

RESEARCH ARTICLE

Widespread imprinting of transposable elements and variable genes in the maize endosperm

Sarah N. Anderson^{1*}, Peng Zhou², Kaitlin Higgins¹, Yaniv Brandvain², Nathan M. Springer²

1 Department of Genetics, Development, and Cell Biology; Iowa State University; Ames, Iowa, United States of America, **2** Department of Plant and Microbial Biology; University of Minnesota; St. Paul, Minnesota, United States of America

* sna@iastate.edu



OPEN ACCESS

Citation: Anderson SN, Zhou P, Higgins K, Brandvain Y, Springer NM (2021) Widespread imprinting of transposable elements and variable genes in the maize endosperm. *PLoS Genet* 17(4): e1009491. <https://doi.org/10.1371/journal.pgen.1009491>

Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, UNITED STATES

Received: December 17, 2020

Accepted: March 15, 2021

Published: April 8, 2021

Copyright: © 2021 Anderson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: RNA-seq data files have been uploaded to NCBI SRA under BioProject ID PRJNA623806. The intermediate data files used to generate all figures are available at <https://github.com/SNAnderson/Imprinting2020>. Scripts and data files used to process results are available at <https://github.com/SNAnderson/Imprinting2020> and https://github.com/kmhiggins/Imprinting_2020. There are no restrictions on data or code availability.

Abstract

Fertilization and seed development is a critical time in the plant life cycle, and coordinated development of the embryo and endosperm are required to produce a viable seed. In the endosperm, some genes show imprinted expression where transcripts are derived primarily from one parental genome. Imprinted gene expression has been observed across many flowering plant species, though only a small proportion of genes are imprinted. Understanding how imprinted expression arises has been complicated by the reliance on single nucleotide polymorphisms between alleles to enable testing for imprinting. Here, we develop a method to use whole genome assemblies of multiple genotypes to assess for imprinting of both shared and variable portions of the genome using data from reciprocal crosses. This reveals widespread maternal expression of genes and transposable elements with presence-absence variation within maize and across species. Most maternally expressed features are expressed primarily in the endosperm, suggesting that maternal de-repression in the central cell facilitates expression. Furthermore, maternally expressed TEs are enriched for maternal expression of the nearest gene, and read alignments over maternal TE-gene pairs indicate that these are fused rather than independent transcripts.

Author summary

While parents contribute equally to the DNA of offspring, some genes are imprinted, or differentially expressed depending on which parent the allele is passed from. In plants, imprinted genes are expressed in the endosperm, and prior studies have missed many imprinted genes. Here, we present a new method which identifies imprinted genes and transposable elements using all parental differences in sequence, including presence-absence variation between parents, rather than single nucleotide polymorphisms alone. We find that the majority of imprinted genes with maternal expression are genes and transposable elements that are present in only one genome and that are not shared with other grass species. These transcripts are often expressed primarily in the endosperm

Funding: This work was funded by grants from USDA-NIFA ([usda.gov](https://www.usda.gov)) 2016-67013-24747 (S.N. A., and N.M.S.), NSF ([nsf.gov](https://www.nsf.gov)) IOS-1546899 (P.Z. and N.M.S.), NSF ([nsf.gov](https://www.nsf.gov)) DGE-1545453 (K.H.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

where epigenetic marks on the maternal alleles are distinct from the rest of the plant. We also find evidence of fused transcripts with transposable elements upstream of genes, implicating common regulation for imprinted transposable elements and some nearby genes.

Introduction

Imprinted genes showing parent-of-origin based patterns of expression were first identified in maize [1] and have since been identified in a variety of flowering plants [2–6]. In plants, imprinted expression is primarily observed in the endosperm, which is a nutritive tissue of the seed that is formed when the diploid central cell is fertilized by one of the two sperm cells delivered by the pollen tube. The central cell is epigenetically distinct from most vegetative cells in the plant due to DNA demethylation targeted primarily to Transposable Elements (TEs) [7–9]. This demethylation acts as a primary imprint that distinguishes the female and the male alleles in the endosperm. Maternal and paternal alleles are further distinguished through differential accumulation of histone modifications such as H3K27me3 [10,11] which often marks the maternal allele of paternally expressed genes (PEGs) while maternally expressed genes (MEGs) often show differences in DNA methylation alone [12].

Imprinting has been studied at the genomic level in many plant species [2–6]. While some important genes have conserved imprinting in many species [13], several studies have observed variable imprinting for other genes, with inconsistent imprinting within genotypes of a single species or across species [14,15]. However, understanding the rate of turnover and the origins of imprinted expression patterns has been challenging due in part to methodological inconsistencies across studies and the limitations of available SNPs for allele calls. In Arabidopsis, applying consistent methods and cutoffs across studies reduces apparent variability in imprinting calls [16,17], however many genes cannot be assessed due to a lack of informative SNPs. A lack of SNPs can be due to identical sequence or unalignable regions resulting from large structural changes or presence-absence variation (PAV) of whole genes or features. In maize, many genes and TEs exhibit PAV among genotypes [18–20]. This limits the ability to use SNP-based allele-specific expression analyses to study imprinting, especially for transposons and variable genes. In this study, we develop an alternative approach that relies upon comparisons of expression in reciprocal crosses to assess the imprinting of both conserved and variable genes and TEs across maize genotypes with whole genome assemblies, revealing imprinting for many transposable elements and variable genic sequences.

Results and discussion

Reciprocal crosses for every pairwise contrast between three maize genotypes with whole genome assemblies (B73 [21], W22 [22], and PH207 [23]) were performed, and 14 days after pollination, endosperm was isolated in triplicate for RNA-sequencing (S1 Table). Two approaches were applied to identify imprinted expression (Fig 1A). The traditional approach for calling imprinting uses Single Nucleotide Polymorphisms to call Allele Specific Expression (SNP-ASE) followed by comparison of biases across reciprocal crosses (methods). The SNP-ASE ratio is calculated by assigning SNP-containing reads to one allele and determining the proportion of informative reads from each allele, providing an estimate of the expression of two alleles within a single sample. This approach can be applied to a single sample and the assessment of imprinting is typically based on consistent bias for the SNP-ASE value in both reciprocal crosses. We developed and implemented an alternative approach where reads are

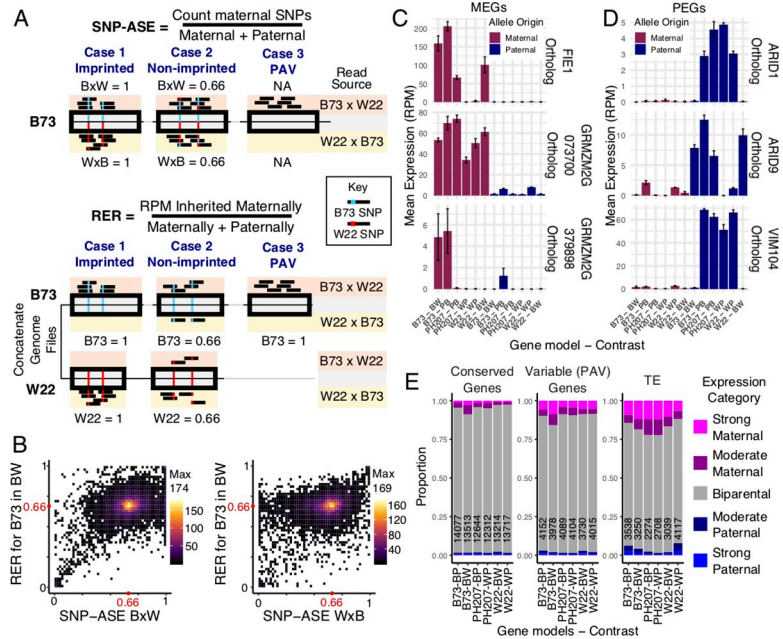


Fig 1. Assessing imprinted expression pattern in maize. A) The method for defining imprinting using SNP-ASE versus RER illustrating an example of an imprinted gene distinguishable by SNPs (Case 1), a non-imprinted gene (Case 2), and an imprinted PAV gene (Case 3). In SNP-ASE, reads mapping to a SNP-corrected reference genome are assigned to alleles based on the SNP supported. In RER, reads are assigned to a concatenated reference genome and retained at unique positions. Both methods can be used to assess imprinting for shared genes distinguished by SNPs, but only RER can assess imprinting for PAV features. B) Comparison of SNP-ASE and RER for B73 alleles assessable with at least 10 unique (RER) and informative (SNP-ASE) reads in the B73 x W22 crosses. The values plotted show the averages across three biological replicates. SNP-ASE is calculated separately for the B73 x W22 cross (left) and reciprocal W22 x B73 cross (right), while RER is calculated with information from both reciprocals and plotted as the y-axis in both plots. The heat represents the number of genes in each pixel of the plot, and red dots indicate the expected ratios for non-imprinted expressed genes. C-D) Expression patterns of known MEGs (C) and PEGs (D) in this dataset using mapping from concatenated genomes. X-axis shows the source of the gene model and the genome contrast (where B = B73, W = W22, and P = PH207), and bars are colored based on whether the allele plotted is inherited maternally (maroon) or paternally (navy). The y-axis shows the mean expression calculated from unique mapping reads using the mapping strategy used for RER. RER values are calculated by dividing the expression when inherited maternally by the sum of expression from both directions of the cross. Error bars represent standard error, and gene IDs for the orthologs in each genome are listed in S2 Table. E) The distribution of RER values for different features across contrasts. RER cutoffs for strong maternal and strong paternal are > 0.9 and < 0.1, respectively, and cutoffs for moderate maternal and paternal are > 0.8 or < 0.2, respectively. The total number of features across all expression categories are shown on each bar.

<https://doi.org/10.1371/journal.pgen.1009491.g001>

aligned to concatenated genome files composed of the two parental genotypes and the Reciprocal Expression Ratio (RER) was calculated to describe the ratio of expression for features in each genome when inherited maternally versus paternally. Unlike SNP-ASE, the RER is a comparison of expression of a feature in reciprocal crosses and cannot be calculated for a single sample. Calculations of RER rely on the ~15% of reads that map uniquely to a single location in the concatenated genomes (S1 Table). While many reads map equally well to both genomes and are therefore discarded, unique mapping reads are only found in places of the genome with variants distinguishing the alleles (SNPs or indels) or in regions unique to one genome. After assigning unique reads to features including genes and TEs using HTseq, RER was calculated by dividing the expression level (RPM) when inherited maternally by the sum of expression when maternally or paternally inherited. Given that endosperm is composed of two copies of the maternal genome and one copy of the paternal genome, the null expectation for a transcript's expression is that it will be twice as highly expressed when inherited from the

maternal parent compared to the paternal parent. For both SNP-ASE and RER, the average value representing a biparentally expressed gene is 0.67, allowing direct comparison of the methods. A comparison of SNP-ASE and RER reveals general agreement between these two approaches for genes that could be analyzed with SNPs, with the majority of genes expressed at the ratio expected by 2:1 dosage (Fig 1B). Many of the genes showing disagreement between methods in Fig 1B result from genotype-biased expression which exhibits a strong bias in SNP-ASE for a single sample but doesn't result in bias for RER (S1 Fig). Inaccuracies in SNP-ASE can also arise from erroneous mapping of paralogs when mapping to a single reference genome. To further assess accuracy of RER, expression patterns for three MEGs and three PEGs with conserved imprinting status in maize, rice, and Arabidopsis [4] were assessed (Fig 1C, S2 Table). In most cases with informative reads, clear parental bias in the expected direction was observed for all genes (Fig 1C).

While both methods can be used to define imprinting for shared genes distinguishable by SNPs, only the RER method can capture imprinting for portions of the genome that exhibit PAV. This provides new opportunities to study parent-of-origin biased gene expression for TEs and variable genes. The distribution of RER values was assessed across contrasts for different feature types (S2A Fig), and the proportion of each set that showed parentally-biased expression was summarized based on RER (Figs 1E, S2B and S2C). This revealed that across all contrasts, genes conserved within maize rarely exhibit parent-of-origin biased expression (Figs 1E and S2). On average, < 3% of expressed genes that are present in all three maize genotypes in this study show a strong parental bias (Fig 1E). For genes that exhibit presence/absence variation among maize lines, a higher proportion (> 6%) of expressed genes show high parental bias, with this set representing genes that are accessible using RER but not SNP-ASE. Strikingly, > 11% of expressed TEs show a strong parental bias, with the majority of strongly biased TEs expressed maternally (Fig 1E).

In order to identify imprinted transcripts, we applied the `lfcThreshold` option within DESeq2 to test for significance (adjusted p-value < 0.05) over the expected 2:1 gene dosage across reciprocals using three biological replicates. To increase the stringency of imprinting calls, significant hits were further filtered by RER values. Maternally Expressed Genes (MEGs) and Maternally Expressed TEs (matTEs) were filtered for RER > 0.9, while Paternally Expressed Genes (PEGs) were filtered for RER < 0.1. It can be difficult to remove all maternal tissues when isolating endosperm tissue and therefore it is important to limit potential false-positive calls of maternal expression that may result from genes expressed in the maternal seed coat [24]. Previously published RNA-seq data [25] was used to filter out genes whose maternal expression could result from seed coat contamination rather than maternal expression in the endosperm. Pericarp-preferred genes were defined where the mean expression in pericarp was >2-fold higher than the expression in endosperm (S3 Fig). After implementing these criteria and filters, we identified an average of 182 total imprinted genes across all hybrid combinations, with an average of 112 MEGs and 70 PEGs in each (Fig 2A, S1–S3 Data).

Since RER considers alleles from each genotype independently, comparative genomic approaches were used to compare the consistency of imprinting for genes discovered in each genome. A comparison of imprinted features in the B73 x W22 reciprocal hybrid endosperm tissue identifies 17 MEGs, 39 PEGs, and 4 matTEs that were consistently imprinted in both genomes (Figs 2B and 3B). A subset of the genes that do not exhibit consistent imprinting are shared between the two genomes. For example, there are 26 MEGs observed only in B73 and 11 only observed in W22 despite the fact that both genomes retain a syntenic ortholog for these genes. For the majority of these shared genes with variable imprinting, the lack of overlap is due to cutoff stringency or lack of coverage rather than true gain or loss of imprinting (S4 Fig). There are many additional cases where imprinted genes are only present in one genome.

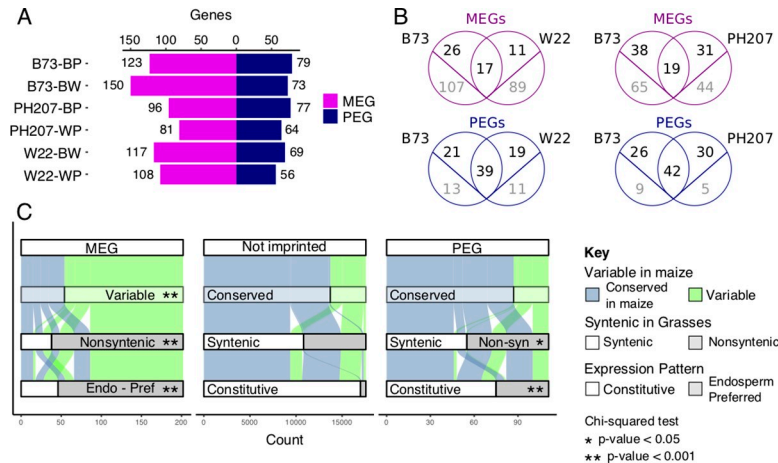


Fig 2. Imprinting of genes defined by RER. A) The number of imprinted genes identified across contrasts using the RER method (see [Materials and Methods](#)). MEGs are shown in magenta and PEGs are shown in blue. B) The overlap between imprinted genes across pairwise contrasts. Genes that are shared between genotypes that could be assessed for imprinting are shown in black above the line while imprinted genes unique to one genome are shown in gray below the line. C) Comparison of features for MEGs, PEGs, and non-imprinted B73 genes. Genes are defined as conserved when they are shared with all genotypes present in this study, syntenic when a syntenic ortholog exists in sorghum, rice, foxtail millet, and brachypodium, and endosperm-preferred if expression is primarily restricted to the endosperm ([S5 Fig](#)). Asterisks denote significance relative to the Not Imprinted set (chi-squared test).

<https://doi.org/10.1371/journal.pgen.1009491.g002>

For PEGs, variable genes represent the minority of non-conserved imprinted genes, with only 13 of 34 B73 PEGs that are not imprinted in W22 variable across genomes. In contrast, for the majority of MEGs with inconsistent imprinting (i.e. 107 of 133 B73 genes in the B73 by W22 contrast), the genes themselves are absent from the other genome. Similar patterns are observed for the B73 by PH207 contrast, though a higher proportion of genes are shared in this contrast. The large number of maternally expressed transcripts with variability in maize suggests that imprinting of non-conserved elements may be far more prevalent than previously detected due to the limitations of SNP-based allele calls.

To understand additional features of imprinted genes, we focused on the B73 genes that were called imprinted in at least one contrast, which included 202 MEGs and 111 PEGs. B73 was selected as the central genotype because it has substantially more expression datasets, syntenic gene information, and functional gene annotations than other genomes. For the genes identified as imprinted, we compared several characteristics relative to genes that were expressed but were not classified as imprinted ([S4 Data](#)). First, genes were assessed for variability across maize inbred lines by defining conserved genes as those with syntenic orthologs in B73, W22, and PH207 and variable genes as those without a corresponding gene in at least one genome [26]. This revealed a clear enrichment for variable genes among MEGs (p-value < 0.001, chisq test), but not PEGs, compared to genes that are not imprinted but have enough unique reads to be assessed for imprinting ([Fig 2C](#)). We then expanded our evolutionary distance and assessed how many genes in each set are syntenic with other grasses as defined by having a syntenic ortholog in sorghum, rice, foxtail millet, and brachypodium. For genes without imprinting, the majority (62%) are syntenic with other grasses. However, MEGs are highly depleted for syntenic genes (19%) and PEGs show a minor depletion (50%, p-value < 0.05, chisq test). Next, the expression pattern across B73 development was assessed using published RNA-seq data [25]. Since imprinting can arise from either silencing of one parental allele specifically in the endosperm or de-repression of one parental allele in the endosperm, the pattern of expression across tissues was defined as either constitutive or

endosperm-preferred (see [Materials and Methods](#), [S4 Fig](#)). While only 3% of non-imprinted genes are expressed preferentially in the endosperm, 77% of MEGs and 32% of PEGs show this expression pattern ([Fig 2C](#), p -value < 0.001, chisq test). Many of the MEGs (38%) have no assigned GO term, a 2.8-fold enrichment compared to genes that are not imprinted (p -value < 0.001, chisq test). Since TEs are a common source of new genes and a driver of gene content variation among maize lines, we intersected our imprinted genes with annotated TEs, identifying 26 MEGs and 1 PEG completely within an annotated transposable element. While MEGs and PEGs are annotated as genes in the B73v4 annotation, transcription of a locus does not imply the production of a functional gene product and many “genes” annotated in B73 have TE origins. Evolutionarily conserved genes with synteny to other grasses may be the best candidates for functional genes capable of conferring phenotypes [27], there are certainly some examples of non-conserved genes that can be important for functions such as disease resistance [28].

To further investigate the imprinting of TEs themselves, the RER method was used to define imprinted TEs, with an average of 95 matTEs identified across contrasts ([Fig 3A](#)). There are a very small number of paternally expressed TEs, however these were excluded from further analyses due to the low number detected and potential technical complications ([Figs 3A and S2](#)). Consistent with the large amount of TE variability among genotypes, the majority of imprinted TEs were unique to one genome ([Fig 3B](#)). There are 145 maternally expressed TEs in B73 relative to at least one other genotype, including 72 LTR retrotransposons, 52 Helitrons, 9 TIR transposons, and 2 LINEs ([Fig 3C](#)). The vast majority of these TEs (93%) represent specific TE insertions that are polymorphic among the three maize genotypes [20]. Given the high tissue-specificity of TE expression observed previously [29], the tissue-specific expression

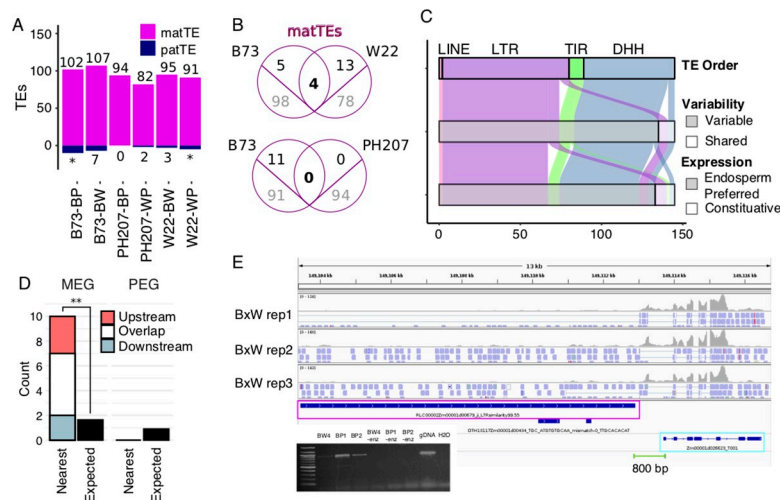


Fig 3. Imprinted TEs defined by RER. A) The number of imprinted TEs across contrasts. matTEs are marked in magenta and paternally expressed TEs are marked in navy. Asterisks denote contrasts where paternally expressed TEs could not be defined ([S2 Fig](#)). B) The overlap between matTEs across pairwise contrasts. TEs that are shared between genotypes that could be assessed for imprinting are shown in black above the line while imprinted TEs unique to one genome are shown in gray below the line. C. Features of matTEs in B73. TE orders are abbreviated: DHH = Helitron, TIR = terminal inverted repeat transposon, LTR = LTR retrotransposon, and LINE = long interspersed nuclear element. TE variability is defined by prior work (Anderson et al. 2019). Endosperm-preferred expression is described by patterns across development ([S5 Fig](#)). D) Imprinting status of closest gene to matTEs. Expected number is based on the number of MEGs and PEGs that were assessed for imprinting. ** p -value < 0.001 (binomial test) E) IGV view showing a representative example of reads aligning to a matTE near a MEG. Reads are colored by the strand of alignments, where blue = forward strand. Green line shows primer positions for RT-PCR, with positive result shown in gel image.

<https://doi.org/10.1371/journal.pgen.1009491.g003>

patterns for matTEs were also assessed. We found that 92% of matTEs are expressed preferentially in the endosperm, suggesting that imprinting is established through de-repression of the maternal allele preferentially in the endosperm and that this is the only stage of development for expression of these elements (Figs 3C and S5). We also assessed some features of these matTEs, finding that matTEs are closer to the nearest gene than non-imprinted genes (p-value < 0.0001, t-test), with a mean distance of 15kb for matTEs vs. 35 kb for non-imprinted genes. For LTR retrotransposons, we are able to confidently assess TE age using LTR similarity as a proxy, where younger elements have more similar LTRs. We find that maternal LTRs are on average older than non-imprinted LTRs (p-value < 0.0001, t-test), with an average of 93% and 95% similarity, respectively.

Since TE families have the potential for coordinated expression responses among members, the families for matTEs were assessed. matTEs are in 84 families, with only one Helitron family containing more than 5 imprinted elements. This family, DHH00002 (DHH2), contains 44 maternally expressed members and is the only Helitron family in B73 that is predicted to have autonomous members. Since prior work has suggested that Helitrons are responsible for creating imprinting by moving PHE1 binding sites around the genome [30], the proportion of DHH2 Helitrons with predicted PHE1 motifs was assessed (S6 Fig). We found that matTEs of this family more frequently have a binding site than elements that are not detected in our analysis (p-value < 0.001, chi square test). However, the proportion with binding sites is similar for imprinted and non-imprinted expressed family members (p-value = 0.38, chi square test), so it is unlikely that PHE1 sites alone are sufficient to confer imprinting of DHH2 Helitrons.

TEs have been proposed as a source of imprinted gene expression [7,31] and coordinate expression of genes and TEs has been observed in other reproductive tissues [32], so we investigated the relationship between imprinted genes and nearby TEs. For every matTE in B73, the closest gene was identified and assessed for imprinting. For 13% of matTEs, the nearest gene is a MEG, which is a significant enrichment (p-value < 0.001, binomial test) and 11.6 times more common than expected based on the proportion of expressed genes that are called MEGs (Fig 3D). In contrast, there were no identified examples of matTEs where the closest gene is a PEG. There were 19 matTEs where the closest gene is one of 15 MEGs (S3 Table). Since TE and gene annotations can overlap and some annotated genes code for TE proteins, we carefully inspected these loci to understand the relationship between these annotated features. In all but one case, the genes overlap (N = 7) or are downstream of the nearest TE (N = 7). In addition to the seven genes overlapping TEs, two downstream genes (Zm00001d046395 and Zm00001d041755) have helicase domains consistent with helitron origin. None of the 15 genes have syntenic orthologs in another grass species, and most (N = 13) exhibit variable presence within maize genotypes. To understand the nature of transcripts, read alignments for matTE-MEG pairs were visualized with IGV. In all cases, reads aligning to both the matTE and corresponding MEG mapped to the same strand without clear separation in read alignments, suggesting that many of these clusters may actually represent single transcripts overlapping multiple features (Figs 3E and S7) rather than clusters of independently imprinted transcripts [33]. To test this experimentally, we performed RT-PCR using primers aligning to the gene and TE, confirming fused transcripts for 4 of 5 loci tested (Figs 3E and S7). Combined with the stranded RNA-seq confirming that the TEs are upstream of the genes within the transcripts, this suggests that some imprinted promoters can drive expression of transcripts overlapping features annotated as TEs and genes.

In summary, we developed the RER method to use information from shared and variable portions of maize whole genome assemblies to identify imprinted expression of genes and TEs in maize. This revealed imprinting of many genes that were undetectable by traditional methods that rely on diagnostic SNPs between parental alleles. The majority of maternally expressed

features (genes and TEs) represent young portions of the genome that are variable within maize and non-syntenic with other grasses. It is worth noting that while matTEs are not necessarily younger than non-imprinted TEs, all TE insertions accumulated after the divergence of maize from other grasses. We also observe strong enrichment for MEGs near maternally expressed TEs, suggesting a connection between imprinted TEs and transcripts containing open reading frames. In particular, this may suggest that polymorphic TE insertions can generate novel imprinted expression patterns for nearby sequences, highlighting a potential mechanism for the birth of imprinting. Notably this could occur for an existing gene or could result in imprinting of a novel transposed sequence resulting in a novel “gene.” The observation that many MEGs and matTEs are sequences that have variable presence in different maize genomes is interesting. Imprinting has been proposed to be a phenomena intended to silence TEs in the germline, and maternal expression of TEs and TE-like genes is perhaps an unsurprising outcome of this study. While it is tempting to speculate that the variability in sequence content itself could imply a mechanistic tie between variable sequence and imprinted expression, further studies would be required to better understand the nature of these transcripts and potential cross-talk between parental genomes to create imprinting in a subset of cases.

We previously hypothesized that TE demethylation could result in either MEGs or PEGs [34], however our results suggest that the majority of imprinted TEs are MEGs. The mechanistic bias for this enrichment of MEG TEs is unclear. We speculate that this could arise as a consequence of many PEGs requiring H3K27me3 to silence the demethylated maternal alleles [11,12,35]. In general, H3K27me3 is rarely found over TEs [10,36] and most TEs may lack the sequences necessary for recruitment of H3K27me3 even when they lose DNA methylation. In mammals, imprinting in the placenta has been proposed to result from different defense mechanisms used by male and female germlines to reduce retrovirus proliferation in the germ line [37], and turnover of imprinting could have similar host defense explanations in plant endosperm and animal development [38]. In plants, there are genes with conserved imprinting across plant species that support theories of parental conflict [39] or dosage [40], however the majority of imprinted loci are variable within and across species. By studying imprinting using whole genome assemblies, we are able to better understand the imprinted expression of both shared and variable portions of the maize genome.

Materials and methods

Materials

Three maize inbred lines, B73, W22, and PH207, were grown in the field in Saint Paul, MN in the summer of 2018. These genotypes are available through the US National Germplasm System (GRIN). Reciprocal crosses between each pair of genotypes were performed. Ears were collected 14 days after pollination and endosperm was isolated using manual dissection, with approximately 10 kernels per ear pooled for each biological replicate. Paired-end, stranded RNA-seq libraries were created using the Illumina TruSeq Stranded mRNA kit and PE125 sequencing was performed on the Illumina HiSeq 2500 at the University of Minnesota Genomics Center. On average, > 45 million reads were generated per library (S1 Table).

Sequence alignments for RER

Concatenated genome files were created for each pairwise contrast of parental genomes and assemblies used included B73v4 [21], W22 [22], and PH207 [23]. When necessary, chromosome designations were altered to ensure non-redundant sequence names across parents. Hisat2 index files were created using genome sequences only for each contrast. Gene annotations and disjointed filtered TE annotations available at <https://github.com/SNAnderson/>

[maizeTE_variation](#) were combined by first subtracting exon regions from the TE annotations and then combining full gene and TE annotations for each genome. This resulted in a file where reads aligned to exon regions of gene annotations are assigned to genes while reads aligned to TE-containing introns are assigned to TEs. Ambiguous reads are not counted to either feature. Concatenated annotation files were then created for each pairwise contrast using the same chromosomal designation as for the genome files. RNA-seq reads were trimmed using cutadapt [41] and aligned to the concatenated genomes corresponding to the parents using hisat2, using default parameters [42]. Unique-mapping reads to the concatenated genome files were then assigned to features (genes and TEs) using HTseq [43]. Counts to each feature were normalized as reads per million using library size estimates derived from the SNP-ASE method (described below). RER for each annotation (gene and TE) was calculated by dividing the mean expression when inherited maternally by the sum of the expression when inherited maternally and paternally. RER itself does not include any information about the corresponding locus in other genomes.

Sequence alignments for SNP-ASE

In parallel to the above method of mapping reads, we also ran the standard, SNP-based allele specific expression pipeline by mapping reads to the B73 AGPv4 reference assembly using a variant-aware aligner HiSat2 trained with a set of known SNPs as described in [44]. The number of reads supporting each parental genotype were used to calculate the proportion of maternal reads for each gene. For comparison across mapping methods, genes were filtered for only those with at least 10 informative reads in both methods. SNP-ASE ratios were calculated for each gene in each direction of the reciprocal cross separately by dividing the number of reads matching the maternal allele by the total number of informative reads. Genes with parent-specific expression were defined as those with a SNP maternal ratio > 0.85 in one direction and < 0.15 in the reciprocal direction.

Defining imprinting

To define imprinted features using RER, count tables for genes and TEs in each library were loaded into R. For each of the three reciprocal crosses performed in triplicate, DESeq2 [45] was applied using the `lfcThreshold = 1` and `altHypothesis = "greaterAbs"` options to identify features with significant deviations from the 2:1 expected expression difference based on dosage. Each contrast includes features from both parental genomes, so maternal and paternal expression was determined by the direction of the differential expression plus the genome where the feature was annotated. Significant features were further filtered to only strong cases of imprinting where RER was > 0.9 for MEGs and `matTEs` and < 0.1 for PEGs. To create the final list of imprinted features, maternal features with pericarp-preferred expression were filtered out (see Tissue Dynamics).

Tissue dynamics

The expression profile of genes and TEs was analyzed for B73 features using previously published analysis [29] using data from [25]. To filter out genes where expression is higher in the pericarp than the endosperm and could thus result in inaccurate imprinting calls [24], expression was compared for 14 dap seeds (the time point used in this study) and 18 dap pericarp. Genes with expression over twice as high in the pericarp over the endosperm were excluded from MEG calls. W22 and PH207 genes corresponding to genes expressed higher in the pericarp were also excluded from MEG calls. No `matTEs` were identified as potential contaminants using this method. Expression data across all tissues was also used to identify endosperm-

preferred expression. Endosperm-preferred expression was defined as genes and TEs where the sum of expression in endosperm and whole seed libraries (26% of libraries) was more than 60% of the sum of expression across all libraries.

Descriptors

To identify genes that are shared between genome assemblies and annotations, the file `gene_model_xref_v4.txt` was downloaded from MaizeGDB [26] on 2020/01/22. This file is B73-based and genes with a single corresponding gene in either the pairwise contrast (for venn diagrams) or in both W22 and PH207 (all other analyses) were defined as conserved in maize while remaining genes were defined as variable. This file was also used to define genes that are syntenic with other grasses, with syntenic genes being defined as any gene with a syntenic ortholog in foxtail millet, rice, brachypodium, and sorghum. To identify the nearest gene to each matTE, bedtools closest was used and distances between TE and gene were reported relative to the orientation of the gene.

RT-PCR

RT-PCR was used to identify cases where TE and gene annotations contribute to a fused imprinted transcript. RNA from B73 x PH207 replicates 1 and 2 along with unsequenced replicate 4 of B73 x W22 was cleaned with the Monarch RNA cleanup kit (T2030L, NEB) and converted into cDNA using the ProtoScript First Strand cDNA synthesis kit using oligo-dT primers (E6300S, NEB). RT-PCR was performed using primers aligning to both the TE and the gene. These sequences were: TGGACATCTTACATTGCTCCAC and CGGGCTTCAAACCAAAGA for DHH00002Zm00001d02687, GTTGATCTCTGGAACACCAACA and ATGCCCTTGTGCACCTAGTAGT for DHH00002Zm00001d02512, ATACAGCGTGACATTCATTTGC and TATTGAAGTTGGCAGGAAAGGT for RLX11772Zm00001d00001, AATTTCAGTTTCGCCCTATTCA and CCACTGAGCTCCCTCAGTATAAAA for RLC00002Zm00001d00679, and AGCTTTCTCCTCCCTTCCTCTA and CGCAGTATTCATCGTCATCATC for DHH00002Zm00001d01432.

Code availability

Scripts and data files used to process results are available at <https://github.com/SNAnderson/Imprinting2020> and https://github.com/kmhiggins/Imprinting_2020. There are no restrictions on code availability.

Supporting information

S1 Fig. Comparison of SNP-ASE and RER for B73 genes accessible using both methods in the B73 x W22 cross. A) The SNP-ASE method applied across both directions of reciprocal crosses, with points colored based on imprinting calls using RER (magenta and blue) or inconsistency across reciprocals using SNP-ASE (green and purple). Colors are consistent across panels. B-C) Comparison of the two SNP-ASE values to RER. Plots are the same as shown in Fig 1B but colored to indicate imprinted genes and genes with genotype-biased expression using SNP-ASE. Genotype bias does not impact RER since the calculation is performed across reciprocals instead of across genotypes. For all panels, values plotted show the average across three biological replicates. Genes are color-coded based on expression pattern, with magenta = MEG, blue = PEG, green = B73 biased, and purple = W22 biased. Genotype bias was defined by SNP-ASE ratios > 0.85 in one direction and < 0.15 in the other direction of

reciprocal crosses.

(TIFF)

S2 Fig. Distribution of RER for genes and TEs across contrasts. A) Most genes and TEs are expressed near the expected ratio given genome dosage (horizontal line, 0.67), with maternal and paternal expression approaching 1 or 0, respectively. TEs have a more pronounced peak of maternal expression than genes. For TEs, a bimodal peak near 0.67 and 0.33 in contrasts with PH207 suggests that some PH207 TEs may be mapping better to B73 or W22 assemblies due to poor assembly quality of PH207 in intergenic space. For this reason, paternal TEs were not counted in contrasts with PH207 in Fig 3A. B-C) The proportion of parentally biased RER values for different features across contrasts. RER cutoffs for strong maternal and strong paternal are > 0.9 and < 0.1 , respectively, and cutoffs for moderate maternal and paternal are > 0.8 or < 0.2 , respectively.

(TIFF)

S3 Fig. Expression profile of B73 MEGs and matTEs in the endosperm and pericarp using RNA-seq data from Stelpflug et al. Genes with mean expression in the pericarp $> 2x$ the mean expression in the endosperm were filtered from the MEG list due to the potential for seed coat contamination. W22 and PH207 genes corresponding to B73 pericarp-preferred genes were also removed from MEG counts. There were no matTEs filtered out using this method. Heat of each pixel represents the expression value compared to the max in the row.

(TIFF)

S4 Fig. RER bias for genes with inconsistent imprinting in Fig 2B. For both MEGs (right) and PEGs (left), the majority of genes do not overlap due to low coverage or have RER values in the same direction as imprinted genes but failed to meet our strict statistical and/or RER threshold. X-axis labels denote the genotype where the gene is imprinted, the type of imprint, and the cross where the imprinting is variable (B = B73, W = W22, P = PH207).

(TIFF)

S5 Fig. Expression pattern of imprinted B73 genes and TEs across development using data from Stelpflug et al. Endosperm and seed tissues are shown on the left side of the break, and all other tissues sampled are shown on the right of the break. Endosperm-preferred expression was defined where the sum of the expression across endosperm and seed libraries were more than 60% of the sum of the expression across all libraries. For each plot, endosperm-preferred features are shown above the break and constitutive features are shown below the break. Heat of each pixel represents the expression value compared to the max in the row.

(TIFF)

S6 Fig. Presence of PHE1 binding sites within DHH2 family helitrons. Binding motifs from Batista et al. 2019 were identified in all DHH2 helitrons, and the distribution of members with zero, one, or both of the sites were determined. A similar distribution of motifs were found for matTEs and non-imprinted family members, with a lower proportion of elements with at least one motif identified in the not detected set. TEs in the not detected set include TEs that are not expressed and TEs without unique sequence that could be assessed for imprinting.

(TIFF)

S7 Fig. IGV views showing examples of reads aligning to matTEs (magenta) near or overlapping MEGs (cyan). Reads are colored by the strand of alignments, where blue = forward strand and red = reverse strand. Green bars show primer positions for the RT-PCR amplifications shown below each example.

(TIFF)

S1 Table. Reads and mapping statistics for RNA-seq libraries in this study.

(PDF)

S2 Table. Gene IDs for conserved imprinted genes plotted in Fig 1C.

(PDF)

S3 Table. IDs and features of matTEs near MEGs.

(PDF)

S1 Data. RER output table for B73 and W22 contrast.

(XLSX)

S2 Data. RER output table for B73 and PH207 contrast.

(XLSX)

S3 Data. RER output table for W22 and PH207 contrast.

(XLSX)

S4 Data. Summary table for features of B73 genes and TEs used in this study.

(XLSX)

Acknowledgments

We thank Peter Hermanson for technical assistance. This work was supported by the Minnesota Agricultural Experiment Station, and the Minnesota Supercomputing Institute (MSI) at the University of Minnesota provided computational resources that contributed to this research.

Author Contributions

Conceptualization: Sarah N. Anderson, Yaniv Brandvain, Nathan M. Springer.

Data curation: Sarah N. Anderson.

Formal analysis: Sarah N. Anderson, Peng Zhou, Kaitlin Higgins.

Funding acquisition: Nathan M. Springer.

Methodology: Sarah N. Anderson, Peng Zhou.

Visualization: Sarah N. Anderson, Kaitlin Higgins.

Writing – original draft: Sarah N. Anderson, Nathan M. Springer.

Writing – review & editing: Sarah N. Anderson, Peng Zhou, Kaitlin Higgins, Yaniv Brandvain, Nathan M. Springer.

References

1. Kermicle JL. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics*. 1970; 66: 69–85. PMID: [17248508](https://pubmed.ncbi.nlm.nih.gov/17248508/)
2. Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, et al. Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A*. 2011; 108: 1755–1762. <https://doi.org/10.1073/pnas.1019273108> PMID: [21257907](https://pubmed.ncbi.nlm.nih.gov/21257907/)
3. Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, et al. A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet*. 2011; 7: e1002125. <https://doi.org/10.1371/journal.pgen.1002125> PMID: [21731498](https://pubmed.ncbi.nlm.nih.gov/21731498/)

4. Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W, et al. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell*. 2011; 23: 4221–4233. <https://doi.org/10.1105/tpc.111.092668> PMID: 22198147
5. Zhang M, Li N, He W, Zhang H, Yang W, Liu B. Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J*. 2016; 85: 424–436. <https://doi.org/10.1111/tpj.13116> PMID: 26718755
6. Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C. Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. *Plant Cell*. 2016; 28: 1815–1827. <https://doi.org/10.1105/tpc.16.00304> PMID: 27465027
7. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*. 2009; 324: 1447–1451. <https://doi.org/10.1126/science.1171609> PMID: 19520961
8. Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*. 2012; 337: 1360–1364. <https://doi.org/10.1126/science.1224839> PMID: 22984074
9. Park K, Kim MY, Vickers M, Park J-S, Hyun Y, Okamoto T, et al. DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc Natl Acad Sci U S A*. 2016; 113: 15138–15143. <https://doi.org/10.1073/pnas.1619047114> PMID: 27956642
10. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet*. 2010; 6. <https://doi.org/10.1371/journal.pgen.1001152> PMID: 20949070
11. Moreno-Romero J, Jiang H, Santos-González J, Köhler C. Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *EMBO J*. 2016; 35: 1298–1311. <https://doi.org/10.15252/embj.201593534> PMID: 27113256
12. Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, et al. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res*. 2014; 24: 167–176. <https://doi.org/10.1101/gr.155879.113> PMID: 24131563
13. Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A. Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci U S A*. 2000; 97: 10637–10642. <https://doi.org/10.1073/pnas.170292997> PMID: 10962025
14. Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, et al. Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc Natl Acad Sci U S A*. 2013; 110: 19639–19644. <https://doi.org/10.1073/pnas.1309182110> PMID: 24218619
15. Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife*. 2014; 3: e03198. <https://doi.org/10.7554/eLife.03198> PMID: 24994762
16. Wyder S, Raissig MT, Grossniklaus U. Consistent Reanalysis of Genome-wide Imprinting Studies in Plants Using Generalized Linear Models Increases Concordance across Datasets. *Sci Rep*. 2019; 9: 1320. <https://doi.org/10.1038/s41598-018-36768-4> PMID: 30718537
17. Picard CL, Gehring M. Identification and Comparison of Imprinted Genes Across Plant Species. *Methods Mol Biol*. 2020; 2093: 173–201. https://doi.org/10.1007/978-1-0716-0179-2_13 PMID: 32088897
18. Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet*. 2009; 5: e1000734. <https://doi.org/10.1371/journal.pgen.1000734> PMID: 19956538
19. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, et al. Insights into the maize pan-genome and pan-transcriptome. *Plant Cell*. 2014; 26: 121–135. <https://doi.org/10.1105/tpc.113.119982> PMID: 24488960
20. Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, Hirsch CD, et al. Transposable elements contribute to dynamic genome content in maize. *The Plant Journal*. 2019. <https://doi.org/10.1111/tpj.14489> PMID: 31381222
21. 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
22. Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F, Barad O, et al. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat Genet*. 2018. <https://doi.org/10.1038/s41588-018-0158-0> PMID: 30061736
23. Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, et al. Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize. *Plant Cell*. 2016; 28: 2700–2714. <https://doi.org/10.1105/tpc.16.00353> PMID: 27803309

24. Schon MA, Nodine MD. Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets. *Plant Cell*. 2017; 29: 608–617. <https://doi.org/10.1105/tpc.16.00845> PMID: 28314828
25. Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, et al. An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. *Plant Genome*. 2016; 9. <https://doi.org/10.3835/plantgenome2015.04.0025> PMID: 27898762
26. 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
27. Schnable JC. Genome evolution in maize: from genomes back to genes. *Annu Rev Plant Biol*. 2015; 66: 329–343. <https://doi.org/10.1146/annurev-arplant-043014-115604> PMID: 25494463
28. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol*. 2011; 30: 105–111. <https://doi.org/10.1038/nbt.2050> PMID: 22158310
29. 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
30. Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, et al. The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *Elife*. 2019; 8. <https://doi.org/10.7554/eLife.50541> PMID: 31789592
31. Fujimoto R, Kinoshita Y, Kawabe A, Kinoshita T, Takashima K, Nordborg M, et al. Evolution and control of imprinted FWA genes in the genus Arabidopsis. *PLoS Genet*. 2008; 4: e1000048. <https://doi.org/10.1371/journal.pgen.1000048> PMID: 18389059
32. Warman C, Panda K, Vejilupkova Z, Hokin S, Unger-Wallace E, Cole RA, et al. High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements. *PLoS Genet*. 2020; 16: e1008462. <https://doi.org/10.1371/journal.pgen.1008462> PMID: 32236090
33. Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, et al. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A*. 2011; 108: 20042–20047. <https://doi.org/10.1073/pnas.1112186108> PMID: 22114195
34. Anderson SN, Springer NM. Potential roles for transposable elements in creating imprinted expression. *Curr Opin Genet Dev*. 2018; 49: 8–14. <https://doi.org/10.1016/j.gde.2018.01.008> PMID: 29453082
35. Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, et al. High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. *PLoS Genet*. 2011; 7: e1002126. <https://doi.org/10.1371/journal.pgen.1002126> PMID: 21698132
36. Noshay JM, Anderson SN, Zhou P, Ji L, Ricci W, Lu Z, et al. Monitoring the interplay between transposable element families and DNA methylation in maize. *PLoS Genet*. 2019; 15: e1008291. <https://doi.org/10.1371/journal.pgen.1008291> PMID: 31498837
37. Haig D. Retroviruses and the placenta. *Curr Biol*. 2012; 22: R609–13. <https://doi.org/10.1016/j.cub.2012.06.002> PMID: 22877784
38. Ondičová M, Oakey RJ, Walsh CP. Is imprinting the result of “friendly fire” by the host defense system? *PLoS Genet*. 2020; 16: e1008599. <https://doi.org/10.1371/journal.pgen.1008599> PMID: 32271759
39. Haig D. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity*. 2014; 113: 96–103. <https://doi.org/10.1038/hdy.2013.97> PMID: 24129605
40. Dilkes BP, Comai L. A differential dosage hypothesis for parental effects in seed development. *Plant Cell*. 2004; 16: 3174–3180. <https://doi.org/10.1105/tpc.104.161230> PMID: 15579806
41. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17: 10–12.
42. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015; 12: 357–360. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142
43. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31: 166–169. <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700
44. Zhou P, Hirsch CN, Briggs SP, Springer NM. Dynamic Patterns of Gene Expression Additivity and Regulatory Variation throughout Maize Development. *Mol Plant*. 2019; 12: 410–425. <https://doi.org/10.1016/j.molp.2018.12.015> PMID: 30593858
45. 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