High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements

Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Sam Hokin, Erica Unger-Wallace, Rex A. Cole, Antony M. Chettoor, Duo Jiang, Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler*

1 Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, United States of America, 2 Donald Danforth Plant Science Center, St. Louis, Missouri, United States of America, 3 Department of Plant Biology, Carnegie Institution for Science, Stanford, California, United States of America, 4 Department of Genetics Development and Cell Biology, Iowa State University, Ames, Iowa, United States of America, 5 Department of Statistics, Oregon State University, Corvallis, Oregon, United States of America, 6 Bioinformatics and Computational Biology, Iowa State University, Ames, Iowa, United States of America, 7 Interdepartmental Genetics, Iowa State University, Ames, Iowa, United States of America, 8 Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon, United States of America

* fowlerj@science.oregonstate.edu

Abstract

In flowering plants, gene expression in the haploid male gametophyte (pollen) is essential for sperm delivery and double fertilization. Pollen also undergoes dynamic epigenetic regulation of expression from transposable elements (TEs), but how this process interacts with gene expression is not clearly understood. To explore relationships among these processes, we quantified transcript levels in four male reproductive stages of maize (tassel primordia, microspores, mature pollen, and sperm cells) via RNA-seq. We found that, in contrast with vegetative cell-limited TE expression in Arabidopsis pollen, TE transcripts in maize accumulate as early as the microspore stage and are also present in sperm cells. Intriguingly, coordinate expression was observed between highly expressed protein-coding genes and their neighboring TEs, specifically in mature pollen and sperm cells. To investigate a potential relationship between elevated gene transcript level and pollen function, we measured the fitness cost (male-specific transmission defect) of GFP-tagged coding sequence insertion mutations in over 50 genes identified as highly expressed in the pollen vegetative cell, sperm cell, or seedling (as a sporophytic control). Insertions in seedling genes or sperm cell genes (with one exception) exhibited no difference from the expected 1:1 transmission ratio. In contrast, insertions in over 20% of vegetative cell genes were associated with significant reductions in fitness, showing a positive correlation of transcript level with non-Mendelian segregation when mutant. Insertions in maize gamete expressed2 (Zm gex2), the sole sperm cell gene with measured contributions to fitness, also triggered seed defects when crossed as a male, indicating a conserved role in double fertilization, given the similar phenotype previously demonstrated for the Arabidopsis ortholog GEX2.
Overall, our study demonstrates a developmentally programmed and coordinated transcriptional activation of TEs and genes in pollen, and further identifies maize pollen as a model in which transcriptomic data have predictive value for quantitative phenotypes.

Author summary

In flowering plants, pollen is essential for delivering sperm cells to the egg and central cell for double fertilization, initiating the process of seed development. In plants with abundant pollen like maize, sperm cell delivery can be highly competitive. In an added layer of complexity, growing evidence indicates expression of transposable elements (TEs) is more dynamic in pollen than in other plant tissues. How these elements impact pollen function and gene regulation is not well understood. We used transcriptional profiling to generate a framework for detailed analysis of TE expression, as well as for quantitative assessment of gene function during maize pollen development. TEs are expressed early and persist, many showing coordinated activation with highly-expressed neighboring genes in the pollen vegetative cell and sperm cells. Measuring fitness costs for a set of over 50 mutations indicates a correlation between elevated transcript level and gene function in the vegetative cell. We also establish a role in fertilization for the maize gamete expressed2 (Zm gex2) gene, identified based on its specific expression in sperm cells. These results highlight maize pollen as a powerful model for investigating the developmental interplay of TEs and genes, as well as for measuring fitness contributions of specific genes.

Introduction

Sexual reproduction enables the segregation and recombination of genetic material, which increases genetic diversity in populations and contributes to the vast diversity of eukaryotes. In flowering plants, sexual reproduction requires the development of reduced, haploid gametophytes from sporophytic, diploid parents. The mature female gametophyte, the embryo sac, includes the binucleate central cell and the egg cell (reviewed in [1,2]), each of which is fertilized by a sperm cell to generate the triploid endosperm and diploid embryo, respectively. The mature male gametophyte, pollen, consists of a vegetative cell harboring two sperm cells (reviewed in [3,4]). In maize, male gametophytes arise from microspore mother cells in the tassel primordium. The transition from diploid sporephore to haploid gametophyte occurs when these cells undergo meiosis, each resulting in four haploid microspores. Each microspore then undergoes two rounds of mitosis to produce the pollen grain, first generating the large vegetative cell and a smaller generative cell via asymmetric division, and then producing the two sperm from the generative cell. After the arrival of the pollen grain on the floral stigma, the vegetative cell transports the two sperm cells to the female gametophyte via pollen tube growth (reviewed in [5,6]). Accurate navigation of the pollen tube as it grows down the style is dependent on the architecture of the style’s transmitting tract [7] and additional signaling and recognition mechanisms that are poorly understood [8]. The final stages of pollen tube growth depend on a complex interplay of signals to guide the pollen tube to the micropyle of the ovule for sperm delivery to the embryo sac (reviewed in [9]).

In maize, a pollen tube must grow up to 30 cm through the silk to reach the female gametophyte, often competing with multiple pollen tubes to eventually enter the embryo sac and release its sperm cells for fertilization (reviewed in [2,5]). Across the angiosperms, this
competitive context for pollen tube development differs, depending on the pollen population as well as sporophytic characters (reviewed in [10]). In a highly competitive environment, successful fertilization is likely enhanced by pollen tubes functioning at full capacity [11–13], as generally only the first tube to reach the micropyle is permitted to enter the female gametophyte. The mechanisms preventing entry of multiple pollen tubes, known as the polytubey block, are not well-understood, but presumably act to reduce polyspermy, which typically leads to sterile offspring [6]. In maize, mutations in the genes MATRILINEAL/NLD/ZmPLA1 and ZmDMP have been linked to pollen-induced production of haploid embryos and other seed defects, which are likely associated with aberrant events at fertilization [14–17] or soon after [18]. Thus, many mechanisms associated with both pollen tube growth and fertilization remain enigmatic.

Given their specialized biological functions and well-defined developmental stages, gametophytes are prime targets for transcriptome analysis. Initial studies of plant gametophytic transcriptomes in Arabidopsis pollen [19,20] and embryo sacs [21,22] described a limited and specialized set of transcripts and identified numerous candidate genes for gametophytic function. In maize, the first RNA-seq study of male and female gametophyte transcriptomes (mature pollen and embryo sacs) similarly identified subsets of developmentally specific genes, with pollen showing the most specialized transcriptome relative to other tissues assessed [23]. More recently, RNA-seq has been carried out on additional stages of maize reproductive development, including pre-meiotic and meiotic anther cells [24–26], as well as sperm cells, egg cells, and early stages of zygotic development [27].

Gametophytic tissues are known to show dynamic expression of transposable elements (TEs). In Arabidopsis, global TE expression is derepressed at the late stages of pollen development, occurring in the pollen vegetative nucleus only after pollen mitosis II [28]. The pollen vegetative nucleus undergoes a programmed loss of heterochromatin, resulting in TE activation, TE transposition and subsequent increased RNA-directed DNA methylation [28–32]. A variety of functions have been ascribed to this male gametophytic "developmental relaxation of TE silencing" (DRTS) event [33], including the generation of TE small interfering RNAs that are mobilized to the sperm cells [34], and control of imprinted gene expression after fertilization [35]. However, the dynamics of TE expression during gametophytic development in a transposable element-rich species such as maize have not been investigated.

To provide a more full description of transcriptome dynamics across maize male reproductive development, including TE transcriptional activity, we generated RNA-seq datasets from tassel primordia, microspores, mature pollen, and isolated sperm cells. Using these data, we describe differential expression patterns of genes and TEs across these stages, uncovering a coordinated regulation of TEs and their neighboring genes in pollen grains. Then, within a framework provided by the transcriptome data, we conducted a functional validation of highly expressed genes by testing over fifty insertional mutations for male-specific fitness effects. Finally, the same transcriptome data guided the discovery of mutant alleles in the sperm cell-enriched gex2, which induces seed development defects when present in the pollen parent, implying a role in fertilization.

Results

Experimental design and gene expression during maize male reproductive development

RNA-seq was performed on four tissues representing integral stages in maize male gametophyte development: immature tassel primordia (TP), isolated unicellular microspores (MS), mature pollen (MP), and isolated sperm cells (SC) (Fig 1A). Techniques were developed to
efficiently isolate RNA from TP, MS, and SC (see Methods). RNA was extracted from the inbred maize line B73, with four biological replicates for each tissue. In addition, a single RNA replicate was isolated for the bicellular stage of pollen development (MS-B). Libraries were sequenced using Illumina sequencing (100 bp paired-end reads) and mapped to the B73 AGPv4 reference genome [36]. Principal Component Analysis (PCA) showed samples from each of the four sequenced tissues clustered with other replicates from the same tissue. TP and MS were clearly separated in principal component space, whereas SC and MP samples displayed less separation from each other. (C) High expression levels are associated with developmental specificity: approximately 2/3 of the genes associated with the highest FPKM values in each of the four sample types are highly expressed in only that sample type.
Differential gene expression was defined in two ways: in the first, gene expression in later developmental stages was compared to the premeiotic, diploid tassel primordia (TP vs MS, TP vs MP, and TP vs SC); in the second, gene expression was compared between all adjacent developmental stages (TP vs MS, MS vs MP, MS vs SC, MP vs SC) (S2 and S11 Tables). Enriched GO terms highlighted the differences in gene expression among developmental stages and suggested consistency with the established functions of each tissue [19,23,27]. GO terms in MS were consistent with a post-meiotic tissue still at an early stage of development, with terms related to protein synthesis and transport, morphogenesis, and reproduction showing enrichment. MP showed more specific enriched GO terms, including those related to pollen tube growth, signaling, and actin filament-based movement. SC shared many GO terms with MP when compared to MS, but was uniquely enriched for GO terms related to epigenetic regulation of gene expression, such as histone H3K9 demethylation and gene silencing by RNA, potential mechanisms involved in differential regulation of TEs.

Comparison of the most highly expressed genes from all four sample types showed that, generally, such transcripts were highly enriched at a single developmental stage (Fig 1C). Notably, no single gene was highly expressed in all four tissues, and fewer than twenty were highly expressed in three out of the four stages assessed. These data suggest a simple hypothesis in which the expression level for protein-coding genes reflects, at some level, functional importance, i.e., a high expression level at a specific developmental stage implies an increased contribution by the gene’s encoded function at that particular stage. Alternatively, or in addition, high expression could be reflective of regulatory mechanisms specific to each stage, each primarily influencing specific subsets of genes. We were interested to explore the possibility of some regulatory linkage between those genes and TEs highly expressed in the male gametophyte.

A subset of transposable elements in the maize genome show developmentally dynamic expression

To obtain a baseline view of TE expression throughout maize development, our RNA-seq data for maize male reproductive development (samples with asterisks, S1 Fig) was combined with publicly available datasets from nine-day old above-ground seedlings, juvenile leaves, ovules, another set of independently isolated sperm cells, and three independent studies of pollen RNA-seq [23,27,37,38] & SRP067853). The complete list of samples, their sequencing statistics, references and data availability can be found in S3 Table. All of the raw data were remapped using the same parameters (see Methods). Principal component analysis demonstrates that replicates of the same tissue and growth state typically group together (S1 Fig).

We aimed to identify the set of dynamically expressed TEs within the tissues sampled, and thus calculated expression levels for each individual TE in the genome located more than 2 kb away from annotated non-TE genes. Our rationale was to avoid false positive signals of TE expression due to a TE residing within a gene, and to minimize the influence of read-through transcription from a nearby gene, which could not be distinguished from TE-initiated transcription. Because we concentrated on individual elements, and not TE families, the majority of annotated TEs were not assessed in this analysis (55%; Fig 2A ‘not covered’), either because no expression was detected in any dataset, or because their sequence lacks the polymorphisms necessary for mapping reads to a specific TE. To relate TE expression comparatively across development, we used seedling tissue as a baseline against which other tissues were measured. Seedling was chosen for several reasons: it is not a reproductive tissue, it has low to average levels of TE expression, and a large number of TEs show no evidence of expression in this tissue (S2 Fig).
Fig 2. Characterization of developmentally dynamic transcription from transposable elements (TEs). (A) Distribution of different categories of TEs based on their expression. Number of TEs are in parentheses. (B) Length of TEs in the different TE categories from part A. The violin plots around the box show the kernel probability density of the data. The box represents lower and upper quartile, the line is the median, and the whiskers represent 10–90% range. Red asterisk denotes the mean. (C)
Apart from the 18.3% of annotated TEs that are near genes and analyzed separately (see below), we calculated the number of TEs with statistically significant expression differences in each tissue compared to the seedling reference. This identified the subset of TEs that are developmentally dynamic, meaning that they show differential expression in at least one tissue in our dataset compared to the seedling reference. Only 4.4% of all maize annotated TEs are developmentally dynamic, whereas 22.2% of TEs have detectable expression but do not change in our dataset and therefore are developmentally 'static' (Fig 2A). Each TE category was interrogated for feature overrepresentation. Both dynamic and static TEs are longer than the genome average, and longer than the sets of TEs 'not covered' or 'near genes' (Fig 2B). To determine if one long family of TEs was contributing this difference, we performed this analysis for each TE superfamily and found that for dynamic TEs, this observation is not specific to one TE type (S3 Fig). The finding that expressed TEs as a group are longer correlates with Arabidopsis data where longer TE transcripts are overrepresented and differentially regulated when epigenetic repression is lost [39].

Expressed TEs show an under-representation for DNA transposon and SINE families, which are mainly within the ‘near genes’ set (Fig 2C). In contrast, the 'LTR unknown' TE annotation is over-represented in the dynamic TE set (Fig 2C). Since some LTR retrotransposons are enriched in the pericentromere [40], we tested if the dynamic TE set is enriched in the pericentromere compared to the genome average, but did not detect any correlation (Fig 2D). Therefore, we conclude that expressed TEs are generally longer elements, and the subset of developmentally dynamic TEs are enriched for uncharacterized LTR retrotransposons located throughout the genome.

Transposable element transcript levels are up-regulated in the post-meiotic male reproductive lineage

From the developmentally dynamic TE set, we calculated the number of differentially expressed TEs in each tissue/stage compared to the seedling reference. In some tissues, such as tassel primordia and ovules, we observed a similar number of TEs up-regulated and down-regulated (Fig 3A), demonstrating that while there are shifts in which TEs are expressed, a genome-scale change in TE expression does not occur. In other tissues, such as juvenile leaves, there is a skew towards increased TE expression. The largest TE up-regulation occurs in the tissues of the male reproductive lineage, including unicellular and bicellular microspores, mature pollen and isolated sperm cells (Fig 3A). Our data confirm the recent finding that the tissue with the largest number of TE families activated is mature pollen [41]. The number of up-regulated TEs compared to down-regulated TEs in these tissues suggest that there is a genome-wide activation of TE expression, similar to the DRTS event that occurs in Arabidopsis pollen [28,33]. One important distinction is that TE expression is present in maize sperm cells (Fig 3A), whereas it is not detected in Arabidopsis sperm cells [28]. To verify this finding, we compared our sperm cell RNA-seq data to an independent maize sperm cell dataset [27]. We found that TEs are also significantly expressed in this independent dataset, and 70% of those expressed TEs are also detected in our dataset (p < 0.001) (Fig 3B). This shared set of 810 sperm cell-expressed TEs (38% of those detected in our dataset), supports the conclusion that significant expression of TEs occurs in maize sperm cells. Of the sperm cell-expressed TEs, 36% were not observable in total pollen, but rather required the isolation and enrichment of
Fig 3. High TE expression in the maize male gametophyte lineage. (A) Number of differentially expressed TEs in seven tissues compared to seedlings. The inset volcano plot shows for mature pollen how differentially expressed TEs were identified. Green and red numbers within the volcano plot indicate how many TEs were statistically up- or down-regulated, respectively. (B) Number of up-regulated TEs in mature pollen compared to isolated sperm cells from this study and a previously published distinct isolation and sequencing of sperm cell mRNA. (C) Starting with TEs differentially up-regulated in unicellular microspores (boxed, far left volcano plot), we determined how many of these same TEs are expressed at other developmental time points. (D) Raw distribution of expressed TE family annotations. 'Male'
sperm cells for detection (Fig 3B). Overall, we detect 157 TEs expressed in both sperm cell datasets that are not expressed throughout development, but specifically in the sperm cells (sperm-cell exclusive).

A second notable difference between maize and Arabidopsis is the activation of TE expression early in the male gametophytic phase of maize. A genome-wide increase in TE transcript levels is detected at the earliest post-meiotic stage tested, the microspore, in contrast to low TE expression in the sporophytic tassel primordia (Fig 3A). Arabidopsis TE expression occurs only late in pollen development, after pollen mitosis I when the vegetative cell is generated [28]. To determine if TEs were indeed activated early in maize male reproductive development, we asked if the same TEs that we identified as expressed in the unicellular microspore remain active throughout the male reproductive lineage. We used the set of differentially expressed up-regulated TEs in unicellular microspores (3,335) and found that 62% are still expressed in bicellular microspores and 54% in mature pollen (Fig 3C), demonstrating that once TEs are activated early in development, expression and/or steady-state mRNA frequently remains through pollen maturation. Only some of these male-lineage expressed TEs continue to be expressed in sperm cells (32%), raising the possibility that many TEs with active expression in the early gametophytic stages are under negative/repressive regulation in the gametes. This large-scale developmental activation is potentially limited to the male lineage, as ovules express relatively few TEs (Fig 3A) and only 14% of the male lineage-expressed elements (Fig 3C). Together, our data demonstrate conserved activation of TE expression in the male gametophytes of maize and Arabidopsis, with key differences such as the developmental timing and localization of TE expression in the gamete cells.

We determined what types of TEs activate in the male reproductive lineage and sperm cells and compared these to the whole-genome distribution of TEs analyzed. Overall, both male lineage-expressed TEs and sperm cell-expressed TEs reflect the genome-wide TE distribution (Fig 3D). This suggests that TE family type does not have a determining role in the developmental regulation of TE expression. One notable exception is the enrichment of Mutator family TE expression in sperm cells (Fig 3D). When normalized for genome-wide TE distribution, Mutator element expression is highly enriched across the male lineage, including in sperm cells (Fig 3E). The expression of some Mutator TEs in sperm cells is both high confidence (present in both sperm cell datasets) and specific to only that tissue (high confidence sperm cell specific, Fig 3E). LINE L1 elements are also expressed throughout the male lineage and sperm cells, but their expression is general and not specific to these cell types (Fig 3E). Our data demonstrate that there is a general (TE family-independent) activation of TE expression in the male reproductive lineage, with one observable bias towards Mutator family expression in both the male lineage and sperm cells.

**Mature pollen and sperm cells display coexpression of highly expressed genes and their neighboring TEs**

To determine if TEs have an effect on neighboring gene expression, or vice versa, we next analyzed the set of 36,943 assayable TEs within 2kb of genes (Fig 2A). We calculated the absolute expression level of each genic isoform and categorized them into 100 bins of expression levels for each developmental stage (Fig 4). We found no significant relationship between highly expressed genes and the number of up- or down-regulated TEs in tassel primordia or...
microspores (grey bars, top row, Fig 4). In contrast, in both mature pollen and isolated sperm cells there is a statistically significant (p < 1E-6) positive association between highly expressed genes and the number of up-regulated TEs within 2kb of those genes (grey bars, bottom row, Fig 4). Similarly, there is a negative correlation (p < 1E-6) between high gene expression and the number of down-regulated TEs in the same samples (grey bars, bottom row, Fig 4). This relationship is not due to the fact that pollen or sperm cell-expressed genes are more likely to be located nearby a TE (S4A Fig), nor due to sample contamination between these two datasets (S4B Fig). To determine if our analysis is biased by the presence of multiple TEs close to just a few highly expressed genes, we also counted the number of genes with at least one differentially expressed TE within 2kb (black dots, Fig 4 and S4 Fig). We found that a similar number of genes were adjacent to differentially expressed TEs (compared grey bars to black dots, Fig 4 and S4 Fig), demonstrating that a small number of genes was not biasing our dataset. We conclude that specifically in the mature male gametophyte the most highly expressed genes tend to be near actively expressing TEs. However, it remains unclear whether gene expression is

Fig 4. Co-regulation of TE and gene expression in the male gametophyte. For each tissue type, the top 20,000 most highly expressed genes are distributed along the X-axis in bins of 200, with the most highly expressed bin on the far left. For each bin the number of up- and down-regulated TEs near (<2kb) that bin’s genes is then summed on the Y-axis (shown in grey bar). For each bin, the number of genes with at least 1 up- or down-regulated TE within 2kb is displayed as black dots. In unicellular microspores (top right) there is little correlation, whereas in mature pollen and sperm cells (bottom panels) the most highly expressed genes are near primarily up-regulated TEs. To check if the perceived correlations are statistically significant, we performed Kendall’s Tau rank correlation tests and found significant correlations (p < 1E-6) only for mature pollen and sperm cells for both number of TEs and number of genes. For mature pollen, gene expression has a positive correlation with up-regulated TEs (#TEs: \( \tau = 0.65 \), #genes: \( \tau = 0.64 \)) and a negative correlation with down-regulated TEs (#TEs: \( \tau = -0.55 \), #genes: \( \tau = -0.55 \)). Similarly for sperm cells, gene expression has a positive correlation with up-regulated TEs (#TEs: \( \tau = 0.52 \), #genes: \( \tau = 0.52 \)) and a negative correlation with down-regulated TEs (#TEs: \( \tau = -0.42 \), #genes: \( \tau = -0.44 \)). Only TE (not genes) rank correlation coefficients (\( \tau \)) are displayed in the figure. The bin location of gex2 (see Fig 7) is annotated in the sperm cell data.

https://doi.org/10.1371/journal.pgen.1008462.g004
influencing TE expression, or TE expression is affecting gene regulation, or alternatively, some
global regulatory mechanism is influencing both.

**Large-scale insertional mutagenesis supports a relationship between transcript level and fitness contribution for vegetative cell-expressed genes**

Using the quantitative framework provided by our transcriptome dataset, we next tested the simple hypothesis that highly expressed genes contribute to male gametophytic function—i.e., to reproductive success (pollen fitness). The functional validation approach we used relied on a large, sequence-indexed collection of green fluorescent protein (GFP)-marked transposable element (Ds-GFP) insertion mutants [42], enabling assessment of the effects of mutations in select genes (Fig 5). We focused on expression data from the MP and SC stages, as these have

---

**Fig 5. Large-scale tracking of seed marker transmission frequencies was accomplished by generating ear projections with a custom built rotational scanner.**

(A) When crossed either through the male or the female, Ds-GFP mutant allele tsgR107C12 (in gene Zm00001d012382), marked by green fluorescent seeds, shows 1:1 Mendelian inheritance (50% transmission of the GFP seed marker). Images captured in blue light with an orange filter. (B) Mutant alleles in other genes, such as tsgR102H01 (Zm00001d037695), showed non-Mendelian segregation when crossed through the male (37.5% GFP transmission). Segregation through the female remained Mendelian, indicating a male-specific transmission defect. (C) For some mutant alleles (~10% of lines in this study), the anthocyanin transgene C1 was tightly linked to the insertion mutant. In these cases, seeds carrying a mutant allele of a gene of interest could be tracked by their purple color. Here, insertion tsgR96C12 (Zm00001d015901) shows a strong male-specific transmission defect (24.8% C1 transmission through the male). Images captured in full spectrum visible light.

https://doi.org/10.1371/journal.pgen.1008462.g005
distinctive cell fates and roles in reproduction: the vegetative cell generates the pollen tube for competitive delivery of gametes, and the sperm cells accomplish double fertilization. Expression data from seedlings [23] was used to design a sporophytic control. Highly expressed genes, operationally defined as in the top 20% for a tissue by FPKM, were grouped into three mutually exclusive classes: Seedling, Sperm Cell, and Vegetative Cell. The seedling group also excluded any gene highly expressed in either MP or SC. Due to the significant overlap among genes highly expressed in both MP and SC, we compared expression values to assign each of these genes to a single class. Vegetative Cell genes were not only highly expressed in MP, but were also associated with an FPKM greater in MP than in SC, and vice versa for Sperm Cell genes (S5 Table). All genes in these classes were then cross-referenced with Ds-GFP insertion locations to identify potential mutant alleles for study, restricting the search to insertions in coding sequence (CDS), as these were rationalized as most likely to generate loss-of-function effects. Finally, to insure our results were as generalizable as possible, each class list was randomized to identify the specific subset of Ds-GFP lines for study. Insertion locations were verified by PCR for 64 of 83 alleles obtained (S6 Table) (see Methods), of which 56, representing mutations in 52 genes, generated sufficient transmission data to include in our final analysis.

Mendelian inheritance predicts 50% transmission of mutant and wild-type alleles when a heterozygous mutant is outcrossed to a wild-type plant. However, a mutation that alters the function of a gene expressed during the haploid gametophytic phase can result in a reduced transmission rate if that gene contributes to the fitness of the male gametophyte—i.e., to its ability to succeed in the highly competitive process of pollen tube growth, given that 50% of the pollen population will be wild-type for the same gene. Thus, reduced transmission of a mutant through the male (a male transmission defect) provides not only evidence for gene function in the gametophyte, but also a measure of the mutated gene’s contribution to fitness. Transmission rates through the female serve as a control, as 50% transmission through the female would confirm both a single Ds-GFP insertion in the genome and male-specificity for any defect identified. To measure the fitness cost of each Ds-GFP insertion, heterozygous mutant plants were reciprocally outcrossed with a heavy pollen load to a wild-type plant, maximizing pollen competition within each silk. Transmission rates were then quantified by assessing the ratio of the non-mutant to mutant progeny using a novel scanning system and image analysis pipeline (Fig 5) (see Methods) [43]. Mutant alleles were tracked using linked endosperm markers: either the GFP encoded by the inserted transposable element (Fig 5A and 5B), or, in ~10% of the lines, a tightly linked C1+ anthocyanin transgene (present due to the initial Ds-GFP generation protocol) (Fig 5C, S7 Table).

Transmission rates for all groups were tested through quasi-likelihood tests on generalized linear models with a logit link function for binomial counts (see Methods, S8 Table). When crossed through the female, no genes showed significant differences from Mendelian inheritance (Fig 6A). When crossed through the male, no genes with insertion alleles in the Seedling category \( n = 10 \) showed evidence of abnormal transmission rates (Fig 6B). Most Sperm Cell genes \( n = 10, 90\% \) showed no statistically significant transmission defects, with one notable exception (two independent alleles of the gex2 gene, described in detail below) (Fig 6C). However, among Vegetative Cell genes tested \( n = 32 \), a larger proportion of insertion alleles \( 7 \) out of \( 32 \) or 21.9\% showed significant male transmission defects (quasi-likelihood test, adjusted p-value threshold \(< 0.05\) ) (Fig 6D). The proportions of genes with transmission defects in the three classes were not significantly different by Fisher’s exact test (Seedling vs Sperm Cell \( p\text{-value} = 0.500\), Seedling vs Vegetative Cell \( p\text{-value} = 0.125\), Vegetative Cell vs Sperm Cell \( p\text{-value} = 0.374\), likely due to the small number of mutations assessed in the Seedling and Sperm Cell classes. For the insertion alleles tested, a summary description of genes showing non-Mendelian inheritance can be found in Table 1, whereas a description of those showing Mendelian inheritance can be found in Table 2.
The majority of transmission defects in the Vegetative Cell class genes (six of the seven with significant effects) were mild, at approximately 45% transmission, with only one reducing transmission by a moderate amount, to ~30%. Notably, six of the genes associated with significant defects were measured at a log$_2$(FPKM) $> 8$ (i.e., in the top 5% of Vegetative Cell genes by FPKM). Given that twelve genes above this threshold were tested, these most highly expressed Vegetative Cell genes were significantly more likely to be associated with non-Mendelian transmission (6 out of 12) than the group of Vegetative Cell genes below this expression threshold (1 out of 20) (Fisher’s exact test, p-value = 0.00572). Consistent with this observation, an increase in log$_2$(FPKM) was associated with both reduced transmission rate and an increase in -log$_{10}$(p-value) (linear regression, p-value = 0.0120, 0.0255, respectively). Thus, our
data indicate that transcript level in the Vegetative Cell does provide some limited predictive power for identifying gene-specific contributions to male gametophytic fitness (adjusted $R^2 = 0.151, 0.116$, respectively). Vegetative Cell genes associated with non-Mendelian inheritance had a range of predicted cellular functions, including cell wall modification, cell signaling, protein folding, vesicle trafficking, and actin binding (Table 1).

To ensure the experimental design was robust, we examined two potential confounding variables: the presence of the \( \text{wx1-m7}::\text{Ac} \) allele in a subset of lines tested and the potential for epigenetic silencing of GFP transgenes (see S1 Methods). We found no evidence that the presence of \( \text{wx1-m7}::\text{Ac} \) significantly impacted the overall conclusions drawn from the dataset, nor evidence of epigenetic silencing of GFP transgenes.

**Insertional mutants in the sperm cell-expressed \( Zm \text{gex2} \) cause paternally triggered aberrant seed development**

The male-specific transmission defect for the sole affected gene in the Sperm Cell class, \( Zm00001d005781 \) (GRMZM2G036832), was notably more severe than the average defect across all \( Ds\text{-GFP} \) mutants identified with decreased transmission through the male (Table 1). This gene is hereafter referred to as \( Zm \text{gex2} \) or \( \text{gex2} \), for reasons detailed below. The two independent alleles assessed, \( \text{gex2-tdsgR82A03} \) and \( \text{gex2-tdsgR84A12} \), were associated with transmission rates of 33.4% and 23.1%, respectively. Sequencing confirmed that these \( Ds\text{-GFP} \) elements were inserted into their predicted CDS locations (Fig 7A). In addition to the transmission defect, both alleles, when crossed through the male, conditioned unusual phenotypes: underdeveloped or aborted seeds, as well as ovules with no apparent seed development despite heavy pollination (Fig 7B). These features motivated further investigation of this gene.

Across maize tissues, \( \text{gex2} \) is highly and specifically expressed in sperm cells [38] (Fig 7C). Like many highly expressed genes in mature pollen, it is within 2kb of a transcriptionally active

---

**Table 1. Characteristics of genes showing non-Mendelian inheritance.**

| Category   | Gene designation (v3) | Gene designation (v4) | Gene Type | \( Ds\text{-GFP} \) allele | Male transmission rate | Adjusted p-value | Best BLAST Hit, A. thaliana | Predicted Function (B73v4 Gramene) | Cellular process (inferred) |
|------------|-----------------------|-----------------------|-----------|---------------------------|------------------------|---------------------|---------------------------------|---------------------------------|-----------------------------|
| Vegetative Cell | GRMZM2G359879         | Zm00001d028437         | singleton | \( \text{tdsgR04A02} \) | 43.84%                 | 3.29E-04           | AT3G61050 Calcium-dependent lipid-binding (CaLB domain) family protein | Cell signaling                   |                             |
| Vegetative Cell | GRMZM2G350802         | Zm00001d037695         | singleton | \( \text{tdsgR102H01} \) | 45.50%                 | 3.17E-02           | AT1G52080 Actin binding protein family | Cytoskeleton                    |                             |
| Vegetative Cell | GRMZM2G039583         | Zm00001d022250         | singleton | \( \text{tdsgR33F03} \) | 43.95%                 | 1.38E-04           | AT2G02370 SNARE associated Golgi protein family | Vesicle trafficking              |                             |
| Vegetative Cell | GRMZM2G012328         | Zm00001d003431         | syntelog  | \( \text{tdsgR49F11} \) | 43.87%                 | 1.38E-04           | AT3G05610 Pectinesterase 5 | Cell wall modification           |                             |
| Vegetative Cell | GRMZM2G135570         | Zm00001d014731         | singleton | \( \text{tdsgR67C09} \) | 44.09%                 | 1.06E-03           | AT2G29960 Peptidyl-prolyl cis-trans isomerase CYP20-1 | Protein folding                 |                             |
| Vegetative Cell | GRMZM2G139987         | Zm00001d014782         | singleton | \( \text{tdsgR92F08} \) | 45.06%                 | 1.41E-02           | AT1G19940 Endoglucanase 2 | Cell wall modification           |                             |
| Vegetative Cell | GRMZM2G082517         | Zm00001d015901         | singleton | \( \text{tdsgR96C12} \) | 29.51%                 | 0.00E+00           | AT2G24450 Fasciclin-like arabinogalactan protein 3 | Cell wall modification           |                             |
| Sperm Cell    | GRMZM2G036832         | Zm00001d005781         | singleton | \( \text{tdsgR82A03} \) | 33.43%                 | 4.15E-14           | ATSG49150 Protein GAMETE EXPRESSED 2 | Fertilization                   |                             |
| Sperm Cell    | GRMZM2G036832         | Zm00001d005781         | singleton | \( \text{tdsgR84A12} \) | 23.14%                 | 0.00E+00           | ATSG49150 Protein GAMETE EXPRESSED 2 | Fertilization                   |                             |

[38] doi.org/10.1371/journal.pgen.1008462.t001
| Category        | Gene designation (v3) | Gene designation (v4) | Gene Type | Ds-GFP allele | Male transmission rate | Adjusted p-value | Best BLAST Hit, A. thaliana     | Predicted Function (B73v4 Gramene)                                                                 |
|-----------------|-----------------------|-----------------------|-----------|----------------|------------------------|-----------------|---------------------------------|----------------------------------------------------------------------------------------------------|
| Seedling Only   | GRMZM2G000052         | Zm000001d002266       | syntelog  | tdsGR63F09    | 49.37%                 | 6.99E-01        | AT2G43020                      | Lysine-specific histone demethylase 1                                                                 |
| Seedling Only   | GRMZM2G111143         | Zm000001d004768       | singleton | tdsGR80E09    | 49.09%                 | 6.99E-01        | AT5G24318                      | Glycosyl hydrolase superfamily protein                                                                 |
| Seedling Only   | GRMZM2G0007283        | Zm000001d005036       | syntelog  | tdsGR83H05    | 49.22%                 | 6.99E-01        | AT5G01020                      | Serine/threonine-protein kinase                                                                 |
| Seedling Only   | GRMZM2G129209         | Zm000001d007228       | singleton | tdsGR65E02    | 48.00%                 | 6.83E-01        | AT5G05580                      | Omega-3 fatty acid desaturase                                                                    |
| Seedling Only   | GRMZM2G051403         | Zm000001d012395       | singleton | tdsGR46C04    | 51.28%                 | 6.99E-01        | AT3G55250                      | Nuclear pore complex protein Nup214                                                                 |
| Seedling Only   | AC217975.3_FG001      | Zm000001d002274       | singleton | tdsGR106F04   | 49.37%                 | 6.99E-01        | AT3G06483                      | pyruvate orthophosphate dikinase                                                                 |
| Seedling Only   | GRMZM2G100288         | Zm000001d009047      | syntelog  | tdsGR76E07    | 49.79%                 | 8.83E-01        | AT3G51530                      | Receptor-like protein kinase FERONIA                                                                |
| Seedling Only   | GRMZM2G127798         | Zm000001d035925       | singleton | tdsGR53F11    | 48.87%                 | 6.99E-01        | AT3G02360                      | 6-phosphogluconate dehydrogenase                                                                  |
| Seedling Only   | GRMZM2G342243         | Zm000001d036283       | syntelog  | tdsGR52B09    | 48.45%                 | 6.99E-01        | AT1G45688                      | Late embryogenesis abundant protein group 2                                                        |
| Seedling Only   | GRMZM2G044882         | Zm000001d051110       | singleton | tdsGR12H07    | 49.43%                 | 6.99E-01        | AT5G18430                      | GDSL esterase/lipase LTL1                                                                         |
| Vegetative Cell | GRMZM5G876898         | Zm000001d002258       | singleton | tdsGR81G05    | 51.26%                 | 6.80E-01        | AT1G11860                      | Aminomethyltransferase                                                                          |
| Vegetative Cell | GRMZM2G142863         | Zm000001d003947       | syntelog  | tdsGR83B04    | 48.73%                 | 7.03E-01        | AT5G63750                      | 2-oxoglutarate dehydrogenase E1 component                                                          |
| Vegetative Cell | GRMZM5G827174         | Zm000001d007845       | syntelog  | tdsGR52E07    | 47.09%                 | 2.88E-01        | AT1G10020                      | Formin-like protein 18                                                                              |
| Vegetative Cell | GRMZM2G045278         | Zm000001d012382       | singleton | tdsGR107C12   | 48.33%                 | 4.26E-01        | AT3G53990                      | Adenine nucleotide alpha hydrolases-like superfamily protein                                       |
| Vegetative Cell | GRMZM2G045278         | Zm000001d012382       | singleton | tdsGR34C11    | 48.78%                 | 6.80E-01        | AT3G53990                      | Adenine nucleotide alpha hydrolases-like superfamily protein                                       |
| Vegetative Cell | GRMZM2G102912         | Zm000001d015242       | singleton | tdsGR99B02    | 49.27%                 | 9.07E-01        | AT2G24390                      | AIG2-like protein                                                                                 |
| Vegetative Cell | GRMZM2G056252         | Zm000001d017840       | syntelog  | tdsGR41F01    | 49.62%                 | 9.36E-01        | AT3G12120                      | Delta(12)-fatty-acid desaturase                                                                   |
| Vegetative Cell | GRMZM5G872068         | Zm000001d017958       | syntelog  | tdsGR98H09    | 47.47%                 | 1.69E-01        | AT1G66200                      | Glutamine synthetase root isozyme 3                                                                 |
| Vegetative Cell | GRMZM2G136508         | Zm000001d025437       | singleton | tdsGR31H05    | 50.18%                 | 9.68E-01        | AT2G01170                      | Amino acid permease                                                                               |
| Vegetative Cell | GRMZM2G126858         | Zm000001d026303       | syntelog  | tdsGR23D05    | 51.10%                 | 7.26E-01        | AT1G56145                      | Putative leucine-rich repeat receptor-like protein kinase family protein                            |
| Vegetative Cell | GRMZM2G120136         | Zm000001d026445       | syntelog  | tdsGR24D03    | 49.47%                 | 9.36E-01        | AT1G45180                      | E3 ubiquitin-protein ligase MBR2                                                                   |
| Vegetative Cell | GRMZM2G006894         | Zm000001d026490       | syntelog  | tdsGR02D02    | 49.57%                 | 9.68E-01        | AT4G30190                      | proton-exporting ATPase4                                                                          |
| Vegetative Cell | GRMZM2G172751         | Zm000001d027590       | singleton | tdsGR35A08    | 48.21%                 | 6.80E-01        | AT2G33420                      | Protein of unknown function (DUF810 domain)                                                       |
| Vegetative Cell | GRMZM2G035243         | Zm000001d027856       | syntelog  | tdsGR72D11    | 49.84%                 | 9.68E-01        | AT1G14330                      | Kelch motif family protein                                                                       |

Continued...
Table 2. (Continued)

| Category       | Gene designation (v3) | Gene designation (v4) | Gene Type | Male transmission rate | Adjusted p-value | Best BLAST Hit, A. thaliana     | Predicted Function (B73v4 Gramene) |
|----------------|-----------------------|-----------------------|-----------|------------------------|------------------|--------------------------------|-----------------------------------|
| Vegetative Cell| GRMZM2G016734         | Zm000001d028820       | syntelog  | 48.34%                 | 6.80E-01         | AT1G56300                      | Chaperone DnaJ-domain superfamily protein |
| Vegetative Cell| GRMZM2G142249         | Zm000001d032279       | singleton | 49.17%                 | 8.88E-01         | AT3G47730                      | ABC2 homolog 15                     |
| Vegetative Cell| GRMZM2G114093         | Zm000001d032310       | singleton | 49.57%                 | 9.36E-01         | AT3G58950                      | Protein kinase superfamily protein |
| Vegetative Cell| GRMZM2G124434         | Zm000001d032950       | singleton | 48.39%                 | 6.80E-01         | AT5G28840                      | GDP-mannose 35-epimerase            |
| Vegetative Cell| GRMZM5G878153         | Zm000001d034799       | singleton | 49.33%                 | 9.36E-01         | AT3G03320                      | RNA-binding ASCH domain protein    |
| Vegetative Cell| GRMZM2G134054         | Zm000001d034839       | singleton | 50.28%                 | 9.68E-01         | AT3G07130                      | Purple acid phosphatase 15          |
| Vegetative Cell| GRMZM2G134054         | Zm000001d034839       | singleton | 50.00%                 | 1.00E+00         | AT3G07130                      | Purple acid phosphatase 15          |
| Vegetative Cell| GRMZM2G307402         | Zm000001d036330       | singleton | 48.26%                 | 6.80E-01         | AT2G28200                      | C2H2-type zinc finger family protein |
| Vegetative Cell| GRMZM2G012263         | Zm000001d037061       | singleton | 50.08%                 | 9.80E-01         | AT5G14130                      | Peroxidase 64                      |
| Vegetative Cell| GRMZM2G012263         | Zm000001d037061       | singleton | 50.39%                 | 9.36E-01         | AT5G14130                      | Peroxidase 64                      |
| Vegetative Cell| GRMZM2G018372         | Zm000001d041514       | syntelog  | 50.55%                 | 9.36E-01         | AT1G18670                      | Protein kinase superfamily protein IBS1-like |
| Vegetative Cell| GRMZM2G095206         | Zm000001d046483       | singleton | 52.43%                 | 2.94E-01         | AT4G09750                      | NAD(P)-binding Rossmann-fold superfamily protein |
| Vegetative Cell| GRMZM2G089699         | Zm000001d048384       | complex   | 51.04%                 | 8.88E-01         | AT1G65680                      | beta expansin10a                   |
| Vegetative Cell| GRMZM5G845021         | Zm000001d048785       | singleton | 47.30%                 | 4.25E-01         | AT3G03900                      | Adenylyl-sulfate kinase 3          |
| Sperm Cell     | GRMZM2G365613         | Zm000001d066218       | singleton | 51.47%                 | 7.46E-01         | AT5G42560                      | HVA22-like protein i               |
| Sperm Cell     | GRMZM2G100318         | Zm000001d012128       | syntelog  | 50.03%                 | 9.79E-01         | AT1G67710                      | Putative two-component response regulator family protein |
| Sperm Cell     | GRMZM2G172726         | Zm000001d021974       | syntelog  | 51.10%                 | 7.46E-01         | AT1G9360                       | Arabinosyltransferase RRA3         |
| Sperm Cell     | GRMZM2G160069         | Zm000001d025834       | singleton | 47.96%                 | 7.46E-01         | AT4G16480                      | Inositol transporter 4             |
| Sperm Cell     | GRMZM2G114899         | Zm000001d034788       | syntelog  | 49.24%                 | 7.46E-01         | AT1G77280                      | Protein kinase protein with adenine nucleotide alpha hydrorases-like domain |
| Sperm Cell     | GRMZM2G007659         | Zm000001d042810       | syntelog  | 51.03%                 | 7.46E-01         | AT2G29680                      | Cell division control protein 6 homolog B |
| Sperm Cell     | GRMZM2G038252         | Zm000001d043076       | syntelog  | 50.99%                 | 7.46E-01         | AT4G35550                      | WUSCHEL-related homeobox 13a      |
| Sperm Cell     | GRMZM2G099382         | Zm000001d041109       | singleton | 49.33%                 | 7.46E-01         | AT5G47560                      | Tonoplast dicarboxylate transporter |
| Sperm Cell     | GRMZM2G352898         | Zm000001d048434       | singleton | 51.31%                 | 7.46E-01         | AT3G63240                      | Type IV inositol polyphosphate 5-phoshatase 7 |

*Beta expansins have proliferated in the maize genome, including by tandem duplication [77]. Thus, this gene cannot be strictly characterized as syntelog; however, the presence of multiple paralogs in the genome indicate it should not be categorized as a singleton gene.

https://doi.org/10.1371/journal.pgen.1008462.t002

TE, a downstream RLG retrotransposon that displays sperm cell-specific activation (Fig 7C). The Zm gex2 gene was first identified via EST sequencing of maize sperm cells [44], and subsequently used to isolate the Arabidopsis ortholog, named GAMETE EXPRESSED2 (GEX2) [45].
Fig 7. Mutations in the sperm cell-specific gex2 gene cause aberrant seed development. (A) The exon/intron structure of gex2 (Zm00001d005781/GRMZM2G036832), showing the locations of the two independent Ds-GFP insertion mutants. (B) Ear projections of gex2 mutant outcrosses. Top: heterozygote outcrossed as female, showing 1:1 transmission of the GFP-tagged allele. Middle: heterozygote outcrossed as a male, with 26.1% transmission of the mutant allele. Additionally, small seeds and occasional, small gaps between seeds are visible. Bottom: homozygous mutant outcrossed as a male, with many small seeds and large gaps, despite heavy pollination. (C) Genomic neighborhood of the GEX2 locus, with two nearby TEs, and their RNA-seq expression levels across male reproductive development. (D) Predicted domain structure of Zm GEX2; the amino acid sequence shows 44.2% similarity with Arabidopsis GEX2. (E) Quantification of small/aborted seeds resulting from pollination by gex2 mutant plants and controls. Controls included two Ds-GFP lines that did not show transmission defects (tdsgR12H07 and tdsgR46C04) and one Ds-GFP line that showed a strong transmission defect in the vegetative cell group (tdsgR96C12; 29.5% transmission). A higher percentage of small/aborted seeds was present following pollination by heterozygous gex2 plants representing the two mutant alleles (tdsgR82A03 and tdsgR84A12), and pollination by homozygous gex2-tdsgR84A12 plants further increased the percentage of small/aborted seeds.

https://doi.org/10.1371/journal.pgen.1008462.g007
In Arabidopsis, GEX2 is necessary for effective double fertilization, causing seed abortion and empty spaces in the silique when a mutant allele is inherited through the male [46]. Zm gex2 encodes a protein with similar structure and amino acid sequence to its Arabidopsis ortholog (Fig 7D).

Small and aborted seeds were quantified for both gex2 insertion alleles when outcrossed to wild-type plants (S10 Table, S5A Fig). Two different Ds-GFP insertion lines that were not associated with transmission defects (tdsgR12H07, tdsgR46C04), as well as the Ds-GFP associated with the strongest male transmission defect in the Vegetative Cell class (tdsgR96C12), were used as controls. Pollination with both gex2::Ds-GFP insertion alleles was associated with increased percentages of small or aborted seeds, significantly so in gex2-tdsgR84A12 (pairwise t-test against Ds-GFP controls, all p-values < 0.05), and pollination from gex2-tdsgR84A12 homozygotes approximately doubled the percentage of aberrant seeds from heterozygotes (pairwise t-test against Ds-GFP controls, all p-values < 0.01) (Fig 7E). From the heterozygous gex2::Ds-GFP crosses, small seeds with endosperm large enough for DNA preparation were genotyped, and 79.2% were found to harbor the gex2 mutation, whereas the tdsgR46C04 control showed Mendelian segregation in small seeds (S5B Fig). These data support the hypothesis that aberrant seed development is induced by fertilization by gex2::Ds-GFP sperm.

If the Zm GEX2 protein acts to promote double fertilization similarly to its Arabidopsis ortholog, the arrival of a gex2::Ds-GFP pollen tube at the embryo sac could lead to failure of one or both fertilization events. Given an active polytubey block, this could produce the observed gaps between seeds on the ear, resulting from ovules associated with completely failed fertilization, or with very early seed abortion due to single fertilization. Consistent with this possibility, pollination with both heterozygous and homozygous gex2-tdsgR82A03 alleles resulted in increased seedless area relative to controls (S6 Fig). To test for aberrant fertilization more directly, seed development was assessed at 4 days post-pollination with either wild-type or gex2-tdsgR84A12 homozygous pollen (Fig 8, Table 3). Typical embryo and endosperm development, as well as indication of the polytubey block (i.e., arrival of only single pollen tubes at the embryo sac), was observed in all ovules assessed from wild-type pollination. In contrast, half of the ovules assessed following pollination with gex2::Ds-GFP showed significant evidence of abnormal double fertilization, demonstrating single fertilization of either embryo or endosperm or indication of arrival of more than one pollen tube at the embryo sac (Fisher’s exact test, p-value = 0.000241). We conclude that in maize, similarly to Arabidopsis, Zm GEX2 is part of the sperm cell machinery that helps ensure proper double fertilization.

Discussion

The Zm gex2 gene has a conserved role in promoting double fertilization

The generation of a well-replicated developmental time course of transcriptomic data enabled the targeting of highly expressed genes in vegetative and sperm cells for mutational screening. Two independent insertions in the highly and specifically expressed maize sperm cell gene gex2 led not only to severely reduced transmission through the male, but also, in contrast to other mutations analyzed in this study, to paternally triggered post-fertilization defects. Zm gex2 was first identified in maize by sperm cell EST sequencing [44], which led to identification of the orthologous gene in Arabidopsis, GEX2, and its sperm cell-specific promoter [45]. In Arabidopsis, single fertilization events were observed at increased frequency in GEX2 mutant-pollinated plants, both for the egg cell and the central cell, leading to an observed increase in aberrant seed development and abortion [46]. Our results in maize are similar, with gex2 mutant pollen resulting in unfilled ovules, single fertilization events in embryo sacs, and aberrant early seed development from embryo sacs fertilized by gex2::Ds-GFP sperm cells. In
Arabidopsis, the interaction between the plasma membrane-localized GEX2 and either the female egg or central cells has been suggested to contribute to gamete attachment. The two orthologues share a predicted domain structure, including a large N-terminal non-cytoplasmic

Table 3. Seed development at 4 days after pollination by wild-type or gex2::Ds-GFP pollen.

| Pollen parent | One synergid penetrated by a pollen tube | Both synergids penetrated by a pollen tube |
|---------------|----------------------------------------|------------------------------------------|
|               | Both embryo and endosperm | Endosperm without embryo | Embryo without endosperm | Both embryo and endosperm | Endosperm without embryo | Embryo without endosperm |
| gex2-gsgR84A12/gex2-tsgR84A12 | 6 | 2 | 2 | 0 | 0 | 2 |
| Wild-type | 28 | 0 | 0 | 0 | 0 | 0 |

Fig 8. Pollination by gex2-tsgR84A12 leads to aberrant fertilization events and developing seed phenotypes. (A) Seed development in a typical ovule pollinated by wild-type pollen, with one synergid penetrated by a pollen tube, and both embryo and endosperm development initiated. (B-D) Abnormal phenotypes seen following gex2-tsgR84A12 pollination. (B) Ovule with developing (cellularizing) endosperm but unfertilized egg cell. (C) Ovule with developing embryo but unfertilized central cell. (D) Ovule with both synergids penetrated by a pollen tube, and a developing embryo and unfertilized central cell. emb = embryo; endo = endosperm; ps = synergid penetrated by a pollen tube; PN = polar nuclei.

https://doi.org/10.1371/journal.pgen.1008462.g008

Arabidopsis, the interaction between the plasma membrane-localized GEX2 and either the female egg or central cells has been suggested to contribute to gamete attachment. The two orthologues share a predicted domain structure, including a large N-terminal non-cytoplasmic
region containing filamin repeat domains potentially contributing to this interaction [46], raising the possibility that Zm GEX2 acts similarly during double fertilization. With conserved GEX2-like genes widely distributed throughout the currently sequenced Embryophyta taxa, our results support the idea that, in flowering plants, these genes play a crucial role in double fertilization.

**Maize pollen provides a powerful model for quantifying gene-specific contributions to fitness**

The transcriptome dataset also provided a framework to ask broader questions regarding potential relationships between elevated expression and gene function, i.e., utilizing *Ds-GFP* insertions not merely in a genetic screen, but in a mutational interrogation of gene function guided by quantitative hypotheses. Despite the explosion of omic-scale methods to characterize genomes and to measure molecular characters (e.g., transcript levels), our ability to predict phenotypic relevance for specific genes is limited, particularly in multicellular organisms. A simple hypothesis is that a high transcript level at a particular developmental stage implies functional relevance for the associated gene at that stage, and thus potential for phenotypic influence, a hypothesis supported by observations in mice [47]. This study addresses this hypothesis in plants with a systematic assessment of the functional relevance of highly expressed genes in maize pollen, taking advantage of the ease of reciprocal outcross pollination in maize, the availability of a sufficient number of marked and likely null mutations, and the development of an imaging technique that enables sensitive quantitation.

In the post-pollination progamic stage, pollen grains, as independent, genetically distinct organisms, compete to be the first to deliver the sperm cells to the embryo sac for double fertilization. In an outcrossing plant with an extensive stigma and style like maize, there is likely a heightened context for competition among these individuals [10], thus providing a milieu that may be particularly sensitive to genetic perturbation. We reasoned that genes highly expressed in the vegetative cell would tend to contribute to a competitive advantage at this stage, which is responsible for pollen tube germination and growth. Indeed, we found that CDS-insertion alleles for 7 out of 32 (21.9%) tested genes in this class are associated with mild to moderate male-specific transmission defects, with the majority of these defects classified as mild (~45% transmission) and thus detectable only by assessing large populations. In this class, transcript level was significantly correlated with both reduced transmission rate and an increased likelihood of significant non-Mendelian transmission, although the explanatory power is limited ($R^2 < 0.2$), as expected for a complex biological system.

Genetic redundancy could also be predicted to influence the phenotypic outcome of mutating single genes, and there is a suggestive trend consistent with this idea in our dataset (Fisher’s exact test, $p$-value = 0.2117), with 6/7 non-Mendelian alleles classified as singletons in the maize genome (86%), compared to 14/25 singletons in genes harboring Mendelian alleles (56%) (Tables 1 and 2). In addition, the variety of biological processes predicted for genes with documented fitness contributions (Table 1) is consistent with the idea that competitive pollen tube growth requires an array of cellular functions. It should be noted that our approach relies on the availability of *Ds-GFP*-marked insertion alleles, and it seems likely that such availability is biased against genes with severe transmission phenotypes, as these would be selected against in a transposon-mutagenized population. In fact, two large *Mutator* transposon populations show a statistically significant deficit in insertions in genes associated with gametophyte-enriched expression [23]. Notably, both previously described maize genes associated with severe male transmission defect mutants (*rop2, apt1*; transmission <15% [11,48]) also would be classified as highly expressed in the vegetative cell ($\log_2(FPKM) > 8$), suggesting that the
trend we found is applicable even to the types of genes most likely to be absent from the *Ds-
GFP* insertion population.

Few studies have investigated the relationship between transcriptomic data and quantitative
mutant phenotypes, particularly in multicellular organisms. Large-scale screening of the Ara-
bidopsis T-DNA mutant collection for leaf or reproduction-related phenotypes has identified
effects in ~4% of lines assessed [49,50], but these efforts were not guided by transcriptome
data. Our use of a sensitive phenotypic assay, combined with a focus on sampling mutations in
genes that are most highly expressed at a developmental stage relevant to the phenotype tested,
seem likely to have contributed to the higher frequency of phenotypic effects we found. In
mice, a meta-analysis found a relationship similar to the one we observed, with aberrant phe-
notypes more likely to be associated with genes highly transcribed in the tissue exhibiting the
phenotype [47]. In contrast to these results, genome scale measurement of the fitness costs of
gene knockouts via competitive assays in yeast [51,52] and bacteria [53,54] found that, for par-
ticular environmental conditions, there was little to no correlation between the expression
level of a gene and its impact on fitness in that condition. This could be indicative of differ-
ences between the mechanisms underlying single-celled organisms’ response to the environ-
ment versus those underlying developmental complexity in multicellular organisms.

Interestingly, our results are consistent with a recent meta-analysis that identified higher
mRNA expression levels as a feature distinguishing gene models with known mutant pheno-
types from the overall population of gene models defined by molecular approaches [55].

Transposable element dynamics in the maize male gametophyte

The transcriptomic time course enabled exploration of the dynamic relationships among develop-
mental progression, gene expression levels, and transcriptional activation of TEs. Understanding
TE expression during maize male reproductive development provides an informative comparison
to similar analyses in Arabidopsis, an evolutionarily distant plant with a genome landscape that is
distinct from maize. Although maize has a higher number and percentage of its genome occupied
by TEs compared to Arabidopsis, we found that only a fraction of maize TEs are developmentally
dynamic with regards to transcript accumulation. These ‘dynamic’ TEs tend to be longer elements
than average, which suggests that they have protein coding and transpositional potential. From
this dynamic TE set, we were able to identify individual elements that are expressed in a number
of specific tissues. However, more globally, there is a trend towards activation of TE transcription
over the course of the development of the male gametophyte. This finding confirms that both
monocots and eudicots have developmental activation of TE expression in pollen. Consistent with
our findings, a recent study found that spontaneous retrotransposon mutations are much more
frequent through the male than the female in certain maize lines [31]. This conservation suggests
that the roles of TE and TE-induced small RNAs during reproductive development may also be
conserved between monocots and eudicots [35,56,57].

Although TE activation is conserved in maize and Arabidopsis pollen, we have identified
key differences in the timing and location. Maize TE activation is detected earlier (in the uni-
cellular microspore) compared to when it is thought to occur in Arabidopsis [28]. Transcripts
from these early-activated TEs in the microspores typically remain detectable through pollen
development and in the mature pollen grain, which may be due to continued expression or
transcript stability. A second distinction is the location of activation, which in Arabidopsis is
confined to the pollen vegetative cell nucleus [28,32], whereas in maize also occurs in sperm
cells. *Mutator* family TEs are overrepresented in the pool of sperm-cell transcripts, suggesting
that this family of TEs may have evolved (or co-opted) specific regulatory mechanism(s) such
as an enhancer element that confers expression in this cell type.
Given our results indicating a linkage between elevated gene expression levels and functional relevance, we also assessed whether similar correlations exist between gene and TE expression locally in the maize genome. We found that in mature pollen and sperm cells there is a positive correlation: the more highly expressed a gene is, the more likely it is to have an up-regulated TE nearby. This tissue-specific correlation is a developmentally-specific co-regulation of gene and TE expression. Notably, there does not appear to be any strong trend linking this co-regulation to gene function. We find instances of local gene/TE coordinate regulation are present in similar proportions in genes with documented transmission defects vs. those showing Mendelian segregation when mutant (14% vs 28% respectively, S5 Table). Therefore, although our data indicate the population of highly expressed maize pollen genes has a tendency to contribute to pollen fitness, and a tendency to be adjacent to pollen-expressed TEs, these two characteristics appear to be independent.

Several potential mechanisms may account for the coordinated gene/TE expression. First, programmed activation of TE expression may influence chromatin, enhancer, or other regulatory functions that influence neighboring genes. Second, the genes and TEs may be directly controlled by the same mechanism of large-scale epigenetic activation, limiting the expression of both to a specific tissue or developmental time point. Third, gene activation may influence the expression of the neighboring TE via read-through transcription. Future studies using alternative transcriptomic approaches (e.g., CAGE or long-read RNA sequencing) will enable the dissection of these possible mechanisms.

Methods

Plant materials

Maize inbred line B73 was used for all RNA isolations. Plants were grown in a controlled greenhouse environment (16 hrs light, 8 hrs dark, 80 F day/70 F night) and in the field at the Botany & Plant Pathology Field Lab (Oregon State University, Corvallis, OR) using standard practices. Lines containing Ds-GFP insertion alleles were acquired from the Maize Genetics Cooperation Stock Center.

RNA isolation, library preparation and sequencing

Detailed methods are available in S1 Methods. Briefly, tissue was isolated either by dissection (TP), differential density centrifugation (MS, MS-B and SC), or collection at anthesis (MP). Total RNA from TP, MS, and MP was extracted using a modified Trizol Reagent (Life Technologies) protocol; SC total RNA was extracted via a phenol/chloroform protocol. Poly-A RNA (mRNA) was isolated using streptavidin magnetic beads (New England Biolabs, # S1420S) and a biotin-linked poly-T primer. RNA libraries were prepared and sequenced by the Central Services Lab (CSL) at the Center for Genome Research and Biocomputing (CGRB, Oregon State University) using WaferGen robotic strand specific RNA preparation (WaferGen Biosystems) with an Illumina TruSeq RNA LT (single index) prep kit and run on an Illumina HiSeq 3000 with 100 bp paired-end reads.

Mapping reads to genes, differential expression assessment and GO enrichment analysis

Ribosomal reads (rRNA) were removed from all samples using STAR, version 2.5.1b [58] to map reads to a repository of maize rRNA sequences (parameters: --outSAMunmapped Within --outSAMattributes NH HI AS NM MD --outSAMstrandFieldintronMotif --limitBAMsor-tRAM 50000000000 --outReadsUnmapped Fastx). The number of mappable reads generated
from each sample after rRNA removal ranged from approximately 1 million to approximately 41 million, with an average mappable reads of approximately 18 million per sample. Total reads, mapped reads, rRNA contamination, and other statistics are summarized in S1 Table.

rRNA-filtered sequences were mapped to the maize reference genome, version B73 RefGen_v4.33 [36] using STAR, keeping only unique alignments (parameters: - -outSAMun-mapped Within --outSAMattributes NH HI AS NM MD --outSAMstrandField intronMotif --outFilterMultimapNmax 1 --limitBAMsortRAM 50000000000). Transcript levels of annotated gene isoforms were measured using Cufflinks, version 2.2.1 [59]. FPKM (fragments per kilobase of transcript per million mapped reads) values are shown in S4 Table. Differential expression was calculated between each tissue with Cuffdiff, version 1.0.2, using default parameters. FPKM counts were normalized using the geometric library normalization method. A pooled dispersion method was used by Cuffdiff to model variance. Differential expression results are summarized in S11 Table.

Gene ontology (GO) terms were found for enriched genes in each tissue using the AgriGO 2: GO Analysis Toolkit [60]. Enriched genes were defined as the top 300 significantly differentially expressed genes (q-value) from Cuffdiff output, with ties broken by log2 fold change. Enriched sets were split into up- and down-expressed genes. GO term enrichment was calculated using the singular enrichment analysis method with a Fisher test and Yekutieli multi-test adjustment. GO annotations were based off the maize-GAMER annotation set [61].

Mapping reads to transposable elements

The rRNA-filtered reads were quality trimmed (QC30) and adapter sequences were removed using BBDUK2 [62]. The remaining sequences were mapped to the whole genome using STAR, allowing mapping to at most 100 'best' matching loci. (parameters: - -outMultimapper-Order Random --outSAMmultNmax -1 --outFilterMultimapNmax 100). For paired-end reads, the unmapped reads were re-mapped using single-end approach to maximize the number of mappable reads. The mapping percentage is reported in S3 Table. Because 19% of the total reads in the dataset mapped to more than one location, such reads were mapped to only their best match in the genome, and when multiple best matches existed, they were mapped to all of these loci, and then counted fractionally. For example, if one read maps to 4 TE locations equally well, each TE would receive a weighted value of 0.25 mapped reads. Because the TE expression of the aberrant SC1 biological replicate did not cluster with the other three SC replicates (S1 Fig), it was removed from all subsequent analysis of TE expression.

Principal component analysis (PCA)

Using the maize gene and TE annotation file available from Ensembl Genomes (v38) [36], a combined annotation file was generated for both genes and TEs to run PCA for all samples. FeatureCounts [63] was used to calculate the accumulation of each gene and TE in all samples following fractional assignment of reads (parameters: -O - -largestOverlap -M - -fraction -p -C). This counts file was used in DESeq-2 [64] to generate the PCA plot.

Analysis of transposable elements

The featureCount file (described above) was used as input for differential expression analysis using DESeq2. Since DESeq2 accepts only integers as raw counts, we used ‘round’ function of R to round the counts to their nearest integers. For differential expression using default parameters for normalization in DESeq2, we only included TEs with a sum total of > = 10 read counts across samples; the rest were categorized as ‘not covered’ TEs. First, normalized read counts for all TEs were obtained (data in S2 Fig) and then, only TEs (farther than 2kb from
genes) were filtered to be investigated further whereas TEs less than 2kb away from a gene were categorized as ‘near genes’.

After selecting seedling as the reference tissue, pairwise volcano plots were generated for all samples against the reference seedling tissue. Each pairwise comparison with the seedling tissue yielded a set of TEs with adjusted p-value of either ‘NA’ or a real number. The set of TEs with a p-value of ‘NA’ in all pairwise comparisons was added to the count of ‘not covered’ category since there was not enough statistical power to call differential expression in any of the tissues. The number of TEs statistically significantly up- and down-regulated (adjusted p-value < 0.05) in each tissue was calculated, categorized as ‘dynamic TEs’ and plotted (Fig 3). Additionally, the number of TEs with adjusted p-value > 0.05 were categorized as developmentally ‘static TEs’ as no evidence of TE expression was observed over different developmental time points analyzed. For all categories, the length, family or distance from centromere was calculated based on the published TE annotation file.

Validation of Ds-GFP insertion sites

A FASTA file containing 2 kb of genomic sequence surrounding each Ds-GFP insertion site was used as input to a primer3-based tool to generate a pair of specific primers to genotype individual plants from each line (https://vollbrechtlab.gdcb.iastate.edu/tools/primer-server/). The primers used for each Ds-GFP line are listed in S6 Table.

To genotype the plants, two 7 mm discs of leaf tissue were collected from each plant using a modified paper punch. The samples were collected in 1.2 ml tubes that fit within a labeled 96 well plate/rack (https://vollbrechtlab.gdcb.iastate.edu/tools/tissue-sample-plate-mapper/) (Phenix Research Products, Candler, NC; M845 and M845BR or equivalent). Genomic DNA was isolated from the leaf punches [65] with the following modifications. An additional centrifugation (3,000 g for 10 min.) was added to clear the leaf extracts prior to loading onto a 96-well glass fiber filter plate (Pall, 8032). DNA was eluted from filter plates in 125 μL water, and 2 μL was used as template for PCR. Amplification followed standard PCR conditions using GoTaq Green Master Mix (Promega) with 4% DMSO (v/v) and amplicons were resolved using agarose gel electrophoresis. Lines were genotyped using the pair of Ds-GFP line gene-specific primers plus one Ds-specific primer (JSR01 GTTCGAAATCGATCGGGATA or JGP3 ACCGACCGATCGTATCGG). All lines were also screened by PCR for the presence of wx1-m7::Ac using primers for wx1 (CACAGCACGTTGCG ATTTC) and Ac (CCGGATCG TATCGGTTTTCG). Followup PCR to test for co-segregation of GFP fluorescence with the presence of the insertion used the appropriate set of three PCR primers (two gene-specific and one Ds-specific) and DNA prepared either from endosperm or seedling leaves [66].

Insertional mutagenesis transmission quantification and statistics

Heterozygous lines with PCR-validated Ds-GFP insertion alleles were planted in the Botany & Plant Pathology Field Lab (Oregon State University, Corvallis, OR). All insertions were in coding sequence (CDS) sites. Heterozygous Ds-GFP plants were outcrossed to tester plants (c1/c1 wx1/wx1 or c1/c1 genetic background) through both the female and the male, with male pollinations made with a heavy pollen load on extended silks (silks that had been allowed to grow for at least two days following cutback). Following harvest, resulting ears were imaged using a custom rotational scanner in the presence of a blue light source and orange filter for GFP seed illumination [43]. Briefly, videos were captured of rotating ears, which were then processed to generate flat cylindrical projections covering the surface of the ear (for examples, see Figs 5 and 7). Seeds were manually counted using the Cell Counter plugin of the Fiji distribution of ImageJ [67]. Ears showing evidence of more than a single Ds-GFP insertion (~75% GFP
transmission) were excluded from further analysis. For an allele to be included in the final dataset, we required a minimum of three independent male outcrosses from two different plants.

Seed transmission rates of remaining ears were quantified using a generalized linear model with a logit link function for binomial counts and a quasi-binomial family to correct for over-dispersion between parent lines. By incorporating overdispersion, we allowed for the possibility that seeds on the same ear were not completely independent, and for varying transmission rates between ears associated with a given mutation (e.g. by environmental or maternal effects). A quasi-likelihood approach is more realistic than a simple chi-square test, which assumes that all seeds are independent and transmission rates between ears associated with a given mutation are the same. The dispersion parameter for Sperm Cell and Vegetative Cell categories was approximately 1.8, indicating substantially more heterogeneity among seeds on different ears than is expected in a model which assumes independence. Non-Mendelian inheritance was assessed with a quasi-likelihood test with p-values corrected for multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate at 0.05. Significant non-Mendelian segregation was defined with an adjusted p-value < 0.05. Proportions of genes with male-specific transmission defects in the Seedling, Sperm Cell, and Vegetative Cell categories were compared using a two-sided Fisher’s exact test, with significance defined as a p-value < 0.05. A two-sided test for equality of proportions with continuity correction was used to compare transmission rates in families with partial Ac presence.

A Git repository containing statistical tests and plotting information for this portion of the study can be found at https://github.com/fowler-lab-osu/maize_gametophyte_transcriptome.

## Zm gex2 sequence analysis and phenotype characterization

Protein sequence encoded by Zm gex2 (Zm00001d0005781_T002) was retrieved from the Maize Genetics and Genomics Database (MaizeGDB) hosting of the B73 v4 genome [36,68]. Arabidopsis GEX2 protein sequence (AT5G49150.3) was retrieved from the Arabidopsis Information Portal (ARAPORT) Col-0 Araport11 release [69,70]. Maize and Arabidopsis GEX2 protein domains were predicted by InterPro [71], with transmembrane helix predictions by TMHMM [72]. Prediction of land plant species GEX2 conservation was retrieved from PLAZA, gene family HOM04M006791 [73]. Zm gex2 gene duplication searches were performed using BLAST [74] and the B73 v4 genome. To confirm the predicted insertion sites for the two gex2::Ds-GFP alleles, flanking insertion site fragments were PCR-amplified with a gene-specific primer and a Ds-GFP-specific primer (DsGFP_3UTR–TGCAAGCT CGAGTT TCTCCA) and sequenced via Sanger sequencing.

To quantify small seed phenotype, mature, dried down maize ears were imaged prior to seed removal from the ear. For small seeds selection, the ear was first visually scanned row by row from the top to the bottom of the ear. Seeds that were noticeably smaller than their surrounding (regular-sized) seeds are carefully removed from the ear using a pin tool. This sometimes required the removal of regular-sized surrounding seeds, which were saved for later counting. A second visual inspection of the ear often resulted in additional small seeds and is recommended. All remaining seeds were then removed from the ear by hand or using a hand corn sheller tool (Seedburo Equip. Co., Chicago, IL). The ear was screened again for any small (flat/tiny) seeds that could have been missed previously. The cob was inspected prior to discarding, and if any small seed was left behind it was removed and accounted for. Small/smaller seeds and regular-sized seeds were counted and counts were recorded (S10 Table). To measure
seedless area, ears were scanned as previously described to create flat surface projections. "Seedless area" was defined as ear surface area that lacked mature or partially developed seeds. Seedless area was quantified as a percentage of total area, as measured with the "Freehand selection" tool of the Fiji distribution of ImageJ [67]. A Git repository containing statistical tests and plotting information for this portion of the study can be found at https://github.com/fowler-lab-osu/maize_gametophyte_transcriptome.

For analysis of embryo sacs by confocal microscopy, tissues were stained with acriflavine, followed by propidium iodide staining [75,76]. After staining, samples were dehydrated in an ethanol series and cleared in methyl salicylate. Samples were visualized on a Leica SP8 point-scanning confocal microscope using excitations of 436 nm and 536 nm and emissions of 540 ± 20 nm and 640 ± 20 nm.

Supporting information

S1 Fig. Principal component analysis of gene and transposable element (TE) expression levels. Two major components, on x- and y-axis, explain 89% of the variance in gene and TE expression levels. Asterisk (*) mark indicates the sample generated as part of this study, whereas other datasets are publicly available. For the sperm cells isolated in this study (SC), the TE expression of one biological replicate did not cluster with the other three (SC1), and therefore was removed from subsequent analyses of expression from TEs. MP-2014, SE, and OV are from [23]; MP-WEB is from [38]; LF is from [37]; MP-LM is from NCBI BioProject 306885 (2015); SC-TD is from [27].

S2 Fig. Seedling tissue is the appropriate reference for comparison of TE activity. (A) Steady-state mRNA accumulation of all TEs in different tissues. Datasets generated in this study are marked with an asterisk. (B) The number of TEs with zero or near-zero expression levels in different tissues. Seedlings (SE) have the most TEs with low expression levels.

S3 Fig. Length distribution of categorized TEs subdivided by superfamilies. Length of TEs in the different TE categories from Fig 2A but further subcategorized by different superfamilies. The violin plots around the box show the kernel probability density of the data. The box represents lower and upper quartile, the line is the median, and the whiskers represent 10–90% range. Red asterisk denotes the mean. 'n' shows the number of TEs in each category for each superfamily.

S4 Fig. Abundance of TEs near genes in each tissue. (A) For each tissue type, the top 20,000 expressed genes are distributed along the X-axis in bins of 200, with the highest expressed bin on the far left. The number of TEs near (<2kb) these genes is then counted on the Y-axis (shown in grey bar) and the number of genes with at least 1 TE within 2kb is displayed as black dots. (B) Genes filtered for either higher expression in pollen (MP) over sperm cells (SC) (left) or SC>MP (right) were used to determine if the association in Fig 4 is due to sample contamination between SC and MP. Once genes were filtered, the top expressed genes in that tissue were distributed along the X-axis in bins of 200 based on their expression values, with the highest expressed bin on the far left. The number of up- and down-regulated TEs near (<2kb) these genes is then counted on the Y-axis (shown in grey bar) and the number of genes with at least 1 TE within 2kb is displayed as black dots.
S5 Fig. gex2 mutant pollen is associated with increased small and aborted seeds in outcross progeny. (A) Seeds were removed from ears, arranged according to size, and counted. Images of representative seed populations are shown, with the top two rows in each image showing representative fully developed seeds. Rows below the top two contain all of the smaller or aborted seed from that particular ear. (B) PCR genotyping of small endosperm seeds from two independent crosses for the two gex2 alleles show the majority of small seeds harbor the gex2::Ds-GFP allele, despite overall reduced transmission of the insertion alleles through the male. Small seeds from control tsdgR46C04 crosses segregate in a Mendelian fashion.

(TIF)

S6 Fig. Characterization of gex2 seedless ear area. Seedless area was quantified from scanned ear images for gex2 Ds-GFP alleles and Ds-GFP controls. Pollen from heterozygous gex2 plants did not show significantly increased seedless area (gex2-tdsgR82A03 pairwise t-test p-values relative to GFP line 1, GFP line 2, and VC mutant 0.95, 0.96, and 0.74, respectively; gex2-tdsgR84A12 pairwise t-test p-values 0.19, 0.13, and 0.06, respectively), whereas pollen from homozygous gex2-tdsgR84A12 plants had significantly increased seedless area (pairwise t-test against Ds-GFP controls separately, all p-value < 0.0001).

(TIF)

S1 Methods. Tissue sample preparation, RNA extraction, and analysis of potential confounding variables in insertional mutagenesis lines.

(PDF)

S1 Table. Gene sequencing statistics and availability. Summary statistics for sequencing data generated in the study.

(XLSX)

S2 Table. GO term enrichment results. Differentially expressed genes in developmental categories examined in the study, as well as significantly enriched GO terms associated with these genes.

(XLSX)

S3 Table. Transposable element sequencing statistics and availability. Summary statistics and availability for expression datasets used in the analysis of transposable element expression.

(XLSX)

S4 Table. Genic isoform abundance (FPKM) across developmental stages. Cufflinks output describing isoform expression by developmental stages, separated by biological replicate.

(TXT)

S5 Table. Top 20% transcripts by FPKM in Mature Pollen, Sperm Cell and Seedling datasets. List of top 20% highly expressed genes assigned to the Vegetative Cell, Sperm Cell or Seedling Only classes.

(XLSX)

S6 Table. Insertional mutagenesis alleles and primers. List of alleles tested for the presence of Ds-GFP insertions by PCR, including primers sequences.

(XLSX)

S7 Table. Insertional mutagenesis results by line. Insertional mutagenesis results, separated by line, including marker transmission rates and expression category information.

(XLSX)
S8 Table. Insertional mutagenesis results by allele. Insertional mutagenesis results, separated by allele, including marker transmission rates and expression category information. (XLSX)

S9 Table. Concordance of seed phenotype with DsGFP genotype. PCR results from testing Ds-GFP presence GFP and non-GFP seeds for selected alleles. (XLSX)

S10 Table. gex2 small seed phenotyping. gex2 small seed counting and seedless area results. (XLSX)

S11 Table. Differential expression results. Cuffdiff output comparing expression between tissues examined in this study. (TXT)

Acknowledgments

We thank O. Childress, H. Fowler, B. Galardi, B. Hamilton, R. Hartman, and C. Lambert for their tireless seed counting, genotyping, field work, and other technical assistance; and Dr. Lian Zhou for her contributions to maize field genetics. We also thank K. Wimalanathan and T. Shibamoto for computational support at ISU, and M. Dasenko and the Center for Genome Research and Biocomputing for library preparation, sequencing and computational support at OSU. We thank D. Auger for reading the manuscript.

Author Contributions

Conceptualization: Cedar Warman, Kaushik Panda, Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler.

Data curation: Sam Hokin.

Formal analysis: Cedar Warman, Kaushik Panda, Sam Hokin, Duo Jiang, John E. Fowler.

Funding acquisition: Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler.

Investigation: Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Erica Unger-Wallace, Rex A. Cole, Antony M. Chettoor, Matthew M. S. Evans, John E. Fowler.

Methodology: Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Erica Unger-Wallace, Rex A. Cole, John E. Fowler.

Project administration: Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler.

Software: Cedar Warman, Kaushik Panda, Sam Hokin.

Supervision: Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler.

Visualization: Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Sam Hokin, Matthew M. S. Evans, John E. Fowler.

Writing – original draft: Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Sam Hokin, Erica Unger-Wallace, R. Keith Slotkin, John E. Fowler.

Writing – review & editing: Cedar Warman, Kaushik Panda, Zuzana Vejlupkova, Sam Hokin, Erica Unger-Wallace, Duo Jiang, Erik Vollbrecht, Matthew M. S. Evans, R. Keith Slotkin, John E. Fowler.
References

1. Yang W-C, Shi D-Q, Chen Y-H. Female gametophyte development in flowering plants. Annu Rev Plant Biol. 2010; 61: 89–108. https://doi.org/10.1146/annurev-arplant-042809-112203 PMID: 20192738
2. Zhou L-Z, Juranić M, Dresselhaus T. Germline Development and Fertilization Mechanisms in Maize. Mol Plant. 2017; 10: 389–401. https://doi.org/10.1016/j.molp.2017.01.012 PMID: 28267957
3. McCormick S. Male Gametophyte Development. Plant Cell. 1993; 5: 1265–1275. https://doi.org/10.1105/tpc.5.10.1265 PMID: 12271026
4. Haﬁdh S, Fila J, Honys D. Male gametophyte development and function in angiosperms: a general concept. Plant Reprod. 2016; 29: 31–51. https://doi.org/10.1007/s00497-015-0272-4 PMID: 26728623
5. Dresselhaus T, Sprunck S, Wessel GM. Fertilization Mechanisms in Flowering Plants. Curr Biol. 2016; 26: R125–39. https://doi.org/10.1016/j.cub.2015.12.032 PMID: 26859271
6. Zhou L-Z, Dresselhaus T. Chapter Seventeen—Friend or foe: Signaling mechanisms during double fertilization in flowering seed plants. In: Grossniklaus U, editor. Current Topics in Developmental Biology. Academic Press; 2019. pp. 453–496. https://doi.org/10.1016/bs.ctdb.2018.11.013 PMID: 30612627
7. Lausser A, Kliwer I, Srilunchang K-O, Dresselhaus T. Sporophytic control of pollen tube growth and guidance in maize. J Exp Bot. 2010; 61: 673–682. https://doi.org/10.1093/jxb/erp330 PMID: 19926683
8. Mizukami AG, Inatsugi R, Jiao J, Kotake T, Kuwata K, Ootani K, et al. The AMOR Arabinogalactan Sugar Chain Induces Pollen-Tube Competency to Respond to Ovular Guidance. Curr Biol. 2016; 26: 1091–1097. https://doi.org/10.1016/j.cub.2016.02.040 PMID: 27068416
9. Higashiyama T, Takeuchi H. The mechanism and key molecules involved in pollen tube guidance. Annu Rev Plant Biol. 2015; 66: 393–413. https://doi.org/10.1146/annurev-arplant-043014-115635 PMID: 25621518
10. Williams JH, Reese JB. Evolution of development of pollen performance. Curr Top Dev Biol. 2019; 131: 299–336. https://doi.org/10.1016/bs.ctdb.2018.11.012 PMID: 30612621
11. Arthur KM, Vejlupkova Z, Meeley RB, Fowler JE. maize ROP2 GTPase provides a competitive advantage to the male gametophyte. Genetics. 2003; 165: 2137–2151. PMID: 14704193
12. Cole RA, Synek L, Zarsky V, Fowler JE. SEC8, a subunit of the putative Arabidopsis exocyst complex, facilitates pollen germination and competitive pollen tube growth. Plant Physiol. 2005; 138: 2005–2018. https://doi.org/10.1104/pp.105.062273 PMID: 16040664
13. Huang JT, Wang Q, Park W, Feng Y, Kumar D, Meeley R, et al. Competitive Ability of Maize Pollen Grains Requires Paralogous Serine Threonine Protein Kinases STK1 and STK2. Genetics. 2017; 207: 1361–1370. https://doi.org/10.1534/genetics.117.300358 PMID: 28986443
14. Kelliferi T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, et al. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature. 2017; 542: 105–109. https://doi.org/10.1038/nature20827 PMID: 28114299
15. Gilles LM, Khaled A, Laffaire J-B, Chaignon S, Gendrot G, Laplaige J, et al. Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J. 2017; 36: 105–109. https://doi.org/10.15252/embj.201796603 PMID: 28122843
16. Li C, Li X, Meng D, Zhong Y, Chen C, Dong X, et al. A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase NOT LIKE DAD triggers gynogenesis in maize. EMBO J. 2017; 36: 707–717. https://doi.org/10.15252/embj.201796603 PMID: 28122843
17. Zhong Y, Li C, Qi X, Jiao Y, Wang D, Wang Y, et al. Mutation of ZmDMP enhances haploid induction in maize. Nat Plants. 2019; 5: 575–580. https://doi.org/10.1038/s41477-019-0443-7 PMID: 31182848
18. Kelliferi T, Starr D, Su X, Tang G, Chen Z, Carter J, et al. One-step genome editing of elite crop germplasm during haploid induction. Nat Biotechnol. 2019; 37: 287–292. https://doi.org/10.1038/s41587-019-0038-x PMID: 30833776
19. Honys D, Twell D. Comparative analysis of the Arabidopsis pollen transcriptome. Plant Physiol. 2003; 132: 640–652. https://doi.org/10.1104/pp.103.020925 PMID: 12805594
20. Becker JD, Boavida LC, Carneiro J, Haury M, Feijó JA. Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol. 2003; 133: 713–725. https://doi.org/10.1104/pp.103.028241 PMID: 14500793
21. Steffen JG, Kang I-H, Macfarlane J, Drews GN. Identification of genes expressed in the Arabidopsis female gametophyte: Female gametophyte-expressed genes. Plant J. 2007; 51: 281–292. https://doi.org/10.1111/j.1365-313X.2007.03137.x PMID: 17559508
22. Jones-Rhoades MW, Borevitz JO, Preuss D. Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. PLoS Genet. 2007; 3: 1848–1861. https://doi.org/10.1371/journal.pgen.0030171 PMID: 17937500
23. Chettoor AM, Givan SA, Cole RA, Coker CT, Unger-Wallace E, Vejupkova Z, et al. Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. Genome Biol. 2014; 15: 414. https://doi.org/10.1186/s13059-014-0414-2 PMID: 25084966

24. Zhai J, Zhang H, Arikit S, Huang K, Nan G-L, Walbot V, et al. Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasedRNAs in maize anthers. Proc Natl Acad Sci U S A. 2015; 112: 3146–3151. https://doi.org/10.1073/pnas.1418181112 PMID: 25713378

25. Nelms B, Walbot V. Defining the developmental program leading to meiosis in maize. Science. 2019; 364: 52–56. https://doi.org/10.1126/science.aav6428 PMID: 31378720

26. Begcy K, Nosenko T, Zhou L-Z, Fragner L, Weckwerth W, Dresselhaus T. Male Sterility in Maize after Transient Heat Stress during the Tetrad Stage of Pollen Development. Plant Physiol. 2019. https://doi.org/10.1104/pp.19.00707 PMID: 31378720

27. Chen J, Strieder N, Krohn NG, Cyprys P, Sprunk S, Engelmann JC, et al. Zygotic Genome Activation Occurs Shortly after Fertilization in Maize. Plant Cell. 2017; 29: 2106–2125. https://doi.org/10.1105/tpc.17.00099 PMID: 28814645

28. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA, et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell. 2009; 136: 461–472. https://doi.org/10.1016/j.cell.2008.12.038 PMID: 19680290

29. Schott VK, Chumak N, Mosiolek M, Slusarz L, Komnenovic V, Brownfield L, et al. Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep. 2009; 10: 1015–1021. https://doi.org/10.1038/embor.2009.152 PMID: 19680290

30. Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell. 2012; 151: 194–205. https://doi.org/10.1016/j.cell.2012.09.001 PMID: 23000270

31. Dooner HK, Wang Q, Huang JT, He L, Xiong W, et al. Spontaneous mutations in maize pollen are frequent in some lines and arise mainly from retrotranspositions and deletions. Proc Natl Acad Sci U S A. 2019. https://doi.org/10.1073/pnas.1903809116 PMID: 30992374

32. He S, Vickers M, Zhang J, Feng X. Natural depletion of histone H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation. Elife. 2019; 8. https://doi.org/10.7554/eLife.42530 PMID: 31135340

33. Martinez G, Slotkin RK. Developmental relaxation of transposable element silencing in plants: functional or byproduct? Curr Opin Plant Biol. 2012; 15: 496–502. https://doi.org/10.1016/j.pbi.2012.09.001 PMID: 23022393

34. Martinez G, Panda K, Köhler C, Slotkin RK. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. Nat Plants. 2016; 2: 16030. https://doi.org/10.1038/nplants.2016.30 PMID: 27249563

35. Martinez G, Wolff P, Wang Z, Moreno-Romero J, Santos-González J, Conze LL, et al. Paternal easiRNAs regulate paternal genome dosage in Arabidopsis. Nat Genet. 2018; 50: 193–198. https://doi.org/10.1038/s41588-017-0033-4 PMID: 29335548

36. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017; 546: 524–527. https://doi.org/10.1038/nature22971 PMID: 28605751

37. Lunardon A, Forestan C, Farinati S, Axtell MJ, Varotto S. Genome-Wide Characterization of Maize Small RNA Loci and Their Regulation in the required to maintain repression6-1 (rmr6-1) Mutant and Long-Term Abiotic Stresses. Plant Physiol. 2016; 170: 1535–1548. https://doi.org/10.1104/pp.15.01205 PMID: 26747286

38. Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, Urich MA, et al. Integration of omic networks in a developmental atlas of maize. Science. 2016; 353: 814–818. https://doi.org/10.1126/science.aag1125 PMID: 27540173

39. Panda K, Ji L, Neumann DA, Daron J, Schmitz RJ, Slotkin RK. Full-length autonomous transposable elements are preferentially targeted by expression-dependent forms of RNA-directed DNA methylation. Genome Biol. 2016; 17: 170. https://doi.org/10.1186/s13059-016-1032-y PMID: 27506905

40. Wolfrugber TK, Sharma A, Schneider KL, Albert PS, Koo D-H, Shi J, et al. Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet. 2009; 5: e1000743. https://doi.org/10.1371/journal.pgen.1000743 PMID: 19956743

41. Anderson SN, Stitzer MC, Zhou P,Ross-Ibarra J, Hirsch CD, Springer NM. Dynamic Patterns of Transcript Abundance of Transposable Element Families in Maize. G3. 2019; 9: 3673–3682. https://doi.org/10.1534/g3.119.400431 PMID: 31506319
42. Li Y, Segal G, Wang Q, Dooner HK. Gene Tagging with Engineered Ds Elements in Maize. In: Petersen T, editor. Plant Transposable Elements: Methods and Protocols. Totowa, NJ: Humana Press; 2013. pp. 83–99.

43. Warman C, Fowler JE. Custom built scanner and simple image processing pipeline enables low-cost, high-throughput phenotyping of maize ears. bioRxiv. 2019. p. 780650. https://doi.org/10.1101/780650

44. Engel ML, Chaboud A, Dumas C, McCormick S. Sperm cells of Zea mays have a complex complement of mRNAs. Plant J. 2003; 34: 697–707. PMID: 12787250

45. Engel ML, Holmes-Davis R, McCormick S. Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol. 2005; 138: 2124–2133. https://doi.org/10.1104/pp.105.054213 PMID: 16055690

46. Mori T, Iigawa T, Tamiya G, Miyagishima S-Y, Berger F. Gamete attachment requires GEX2 for successful fertilization in Arabidopsis. Curr Biol. 2014; 24: 170–175. https://doi.org/10.1016/j.cub.2013.11.030 PMID: 24388850

47. Liao B-Y, Weng M-P. Unraveling the association between mRNA expressions and mutant phenotypes in a genome-wide assessment of mice. Proc Natl Acad Sci U S A. 2015; 112: 4707–4712. https://doi.org/10.1073/pnas.1415046112 PMID: 25825715

48. Xu Z, Dooner HK. The maize aberrant pollen transmission 1 gene is a SABRE/KIP homolog required for pollen tube growth. Genetics. 2006; 172: 1251–1261. https://doi.org/10.1534/genetics.105.050237 PMID: 16299389

49. Wilson-Sánchez D, Rubio-Díaz S, Muñoz-Viana R, Pérez-Pérez JM, Rover-Gil S, Ponce MR, et al. Leaf phenomics: a systematic reverse genetic screen for Arabidopsis leaf mutants. Plant J. 2014; 79: 878–891. https://doi.org/10.1111/tpj.12595 PMID: 24946828

50. Rutter MT, Murren CJ, Callahan HS, Bisner AM, Leebens-Mack J, Wolyniak MJ, et al. Distributed phenomics with the unPAK project reveals the effects of mutations. Plant J. 2019. https://doi.org/10.1111/tpj.14427 PMID: 31155775

51. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002; 418: 387–391. https://doi.org/10.1038/nature00935 PMID: 12140549

52. Berry DB, Guan Q, Hose J, Haroon S, Gebbia M, Heisler LE, et al. Multiple means to the same end: the genetic basis of acquired stress resistance in yeast. PLoS Genet. 2011; 7: e1002353. https://doi.org/10.1371/journal.pgen.1002353 PMID: 22102822

53. Price MN, Deutschbauer AM, Skerker JM, Wetmore KM, Thurs T, Mar JS, et al. Indirect and suboptimal control of gene expression is widespread in bacteria. Mol Syst Biol. 2013; 9: 660. https://doi.org/10.1038/msb.2013.16 PMID: 23591776

54. Helmann TC, Deutschbauer AM, Lindow SE. Genome-wide identification of Pseudomonas syringae genes required for fitness during colonization of the leaf surface and apoplast. Proc Natl Acad Sci U S A. 2019. https://doi.org/10.1073/pnas.1908858116 PMID: 31484768

55. Schnable JC. Genes and gene models, an important distinction. New Phytol. 2019. https://doi.org/10.1111/nph.16011 PMID: 31241760

56. Wang G, Jiang H, Del Toro de León G, Martínez G, Köhler C. Sequestration of a Transposon-Derived siRNA by a Target Mimic Imprinted Gene Induces Postzygotic Reproductive Isolation in Arabidopsis. Dev Cell. 2018; 46: 696–705.e4. https://doi.org/10.1016/j.devcel.2018.07.014 PMID: 30122632

57. Borges F, Parent J-S, van Ex F, Wolff P, Martinez G, Köhler C, et al. Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nat Genet. 2018; 50: 186–192. https://doi.org/10.1038/s41588-017-0032-5 PMID: 29335544

58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635 PMID: 23104886

59. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28: 511–515. https://doi.org/10.1038/nbt.1621 PMID: 20436464

60. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 2017; 45: W122–W129. https://doi.org/10.1093/nar/gkx382 PMID: 28472432

61. Wimalanathan K, Friedberg I, Andorf CM, Lawrence-Dill CJ. Maize GO Annotation-Methods, Evaluation, and Review (maize-GAMER). Plant Direct. 2018; 2: e00052. https://doi.org/10.1002/pld3.52 PMID: 31245718

62. Bushnell B. BBTools software package. URL http://sourceforge.net/projects/bbmap. 2014.
63. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30: 923–930. https://doi.org/10.1093/bioinformatics/btt656 PMID: 24227677

64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

65. Gao H, Smith J, Yang M, Jones S, Djukanovic V, Nicholson MG, et al. Heritable targeted mutagenesis in maize using a designed endonuclease. Plant J. 2010; 61: 176–187. https://doi.org/10.1111/j.1365-313X.2009.04041.x PMID: 19811621

66. Vejlupkova Z, Fowler JE. Maize DNA preps for undergraduate students: a robust method for PCR genotyping. Maize Genetics Cooperation Newsletter. 2003; 77: 24–25.

67. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9: 676–682. https://doi.org/10.1038/nmeth.2274 PMID: 22743772

68. Portwood JL 2nd, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, et al. MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res. 2019; 47: D1146–D1154. https://doi.org/10.1093/nar/gky1046 PMID: 30407532

69. Krishnakumar V, Hanlon MR, Contrino S, Ferlanti ES, Karamycheva S, Kim M, et al. Araport: the Arabidopsis information portal. Nucleic Acids Res. 2015; 43: D1003–9. https://doi.org/10.1093/nar/gku1200 PMID: 25414324

70. Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J. 2017; 89: 789–804. https://doi.org/10.1111/tpj.13415 PMID: 27662469

71. Mitchell AL, Atwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019; 47: D351–D360. https://doi.org/10.1093/nar/gky1100 PMID: 30398656

72. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305: 567–580. https://doi.org/10.1006/jmbi.2000.4315 PMID: 11152613

73. Van Bel M, Diels T, Vancaeste E, Kreft L, Botzki A, Van de Peer Y, et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Res. 2018; 46: D1190–D1196. https://doi.org/10.1093/nar/gkx1002 PMID: 29069403

74. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389 PMID: 9254694

75. Vollebrect E, Hake S. Deficiency analysis of female gametogenesis in maize. Dev Genet. 1995; 16: 44–63.

76. Running MP, Clark SE, Meyerowitz EM. Chapter 15 Confocal Microscopy of the Shoot Apex. In: Galbraith DW, Bohmert HJ, Bourque DP, editors. Methods in Cell Biology. Academic Press; 1995. pp. 217–229.

77. Valdivia ER, Sampedro J, Lamb JC, Chopra S, Cosgrove DJ. Recent proliferation and translocation of pollen group 1 allergen genes in the maize genome. Plant Physiol. 2007; 143: 1269–1281. https://doi.org/10.1093/pcp/pcm145 PMCID: 17220362

78. Vollbrecht E, Hake S. Deficiency analysis of female gametogenesis in maize. Dev Genet. 1995; 16: 44–63.

79. Running MP, Clark SE, Meyerowitz EM. Chapter 15 Confocal Microscopy of the Shoot Apex. In: Galbraith DW, Bohmert HJ, Bourque DP, editors. Methods in Cell Biology. Academic Press; 1995. pp. 217–229.

80. Valdivia ER, Sampedro J, Lamb JC, Chopra S, Cosgrove DJ. Recent proliferation and translocation of pollen group 1 allergen genes in the maize genome. Plant Physiol. 2007; 143: 1269–1281. https://doi.org/10.1093/pcp/pcm145 PMCID: 17220362