The Robertson (1980) experiments was simulated assuming a Luria-Delbrück process (constant mutation rate under exponential cell division) for Fig. S15. To simulate the results of one Mu outcross from Robertson (1980), plant growth under mutation was simulated as a series of 16 mitotic divisions, followed by meiosis and two pollen mitotic divisions. The choice of 16 mitotic divisions was because this is the minimum number of divisions required to obtain the population size of a typical pollen sample in this study (~250,000 pollen grains corresponds to 2^16 pollen mother cells after the mitotic divisions and then 4 pollen grains per meiosis). Prior to each cell division, a random number of mutation was simulated according to a poison process with lambda (mutation rate) equal to 2 * (171/1541) / 19 = 0.0117; this mutation rate results in a matched number of mutant F1 offspring as observed by Robertson, where the factor of 2 is because of meiotic reduction (there will be twice as many mutations in the diploid pollen mother cells as the number found in offspring), 171/1541 is the proportion of F1 mutant plants identified by Robertson (171 mutant F1 out of 1541 F1 plants), and the factor of 19 accounts for the total number of simulated cell divisions (16 mitotic divisions + meiosis + 2 pollen divisions). For meiosis, a single round of mutation was followed by two divisions and then mutations were retained with a probability of 0.5 (this simulates meiotic reduction where there is a 50% chance of any given mutation appearing in a meiotic product). Pollen mitoses were simulated with the same mutation rate but without an ensuing cell division as only the lineage leading to the fertilizing sperm cell will contribute to the offspring. After simulating cell division and mutation, 50 'cells' were randomly drawn and the number of mutations shared by 0, 1, or 2+ sibling was calculated. This was repeated 30 times to simulate a single round of the Robertson (1980) experiment, and then the whole experiment was simulated 1000 times to determine the mean and standard error of the mean.

SI REFERENCES

1. deepTools: Effective genome size https://deeptools.readthedocs.io/en/develop/content/feature/effectiveGenomeSize.html.

2. Rockweiler, N.B., Ramu, A., Nagirnaja, L., Wong, W.H., Noordam, M.J., Drubin, C.W., Huang, N., Miller, B., Todres, E.Z., Vigh-Conrad, K.A., et al. (2023). The origins and functional effects of postzygotic mutations throughout the human life span. Science *380*, eabn7113. 10.1126/science.abn7113.

3. Muyas, F., Sauer, C.M., Valle-Inclán, J.E., Li, R., Rahbari, R., Mitchell, T.J., Hormoz, S., and Cortés-Ciriano, I. (2024). De novo detection of somatic mutations in high-throughput single-cell profiling data sets. Nat. Biotechnol. *42*, 758–767. 10.1038/s41587-023-01863-z.

4. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. 10.1093/bioinformatics/bty560.

5. Springer, N.M., Anderson, S.N., Andorf, C.M., Ahern, K.R., Bai, F., Barad, O., Barbazuk, W.B., Bass, H.W., Baruch, K., Ben-Zvi, G., et al. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. Nat. Genet. *50*, 1282–1288. 10.1038/s41588-018-0158-0.

6. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. 10.1038/nmeth.1923.

7. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

8. Robertson, D.S. (1980). The timing of Mu activity in maize. Genetics *94*, 969–978.

9. Levy, A., and Walbot, V. (1990). Regulation of the timing of transposable element excision during maize development. Science *248*, 1534–1537.

10. Eden, M. (1961). A two-dimensional growth process. Dyn. Fractal Surfaces *4*, 598. 10.1111/j.2517-6161.1965.tb00610.x.

11. Poethig, R.S., Coe, E.H., and Johri, M.M. (1986). Cell lineage patterns in maize embryogenesis: A clonal analysis. Dev. Biol. *117*, 392–404. 10.1016/0012-1606(86)90308-8.

SI FIGURES

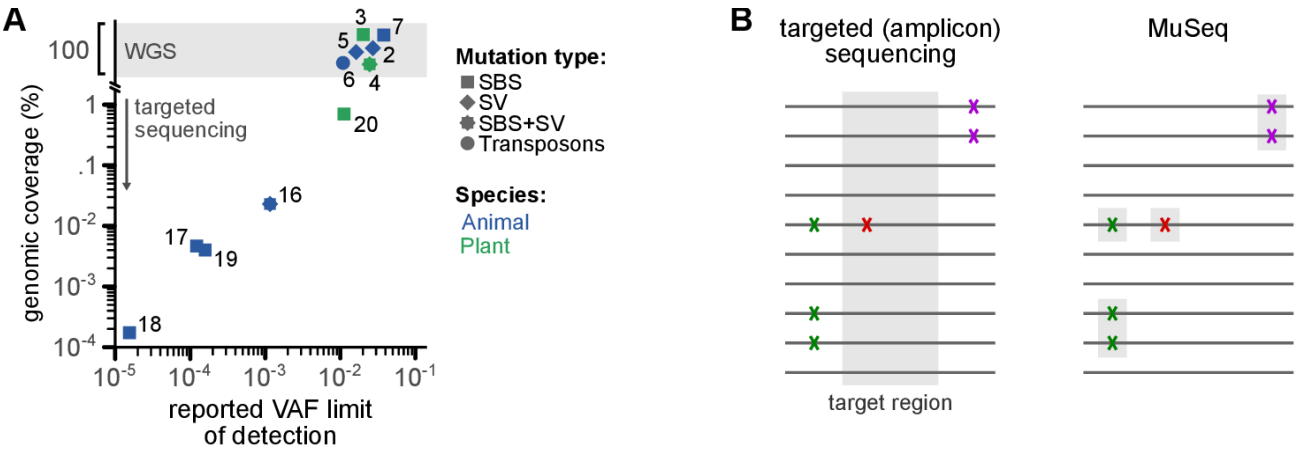

Figure S1. Sequencing depth limitations make it difficult to assess rare *de novo* mutations

(**A**) Relationship between VAF detection limit vs genomic coverage for selected studies. Numbers reflect the citation number in the main text references. Here, '100% genomic coverage' implies there was not intentional selection for a subset of the genome; in practice, this means the 'mappable genome' and excludes regions that are repetitive or otherwise difficult to amplify or sequence. VAF, variant allele frequency; SBS, single-base substitution; SV, structural variant.

(**B**) Comparison between targeted (amplicon) sequencing and MuSeq. While both approaches limit the sequencing to a portion of the genome, they do so in different ways. In these cartoons, 10 example DNA sequences are illustrated as dark gray lines; there are three mutations at different abundances, colored as green, red, and purple 'X's. For targeted sequencing, the 'X's could represent any class of mutation (SBS, SV, transposon); for Museq, these must be Mu transposon insertions. The region targeted by each technique is highlighted in gray.

Targeted sequencing selects a predefined set of genome loci to sequence deeply. Both wild type and mutant alleles are sequenced and any mutations outside of the target region are missed. MuSeq, in contrast, sequences transposon insertion sites throughout the genome, and reduces sequencing depth by avoiding the wild-type (transposon-free) alleles. In this hypothetical example, targeted sequencing would require at least 10 reads but only capture a single mutation; MuSeq would require fewer reads yet would capture all 3 mutations. While MuSeq is limited to transposons (Mu in this case), the opportunity is that it enables orders of magnitude greater sensitivity and dynamic range than is possible for other classes of mutation.

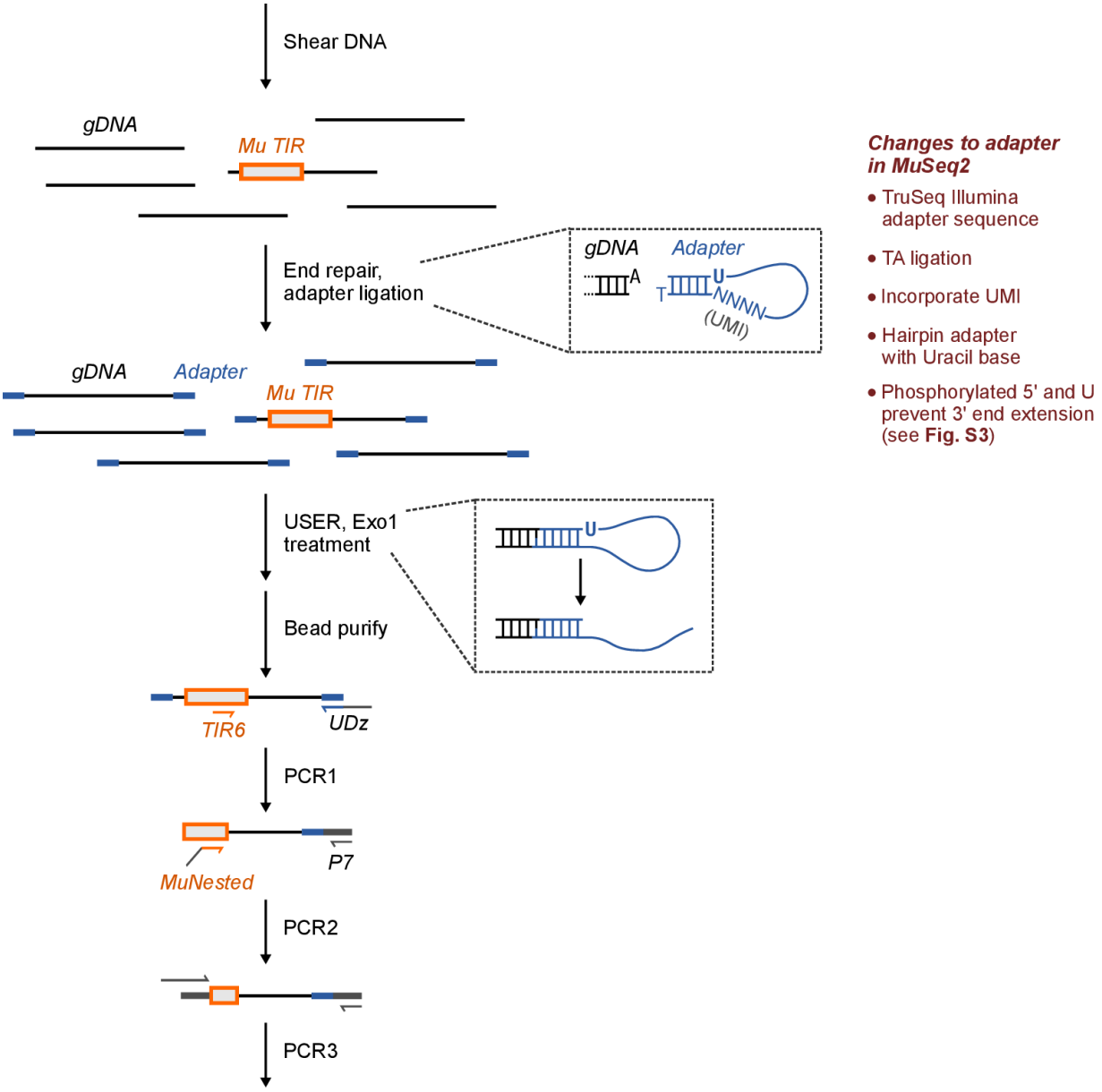

Figure S2. Overview of the MuSeq2 protocol

MuSeq2, similar to MuSeq, uses an adapter ligation followed by a series of nested PCR reactions to specifically amplify fragments spanning the transposon genome junction. Several changes to the adapter were made in MuSeq2, including incorporating a Unique Molecular Identifier (UMI) for molecular counting and updating to modern Illumina adapter sequences. USER treatment cleaves the adapter at the Uracil, making the ends compatible with the downstream PCR reactions.

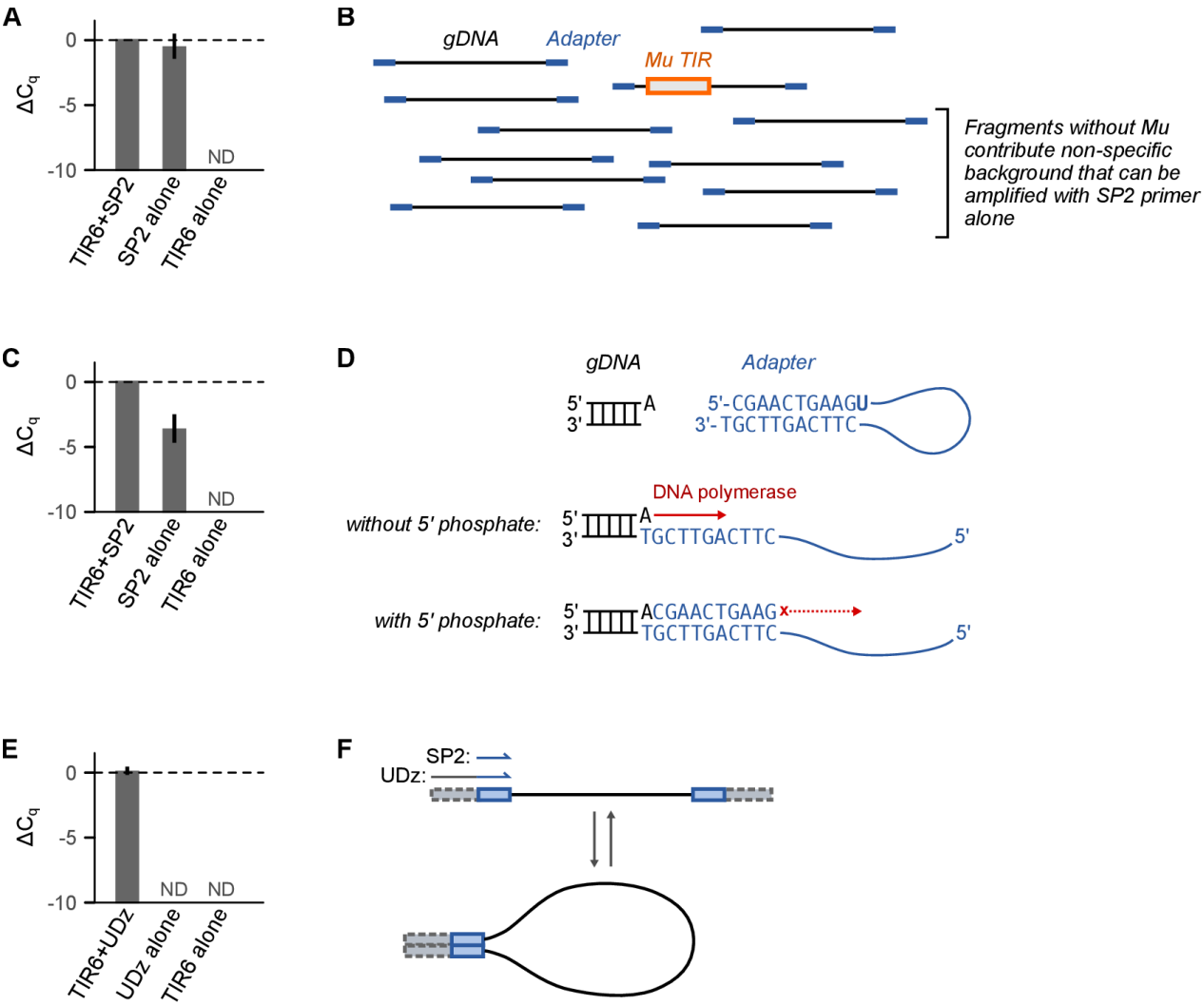

Figure S3. Reducing non-specific amplification through changes to adapter structure and suppression PCR

(**A**) Quantitative PCR for libraries prepared using an adapter design similar to the original MuSeq: the adapter was unphosphorylated and the reverse primer (SP2) matches the adapter-ligated sequence only. TIR6 is a Mu-specific primer; the SP2 primer matches the Illumina sequence ligated onto sheared genomic DNA. Three independent DNA samples were sheared and ligated, then the ligated DNA was split for qPCR with different primer combinations. The amplification is not specific, as SP2 alone amplifies similarly to when the Mu-specific primer was included. Cq was normalized to TIR6+SP2; in this experiment, the average number of cycles at ΔCq = 0 was 16.4.

(**B**) The reason for non-specific amplification with the SP2 primer is that the majority of DNA fragments from sheared genomic DNA do not contain a Mu element. Fragments without Mu are estimated to outnumber Mu-containing fragments by more than 10,000 fold. The background fragments will have adapter DNA on both sides, forming a potential priming site for PCR with the SP2 primer. The adapter structure limits background amplification in part because it has a 5' overhang and does not initially have the sequence needed for primer binding (the primer binds to the reverse complement of the overhang; this was also true in the original MuSeq); however, the 5' overhang can be copied by DNA polymerase at the start of PCR. This is likely an inefficient process, but given the excess of fragments without Mu it still contributed meaningful background (panel A).

(**C**,**D**) One modification to reduce non-specific background in MuSeq2 was to replace the unphosphorylated oligo with a phosphorylated one. With a 5' phosphate on the adapter, both adapter strands can be ligated to the sheared genomic DNA. After a Uracil on the adapter is cleaved to release the hairpin, it leaves a 3' phosphate overhang. By ligating the adapter on both strands, there is no free 3' hydroxyl available – blocking extension by DNA polymerase. Panel **C** shows the same experiment as panel **A**, except using a phosphorylated adapter. Cq was normalized to TIR6+SP2; in this experiment, the average number of cycles at ΔCq = 0 was 16.2.

(**E**,**F**) A second modification in MuSeq2 was to use a longer primer, UDz, in place of SP2 during the first PCR. Fragments with adapter sequence on both sides do not amplify as efficiently because they have self-complementary ends and can form a hairpin (suppression PCR). The UDz primer adds the entire Illumina adapter sequence during PCR. Because this primer makes the self-complementary region longer, it favors hairpin formation and increases the amount of suppression PCR for non-specific fragments. Panel **E** shows qPCR using the same adapter-ligated samples as in panel **C**, except that PCR was performed with the UDz primer. Cq was normalized to TIR6+SP2 from panel **C** and so the ΔCq values are directly comparable between these panels. There was no decrease in specific amplification when switching to the UDz primer, but non-specific amplification was completely suppressed.

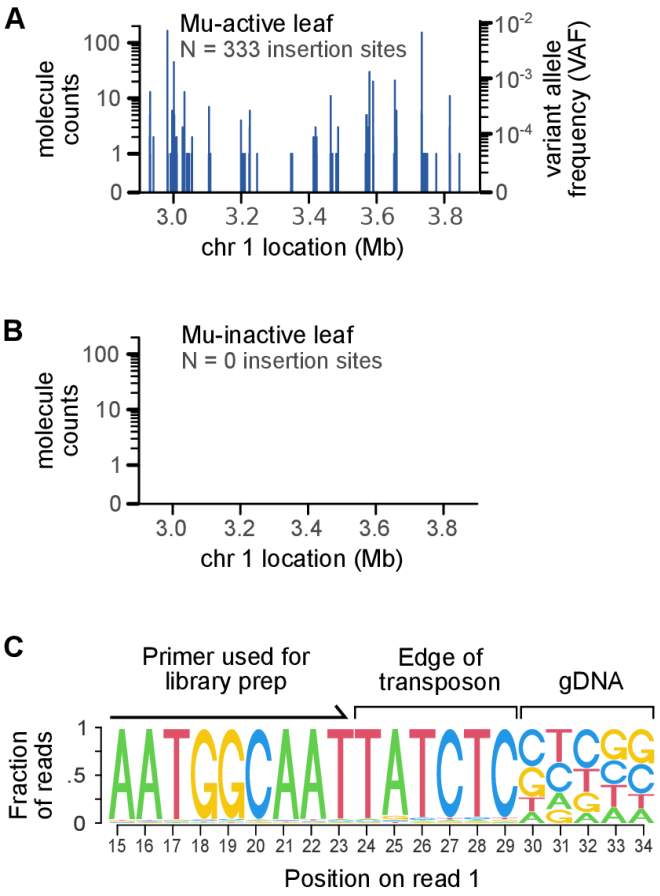

Figure S4. Specificity of MuSeq2

(**A**) A representative 1 Mb region showing all insertion sites identified in a single Mu-active leaf sample. In total, 333 insertion sites were found in this region, covering a range of molecule abundances (UMI counts). The y-axis on the left shows the normalized molecule counts while the axis on the right is normalized to variant allele frequencies (VAF). Molecule counts and VAF are directly proportional to each other; for this sample, the normalization was based on an average of 5860 +/- 240 molecule counts per heterozygous (VAF = 0.5) paternal insertion.

(**B**) The same region as in **A**, but for a Mu-inactive leaf sample. No insertion sites were observed in this region.

(**C**) Sequence composition for a portion of read 1, which covers the transposon-genome junction. The last 6 bp of the transposon were not included in any primer used during library prep, and provides independent validation that the sequencing is specific to Mutator. The 'validation sequence' matches the known transposon sequence TATCTC for the vast majority of reads.

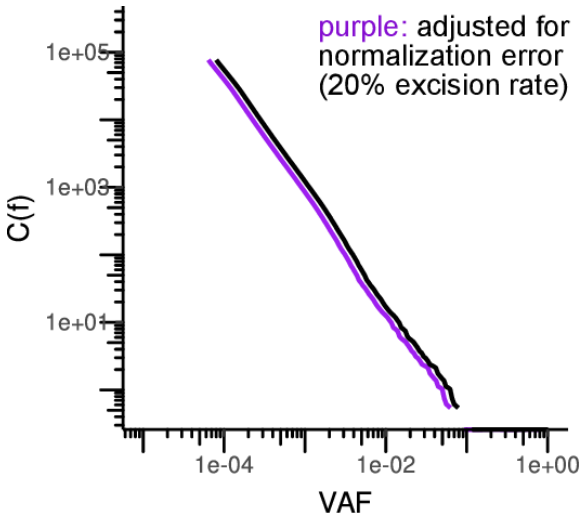

Figure S5. Mu excision rates have a negligible impact on normalization

Allele frequencies were normalized using the paternal insertions, which were assumed to be at their original abundance (VAF = ½ in most tissues, 1/3 in endosperm). In the presence of excisions, the true allele frequency of paternal insertions would be less, resulting in systematic error during normalization. We estimate the endosperm excision rate in our line is ~10%, based on the proportion of endosperm surface that has reverted to purple (this line carries a mutable bz1-Mum9 reporter allele that allows for purple pigment expression after excision). To be conservative, we used double this rate – 20% – and calculated the effect this would have on the measured allele frequencies. This figure shows the allele frequency distribution for a representative endosperm sample before (black line) and after (purple line) adjusting for normalization error due to a 20% excision rate. Even with double the observed excision rate, normalization error has a minimal impact on the allele frequency spectrum. This is because a change on the order of 20% is small when the measured frequencies vary by many orders of magnitude (log-scale). As a result, we did not consider excision further in our analyses; all main text results were not adjusted for excision (e.g. the black line above).

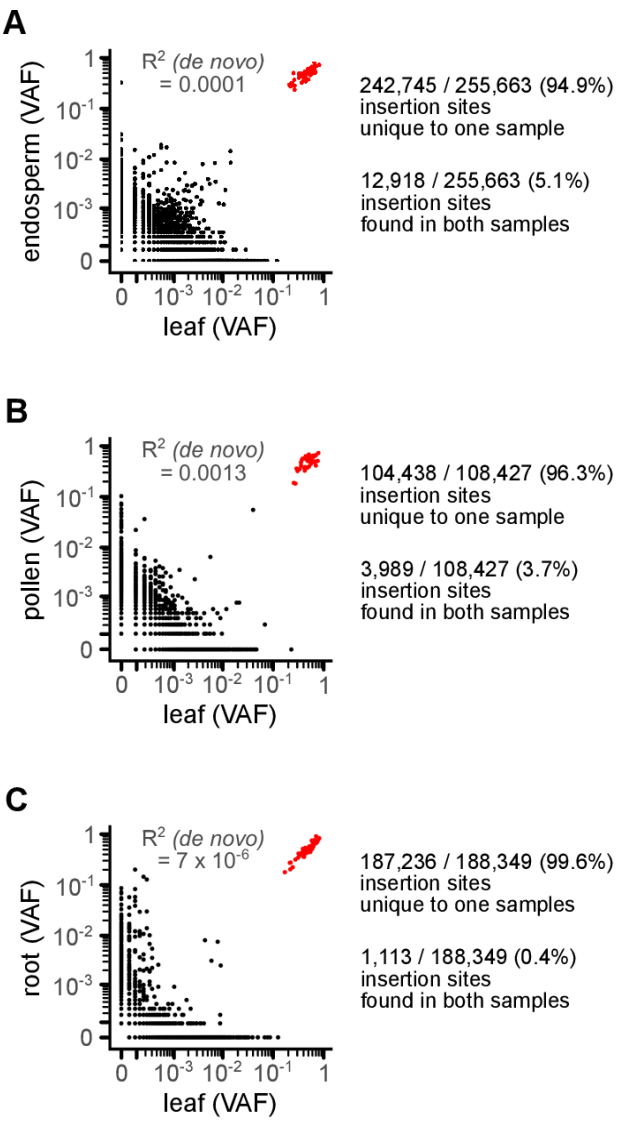

Figure S6. Comparison between insertion sites identified in matched tissues from the same plant

Panel (A) is the same as main text Fig. 1C, except that the number of insertion sites unique to one sample (either leaf or endosperm) vs shared between both are quantified. Panels (B) and (C) are similar, except they show a comparison between leaf and pollen (B) or root (C) from the same plant.

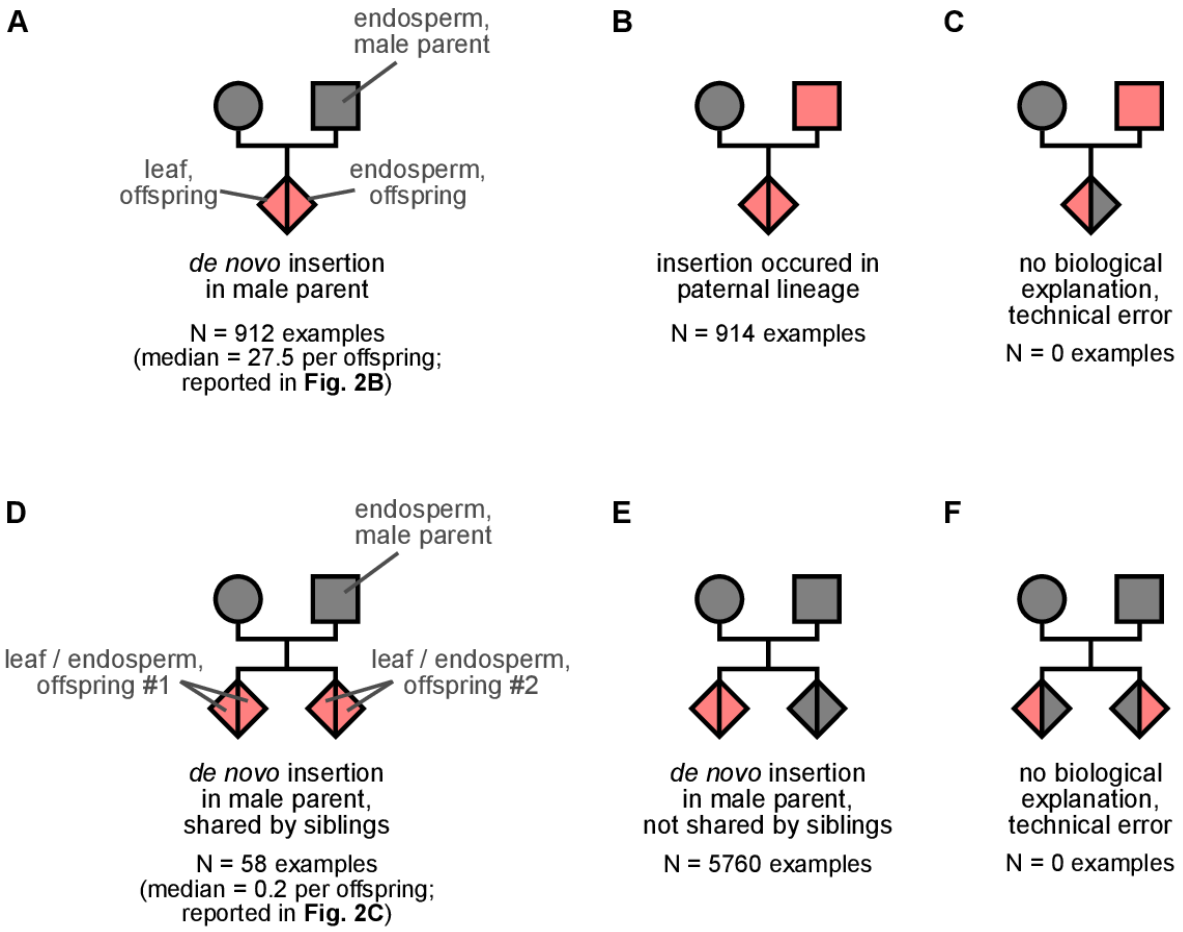

Figure S7. Transmission of *de novo* Mu insertions into the offspring.

Mu insertions were sequenced in the endosperm of Mu-active male parents and both endosperm and leaf of the offspring (N = 7-8 offspring per family, 4 families in total). The female in each of these crosses was Mu-inactive, and so new Mu insertions can be attributed to the male parent.

(**A**) *De novo* Mu insertions that occurred in the male parent were identified based on their absence in paternal endosperm and presence in both endosperm and leaf of the offspring. There was a median of 27.8 ± 6.7 *de novo* Mu insertions per generation.

(**B**) Mu insertions inherited by the offspring but present in the paternal endosperm could be attributed to prior generations (grandparents and beyond) and were not *de novo* in the male parent.

(**D,E**) *De novo* insertions from the male parent were most often transmitted uniquely to a single offspring (**E**) and only occasionally were shared by siblings (**D**). The low rate of *de novo* insertions shared by siblings was reported by Robertson (1980) and also explored in **Fig. 5**.

(**C,F**) To assess the contribution of technical errors to the genotyping calls, we looked for cases where an insertion occurred in two samples that cannot be explained by genetics. For instance, there were 914 cases where an insertion was present in the leaf of an offspring and the endosperm of its parent. This situation can be explained if the parent transmitted the insertion to the offspring, and so we would expect the matched offspring endosperm to also have the same insertion. This was seen in every case (**B**) and we never observed the biologically impossible situation where a matched endosperm did not have the insertion (**C**). Similarly, we observed 5,760 cases where the leaf and endosperm of one plant had an insertion not present in a sibling (**E**) but 0 cases where an insertion was present in the leaf of one plant and the endosperm from the sibling (**F**).

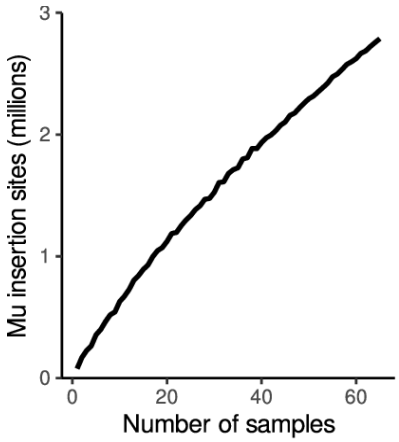

Figure S8. The number of Mu insertion sites has not reached saturation

Random subsets of samples were drawn and then the total number of genomic insertion sites was calculated. The total number of insertion sites has not reached saturation under the conditions of this study.

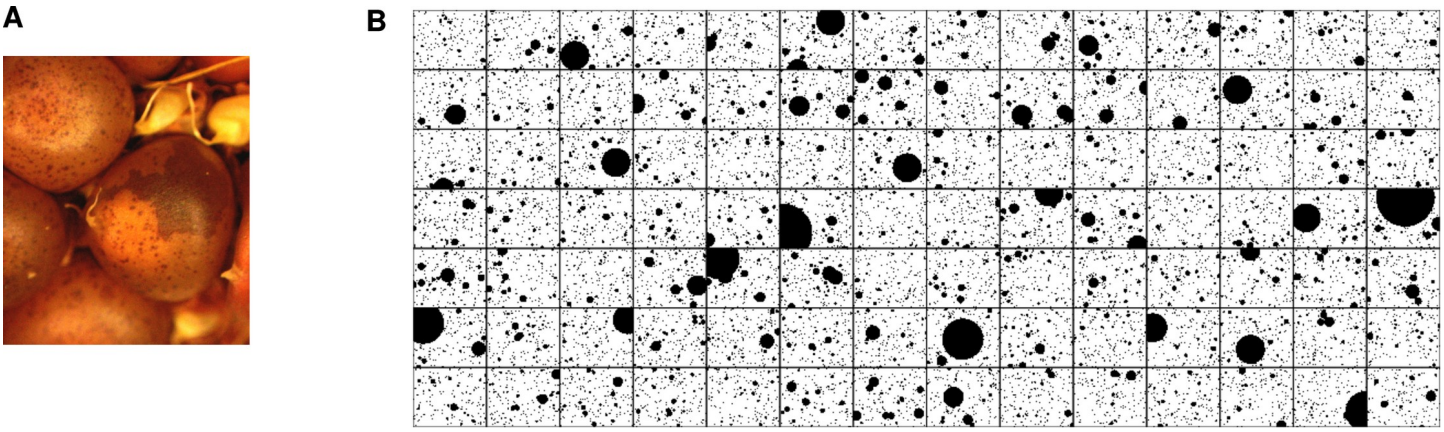

Figure S9. Mu insertions and excisions behave differently in endosperm

(**A**) Example sector sizes in Mu-active kernels with the bz1-Mum9 reporter. The kernel on the top left with many small spots is representative. The very large sector on the kernel in the middle is exceedingly rare; we observed 7 out of 1844 kernels with sectors making up at least 5% of the kernel area. Prior quantitative data on excision spot size found even fewer large sectors, with 0 sectors at a frequency under 2^8 (VAF ~ 10^{-3}) out of 2000 kernels (Levy and Walbot, 1990).

(**B**) A simulated ear with sectors drawn to represent Mu insertions. In this simulation, we assumed all divisions happen within a 2D plane, which may be approximately true for aleurone (the outer cell layer of endosperm where the visible pigment is produced). Spot size was defined by randomly drawing *de novo* insertions based on the measured allele frequencies in endosperm, requiring an average of 10% of the surface to be covered by sectors (10% surface coverage matches the rate measured for Mu excisions from reporter alleles; Levy and Walbot 1990). The frequency of large spots in this diagram is dramatically higher than what has been observed for endosperm excision sectors; in the simulation, 45% of 'seeds' had at least one sector covering more than 5% of the surface area, compared to 7/1844 (0.4%) for excisions (panel A).

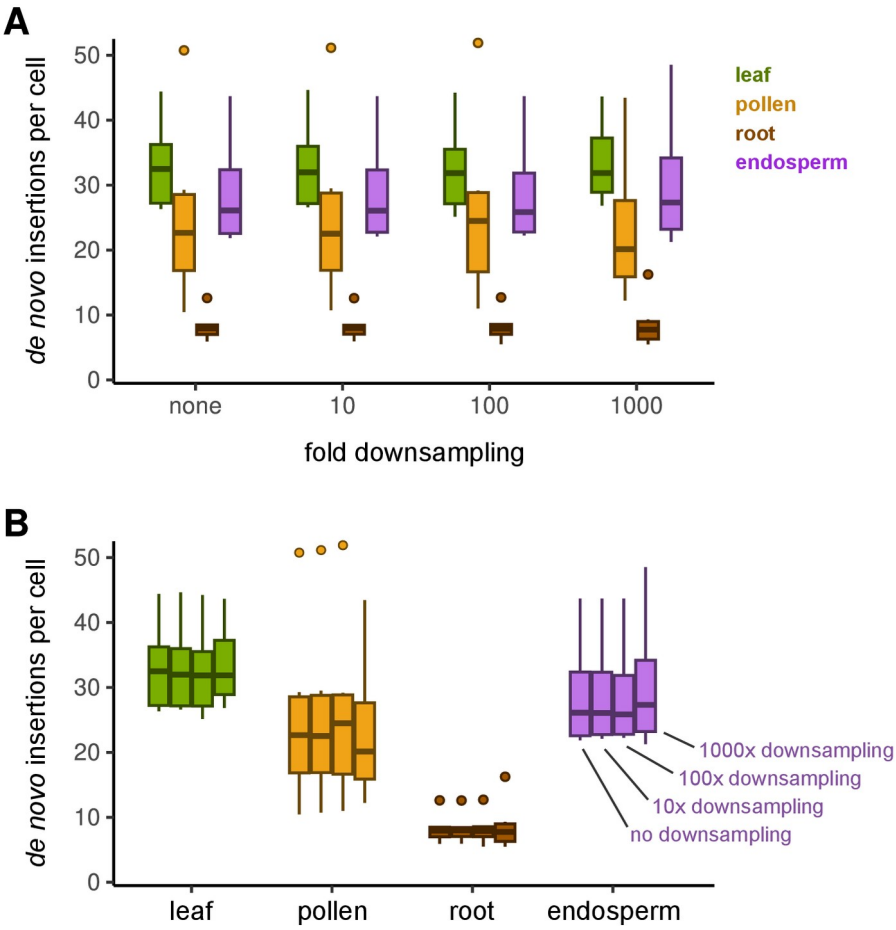

Figure S10. Estimates for the number of Mu insertions per cell are robust to sequencing depth, relevant to Figs. 4B.

Samples were downsampled by randomly removing molecules and then repeating the analysis in Fig. 4B. Panels **A** and **B** are identical, except that panel **B** groups all samples from the same tissue together. The estimated number of *de novo* insertions per cell was unaffected by downsampling as much as 1000-fold.

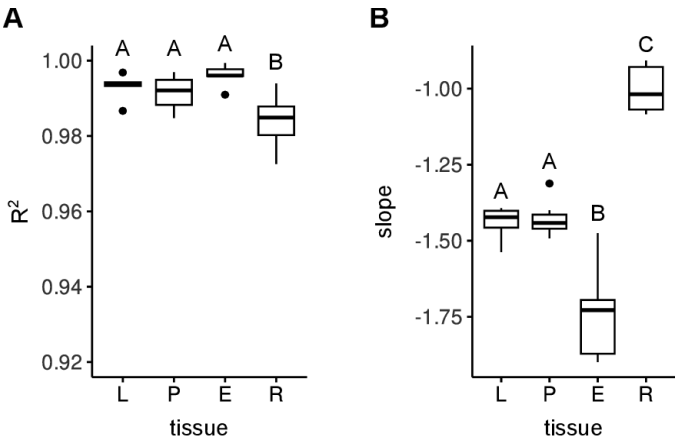

Figure S11. Power-law fits to the allele frequency spectra from various tissues

Best linear fit parameters to log-transformed data. Fitting was performed on individual samples and the results plotted as a boxplot separated by tissue. Letters indicate statistical significance: groups not sharing a letter have a significantly different mean ($p \leq 0.05$; Tukey's honest significant difference test). N = 6 samples for leaf and root; N = 9 samples for endosperm and pollen.

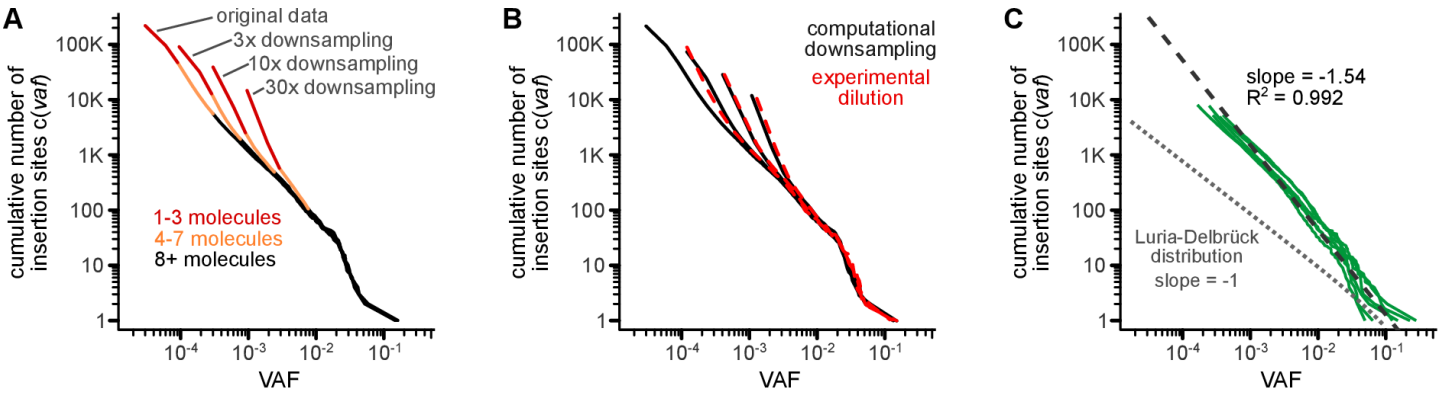

Figure S12. Evaluating the impact of counting error and sampling statistics on the measured allele frequency distributions.

(A) The data for a representative leaf sample was down-sampled 3, 10, or 30-fold by randomly drawing transposon-spanning molecules from the total. The down-sampled data were then normalized and the allele frequency distribution was plotted, as in the main text. The portion of the curve supported by 1-3 molecules are shown in red, 4-7 molecules in orange, and 8+ molecules in black. There is an upward bias in the estimated number of insertion sites for the portion of the curve supported by few molecules. The reason is that many insertions can be present below the detection limit, and while these are individually at low abundance and unlikely to be sequenced, they are collectively numerous (e.g. an insertion present at 1/10 the detection limit has a low chance of being sequenced, but there are so many such insertions that some of them will be sequenced once or twice). Beyond 8+ molecules, this effect is negligible (black curves).

(B) DNA from the leaf sample in (A) was experimentally diluted with Mu-inactive DNA, so that the contribution from the Mu-active leaf sample was 100%, 30%, 10%, or 3% of the total. MuSeq2 libraries were then prepared from these diluted samples and sequenced. The allele frequency distribution for the diluted samples (dotted red lines) closely match the distribution predicted by computationally downsampling (black lines). This shows that sampling statistics, such as approximated by random downsampling, capture the major source of bias in these curves.

(C) To assess whether our conclusions were robust to the upward bias in the number of insertions at low allele frequencies, we repeated the analysis in Fig. 3B after removing the data points supported by fewer than 8 molecules. The R^2 and slope of the best fit line were similar whether using the full data (1+ molecules; Fig. 3B) or truncated data (8+ molecules; this panel).

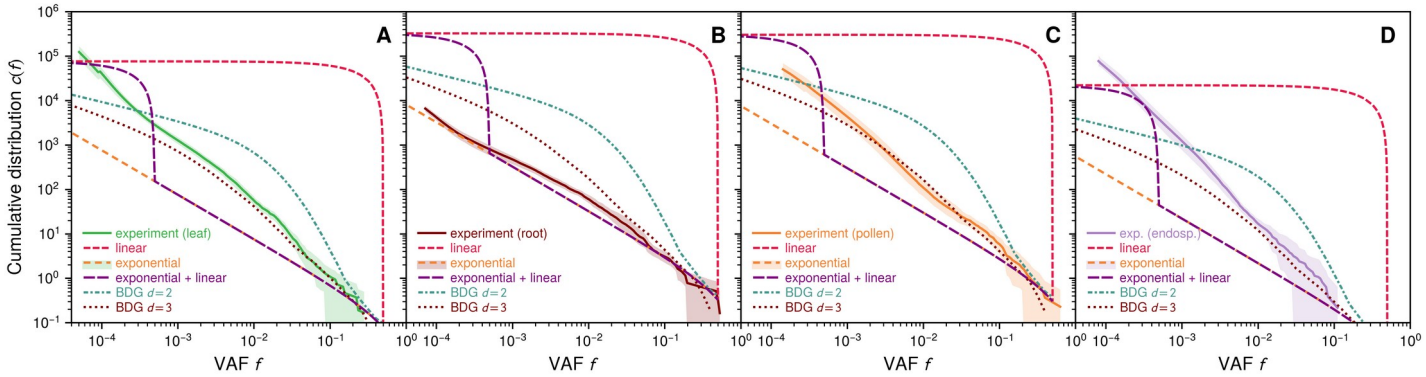

Figure S13. Fit of experimental leaf data to various models of mutation accumulation

Details of the theoretical models are described in the SI Text. All models assume no cell death, a constant mutation probability $\mu = 0.076$ over time (chosen to agree with the experimental curve at the largest frequency), and a final cell population size of $N_{cell} = 10^6$. Linear = linear growth, where after each cell division only one daughter cell is capable of further cell division. Exponential = exponential growth, where both daughters are capable of division (Luria-Delbrück model). Exponential + linear = 10 generations of exponential growth, with the remaining generations linear. BDG = boundary-driven growth simulations based on the Eden model[10], which were carried out on both 2D and 3D square lattices.

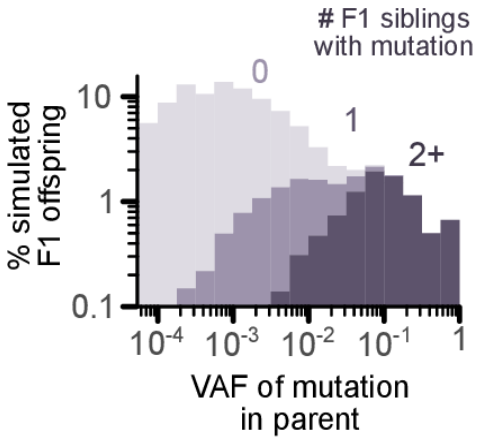

Figure S14. During simulations of Robertson (1980), F1 offspring that share the same mutation as their sibling are most often derived from pollen insertions at high allele frequencies.

For the simulations of Robertson (1980) in **Fig. 5B**, we recorded the VAF for any Mu insertion transmitted to the simulated F1 offspring. In this histogram, the bars are color coded based on whether the pollen mutation was inherited by 0, 1, or 2+ F1 siblings. Simulated F1 offspring were derived from Mu insertions at a wide range of allele frequencies, suggesting these occurred throughout development. For mutations present in larger clusters of F1 offspring (2+ siblings with the same mutation), the average VAF in the parent was 0.13; this corresponds 1 insertion in every 3.8 diploid pollen progenitors, roughly the number of meristematic cells in the seed that ultimately form the maize tassel (the male flower; ref: Poethig, Coe, and Johri, 1986). VAF, variant allele frequency.

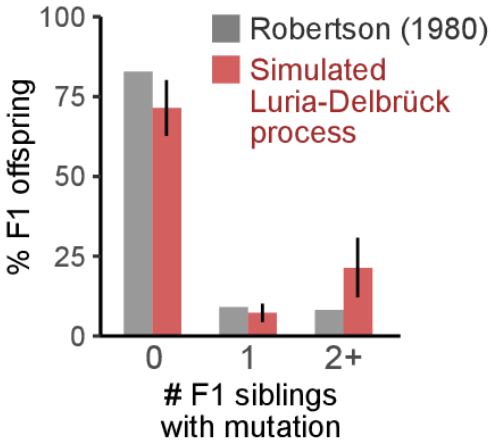

Figure S15. Comparison between Robertson (1980) to simulations with exponential cell division and a constant mutation rate (a Luria-Delbrück process).

The Robertson (1980) experiment was simulated assuming a constant mutation rate during 16 mitotic divisions, followed by meiosis and two pollen divisions. Sixteen mitotic divisions was selected because this is the minimum required to achieve the population size of a pollen sample in our study (~250,000 pollen grains). Simulated pollen grains were then drawn according the experimental design of Robertson (1980), and the frequency that an F1 offspring shared a given mutation with 0, 1, or 2+ siblings was quantified. The Luria-Delbrück process predicts fewer F1 offspring with unique mutations compared to Robertson (1980), but this difference is not statistically significant ($p = 0.198$; two-tailed test based on the result of 1000 simulations).

Table S1. Historical insertion sites

	Position		
Chromosome	***Left border***	***Right border***	**Present in W22 reference genome?**
1	20953342	20955151	Yes
1	30522219	30524307	Yes
1	104590546	104591797	Yes
1	272293892	272298698	Yes
2	47365715	47365707	No
2	73902765	73904517	Yes
2	142959280	142961386	Yes
2	146442692	146444522	Yes
2	150736892	150738762	Yes
2	166467881	166469772	Yes
2	N.D.	180947386	No
2	203236315	203241180	Yes
3	122342380	122342445	No
3	175620033	175621592	Yes
3	241129912	241133491	Yes
4	169355206	169356095	Yes
4	213483884	213485715	Yes
5	120060250	120061749	Yes
5	207534733	207534725	No
6	70998011	70998003	No
6	113180738	113187933	Yes
7	150868758	150868750	No
8	3470744	3470736	No
8	125105523	125105515	No
8	144608207	N.D.	Yes, but shifted 184 bp compared to reference location
9	24011353	N.D.	No
9	94045110	94046981	Yes
9	149235004	149236775	Yes
10	133146086	133147581	Yes

Table S2. Adapter sequences

Name	Sequence
m1	/5Phos/CGAACTGAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTCAGTTCGT
m2	/5Phos/CGTTCATGAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTCATGAACGT
m3	/5Phos/CGAACTCATG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATGAGTTCGT
m4	/5Phos/GCAACTAGTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGACTAGTTGCT
m5	/5Phos/GCAGTTCAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTGAACTGCT
m6	/5Phos/GCATCTACAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTGTAGATGCT
m7	/5Phos/GCATCTCATC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGATGAGATGCT
m8	/5Phos/CGTAGTTGAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTCAACTACGT
m9	/5Phos/GCAGAAGTAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTACTTCTGCT
m10	/5Phos/CGAGTACTTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAAGTACTCGT
m11	/5Phos/GCTGTTGATC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGATCAACAGCT
m12	/5Phos/GCTTCTCTAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTAGAGAAGCT
m13	/5Phos/GCACAAGTTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAACTTGTGCT
m14	/5Phos/CGTTCAACTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAGTTGAACGT
m15	/5Phos/CGTTCTACAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTGTAGAACGT
m16	/5Phos/CGTCTAGTAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTACTAGACGT
m17	/5Phos/CGTACACATC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGATGTGTACGT
m18	/5Phos/CGAAGTCTTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAAGACTTCGT
m19	/5Phos/GCACTAGATC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGATCTAGTGCT
m20	/5Phos/CGTACATGTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCACATGTACGT
m21	/5Phos/CGAAGAGAAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTTCTCTTCGT
m22	/5Phos/CGTTGTTGTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCACAACAACGT
m23	/5Phos/GCATGACAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTGTCATGCT
m24	/5Phos/GCTAGATGTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCACATCTAGCT
m25	/5Phos/CGATGATGAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTCATCATCGT
m26	/5Phos/GCTGTTCTTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAAGAACAGCT
m27	/5Phos/GCTCTTGTTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAACAAGAGCT
m28	/5Phos/CGAACTTCAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTGAAGTTCGT
m29	/5Phos/CGTTCTCTAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTAGAGAACGT
m30	/5Phos/GCTTCAACTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAGTTGAAGCT
m31	/5Phos/GCTAGTCTAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTAGACTAGCT
m32	/5Phos/GCATCATGAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTCATGATGCT
m33	/5Phos/CGACAAGAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTCTTGTCGT
m34	/5Phos/GCAGATGTTG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCAACATCTGCT
m35	/5Phos/GCATCAGTAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTACTGATGCT
m36	/5Phos/CGTAGACAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTGTCTACGT
m37	/5Phos/GCTAGAGATG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATCTCTAGCT
m38	/5Phos/CGTGTACTAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTAGTACACGT
m39	/5Phos/CGTTCTTCTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAGAAGAACGT
m40	/5Phos/GCTACTACAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTGTAGTAGCT
m41	/5Phos/CGTAGTGATG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATCACTACGT
m42	/5Phos/CGTTGAAGAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTCTTCAACGT
m43	/5Phos/GCTTCTGTAC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGTACAGAAGCT
m44	/5Phos/GCTGATCTTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGAAGATCAGCT
m45	/5Phos/CGTGTTGAAG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCTTCAACACGT
m46	/5Phos/GCACTTCATG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATGAAGTGCT
m47	/5Phos/CGTCTACATG/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNCATGTAGACGT
m48	/5Phos/CGTTCATGTC/ideoxyU/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNGACATGAACGT

Table S3. Primer sequences

Name	Sequence
TIR6	AGAGAAGCCAACGCCAWCGCCTCYATTTCGTC
UDz_i7	CAAGCAGAAGACGGCATACGAGATTTTTTTTTTGTGACTGGAGTTCAGACGTG
MuSeq2_NestedTIR	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCBCTCTTCKTCYATAATGGCAAT
P7	CAAGCAGAAGACGGCATACG
SP2	GTGACTGGAGTTCAGACGTG

Table S4. Blacklisted sites

Chromosome	Position		Present in W22 reference genome?
	Left border	*Right border*	
1	N.D.	1465152	No
1	18957410	18957402	No
1	N.D.	20955154	3 bp off from historical insertion site
1	64018346	64018338	No
1	122021265	N.D.	Yes
1	218464797	218464789	No
2	11397110	11397102	No
2	78235838	78240609	Yes
2	N.D.	180947385	1 bp off from historical insertion site
2	203236314	N.D.	1 bp off from historical insertion site
4	232384313	232384305	No
5	179239374	179239366	No
6	166313671	166313663	No
8	44857168	44858977	Yes
8	N.D.	160333759	No
9	24011352	N.D.	1 bp off from historical insertion site
10	24137902	24139637	Yes
10	126828981	126828973	No

Table S5. DNA input and % yield

Sample	Tissue type	Total purified DNA (ng)	DNA used for library prep (ng)	Genome equivalents	Expected number of paternal insertions	Molecules sequenced per paternal insertion	% insertions detected
E105	Endosperm	58	27	10795	3598	252	7.00%
E106	Endosperm	798	363	147672	49224	6192	12.58%
E107	Endosperm	1144	520	211832	70611	6313	8.94%
E42	Endosperm	2514	1143	465420	155140	7644	4.93%
E52	Endosperm	1788	813	330988	110329	22416	20.32%
E61	Endosperm	2827	514	209388	69796	7673	10.99%
E63	Endosperm	2585	470	191464	63821	15519	24.32%
E65	Endosperm	2283	415	169059	56353	15324	27.19%
E69	Endosperm	1293	588	239330	79777	13051	16.36%
E74	Endosperm	501	228	92779	30926	7579	24.51%
E75	Endosperm	946	430	175169	58390	3393	5.81%
E76	Endosperm	1254	456	185761	61920	5701	9.21%
E78	Endosperm	924	420	171095	57032	13795	24.19%
E79	Endosperm	1381	377	153375	51125	8306	16.25%
E83	Endosperm	2090	380	154801	51600	10215	19.80%
E85	Endosperm	908	413	168040	56013	17837	31.84%
E99	Endosperm	550	250	101843	33948	6790	20.00%
L74	Leaf	1848	336	136876	68438	5864	8.57%
L74 (tech rep)	Leaf	1848	336	136876	68438	5538	8.09%
L75	Leaf	1359	371	150931	75465	12065	15.99%
L75 (tech rep)	Leaf	1359	371	150931	75465	11884	15.75%
L76	Leaf	1832	333	135654	67827	5296	7.81%
L78	Leaf	1980	360	146653	73327	14302	19.50%
L79	Leaf	1892	344	140135	70068	10062	14.36%
L83	Leaf	1348	368	149708	74854	16292	21.76%
P105	Pollen	776	353	143598	71799	5652	7.87%
P106	Pollen	1518	552	224868	112434	5271	4.69%
P107	Pollen	188	85	34728	17364	1757	10.12%
P42	Pollen	485	220	89723	44862	3707	8.26%
P52	Pollen	372	169	68846	34423	3511	10.20%
P69	Pollen	300	137	55606	27803	1422	5.11%
P85	Pollen	513	233	94917	47459	5744	12.10%
P99	Pollen	880	320	130358	65179	5348	8.21%
R61	Root	1199	436	177613	88807	24723	27.84%
R63	Root	3476	948	386187	193093	32301	16.73%
R65	Root	405	184	74956	37478	7376	19.68%
R69	Root	1012	460	187390	93695	17379	18.55%
R74	Root	1700	464	188816	94408	5677	6.01%
R75	Root	1265	345	140543	70271	7166	10.20%