TITLE

TE-like DNA methylation of highly expressed endosperm genes in maize

AUTHORS

Yirui Sun, Yibing Zeng, R. Kelly Dawe, and Jonathan I. Gent

ABSTRACT

Unlike CG methylation alone, which is associated with stable and moderate levels of transcription, gene body methylation in both CG and CHG contexts is associated with poorly transcribed and missanotated genes. Recent results show that removal of this TE-like methylation from a set of maize pollen specific genes promotes high-level transcription. Here we report that TE-like methylated genes that are maternally demethylated are expressed toward the extreme high end of the spectrum in endosperm. As with their pollen counterparts, they have few exons, and are enriched for short, secreted proteins. Nearly half encode zeins, the major seed storage proteins of maize. When methylation is high in promoters as well as in gene bodies, their paternal copies are silenced in endosperm, resulting in genomic imprinting. These findings reveal that sporophytic TE-like methylation in gene bodies can be a signature of nearly-constitutive silencing but high and specific expression in endosperm or pollen.

INTRODUCTION

The endosperm is neither sporophyte nor gametophyte. Like the embryo, it is produced by fusion of two gametes; but unlike the embryo, it is a terminal side branch from the alternation-of-generations life cycle (Fig. 1A). Both male and female gametophytes carry two gametes, the monomorphic male sperm cells, and the dimorphic female egg cell and central cell. In most angiosperm species, the central cell is homodiploid. That is, it has two haploid nuclei derived from one megaspore (the haploid product of female meiosis) [1]. In other species it can be heterodiploid, with the two haploid nuclei each derived from different meiotic daughter nuclei. It can also have other ploidies and be derived from one, two, or four meiotic daughter nuclei. The maize central cell is typical of angiosperms in being homodiploid and giving rise to triploid endosperm upon fertilization by a haploid sperm cell (Fig. 1B) [2]. The endosperm is critical to embryo development and germination, and in many species, including maize, to early seedling growth. Central to this is its role in transferring nutrients from maternal cells to the developing embryo and in nutrient storage for later use. Its initial development is coenocytic, with rapid and synchronized nuclear replication without cell division producing 128–512 nuclei within the space defined by the central cell's large cytoplasm.

In addition to the expected two-to-one ratio of maternal-to-paternal gene dosage, maternally genomically imprinted genes can have much higher ratios of maternal-to-paternal

dosage. In effect, this causes imprinted genes to maintain a partially haploid-like expression state after fertilization. The phenomenon is closely connected to repression of transposons/transposable elements (TEs). Not only are core mechanisms of TE silencing shared with mechanisms of imprinting, many TEs are themselves imprinted [3–7]. In fact, the first example of an imprinted gene, the maize gene *r1*, was later discovered to have a CACTA DNA transposon as its promoter [8,9]. In the best-studied form of imprinting in plants, DNA demethylation of maternally-inherited cis-regulatory elements allows binding of activators and preferential maternal expression. Other means of imprinting are also common in plants [10]. Imprinting by demethylation involves both CG and CHG methylation together (H = A,T, or C). Depending on the species and on other details of local DNA sequence context and chromatin context, it can also involve CHH methylation[11–13]. In maize CHH methylation co-occurs with CHG and CG methylation at sites of RNA directed DNA methylation (RdDM) at euchromatin/heterochromatin boundaries and has little connection to imprinting [14,15].

CG and CHG together, and in some cases CHH methylation, are a central feature of heterochromatin, where they function with histone modifications, histone variants, and other factors to repress transcription. DNA methylation does so primarily in the context of genome defense, and its repression is characteristically stable to keep potentially harmful DNA repressed throughout development. This stability of DNA methylation throughout sporophyte development has been well-documented in maize [16–18]. However, tissue-specific and targeted DNA demethylation can allow for dynamic activation of gene expression, as occurs in some imprinted genes with methylated promoters. The enzymes responsible are bifunctional DNA glycosylases/lyases (DNGs for short) [19]. “Glycosylase” refers to the enzymes’ ability to cleave the sugar-base bond to create an abasic site, and “lyase” to their ability to cleave the sugar-phosphate backbone. The end result is base excision repair with an unmethylated cytosine. Mutant screens in Arabidopsis identified the DNG gene *DEMETER* as essential for imprinting of many genes, and mutants of a rice homolog showed a similar imprinting phenotype [20,21]. A maize imprinting mutant identified soon after the discovery of *r1* imprinting called *maternal derepression of r1* (*mdr1*) was recently discovered to also encode a DNG [15,22–24]. DNGs are critical to endosperm development as evidenced by early lethal mutant phenotypes, likely arising from failure to properly express master regulators of endosperm development [25–29]. Imprinting is known to initiate prior to fertilization in Arabidopsis and rice, which could explain why demethylation is mostly limited to the maternal genome [30]. DNG activity, however, is not limited to the central cell. DNGs have minor effects on gene expression in sporophytic tissues [31,32] and on paternal alleles of some loci in endosperm [33] and major effects in pollen [34,35]. It is possible that DNG activity on the maternal genome continues after merging of the two genomes if there are maternal epigenetic marks to guide it.

Imprinting is limited to select genes because of locus-specific rather than genome-wide activity of DNGs. For example, DNGs in Arabidopsis sporophyte targets regions marked by histone H3 lysine 4 trimethylation (H3K4me3) by an unknown mechanism [36]. Little is known about how DNGs are targeted to specific loci in endosperm, though, like in sporophyte, there is an association with transposons near genes [37–39].In contrast to methylation in promoters, methylation in gene bodies does not have a clear relationship to transcriptional repression. In maize as in other flowering plants, its effects depend on both local nucleotide contexts of the cytosines and larger contexts of genes and transposons [40]. Methylation in CG and CHG have

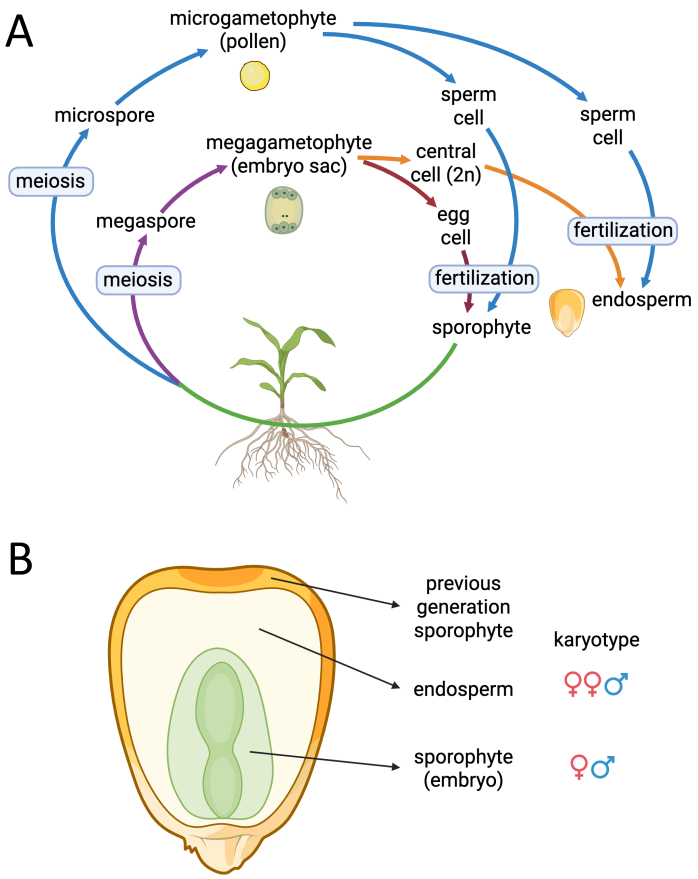

Fig. 1: Endosperm, sporophyte, and gametophyte in the context of an angiosperm life cycle.
a. Endosperm is a terminal offshoot from the angiosperm lifecycle formed by fusion of sperm with a binucleate central cell (female gamete). **b.** The three major parts of a kernel of maize. Endosperm is triploid because it inherits two copies of the maternal genome and one of the paternal genome. The previous generation sporophyte is called the pericarp in maize (not to scale).

little if any repressive effect on the genes that host them when they mark TEs in introns. An extreme example is the TR-1 kinesin gene, which contains over a hundred Kb of TEs in its introns, yet is actively transcribed [41]. CG and CHG methylation in coding exons, however, is strongly associated with poorly expressed, non-conserved, and misannotated genes [40]. Because of its similarity with methylation in TEs, this form of gene body methylation is referred to as TE-like methylation, abbreviated as teM [42]. Most accurately annotated and functional maize genes lack CHG methylation (and lack CHH methylation) in their coding DNA but can have a spectrum of CG methylation levels correlated with constitutive and stable expression [40]. Genes on the higher end of the CG methylation spectrum, are called gene body methylated, abbreviated as gbM [43].

While teM genes are largely silent or poorly expressed in sporophyte cells, a possible exception was observed in pollen early on, where some teM genes have robust pollen expression [44] (At the time, the term “RdDM targets” was used because the term “teM” had not been adopted yet, nor the subtleties of RddM and non-RdDM CHH methylation understood). In a maize double mutant of two DNG homologs *mdr1* (a.k.a. *dng101*) and *dng102*, a set of ~50 identified target genes exhibit ~100-fold reduced expression relative to wild-type pollen [45]. These genes, which are not the complete set of DNG targets in pollen, account for over 10% of the pollen transcriptome because of their unusually high expression levels in pollen.The majority are teM genes and have near-zero expression in sporophytic tissues. They are also unusual in their lack of introns, with most having either none or one. The majority have predicted functions in cell walls. Their high expression and pollen-specificity in pollen led us to ask whether similar teM genes exist with high expression in endosperm, where DNGs are already known to regulate imprinted gene expression. To answer this, we used both gene expression patterns and TE-like methylation to define a list of 67 candidate genes. They encompass a wide variety of genes expressed at different stages of endosperm development, from secreted antifungal proteins and clavata homologs to transcription factors. Most strikingly, however, was the 23 zeins, the major seed storage prolamin genes of maize. Like pollen teM genes, the endosperm teM genes encoded short proteins, had few introns, and were highly expressed. These findings illustrate two broad types of DNG target genes in endosperm, classical imprinted genes with flanking methylation and moderate expression and tissue specificity, and the TE-like methylated genes with both flanking methylation and internal methylation across their gene bodies and unusually high expression and specificity. Surprisingly, only a subset of the TE-like methylated genes exhibited imprinted expression, which could be attributed to the amount of methylation around transcription start sites rather than in gene bodies.

RESULTS

Identification of endosperm genes with TE-like methylation

Using similar methods as previously, we defined teM genes as having both mCG and mCHG levels each equal or greater to 40% in coding DNA [40,45]. Since selecting for TE-like methylated genes also selects for mis-annotated genes, we only included core genes in this analysis, the ones that are annotated in syntenic positions in the B73 genome and all 25 other NAM founder genes [18]. Given the temporal dynamics of gene expression through endosperm

development, we defined each gene's expression by its maximum TPM value over a time course of endosperm development [46]. We then compared this value to its maximum expression in the nine non-endosperm tissues assayed in the NAM founder genome project. Using a minimum threshold of 20 TPM in endosperm and fivefold more expression in endosperm than any of the nine other tissues, we identified 67 genes that we refer to as endosperm teM genes.

Reduced methylation of teM genes in endosperm

If endosperm teM genes are DNG targets, they should be methylated in sporophyte but demethylated in endosperm (at least the for the two maternal genomes). Comparisons of methylation profiles in endosperm and multiple sporophyte tissues showed a decrease in methylation in their gene bodies and in both flanks in endosperm (Fig 2A-C). For these comparisons, we used previously published EM-seq data from mature pollen and from embryos and endosperm collected 15 days after pollination (15-DAP) from W22 inbred stocks [15,45], which required finding homologous gene annotations between W22 and B73 genomes. 42 of 67 endosperm teM genes had syntenic annotations in both genomes. The differences between sporophyte and endosperm were clearest in CG methylation because endosperm has a global reduction in CHG methylation that is not limited to DNG targets. CHH methylation was low in and near endosperm teM genes, similar to other teM genes [40]. We also compared endosperm with EM-seq data from W22 pollen [45] and saw no evidence for demethylation of endosperm teM genes in pollen (Fig 2A-C).

Since many imprinted genes are regulated by DNGs, we asked how methylation patterns in endosperm teM genes compared with imprinted genes. Since imprinting can be mediated by polycomb repressive complex2 (PRC2) and histone methylation without TE-like DNA methylation [10], we created a list of high-confidence DNG-imprinted genes in maize by selecting genes with maternally preferred expression and which overlapped previously identified endosperm demethylated regions (DMRs) [15] in their gene bodies or in their 1-Kb flanking regions. We defined maternal preferred expression using published RNA-seq data from reciprocal W22 × B73 crosses[47]. 61 core genes met these criteria. After removing seven which were also in our endosperm teM gene list, 54 were left. For this study, we refer to them as flank-methylated endosperm genes (FMEGs) because of their methylation and demethylation in flanking cis-regulatory elements regions, as is typical for imprinted genes. For clarity, in sporophyte, both FMEGs and endosperm teM genes can have methylation in their flanks (Fig. 2A-C). Both sets also have reduced methylation in their flanks in endosperm, especially in 5' regions.

As an alternative means of quantifying methylation, we focused on the 1-Kb regions centered on transcription start sites and measured their average CG methylation for each gene. For this analysis, we included only genes which were methylated in embryo to allow for the possibility of detecting demethylation in endosperm. 37 of the 42 regions had average CG methylation of at least 20% in embryo. Of these 37, 36 had decreased CG methylation in endosperm compared to embryo in these regions, with a median value in endosperm being 35.5% of the value in embryo (Fig. 2E,F). In theory, demethylation could be a passive outcome of dilution of methylated DNA with unmethylated DNA during replication rather than active demethylation by DNGs. The best test of this would be to compare wild-type with *mdr1 dng102* double mutant endosperm; however, double mutants fail early in development and are difficult

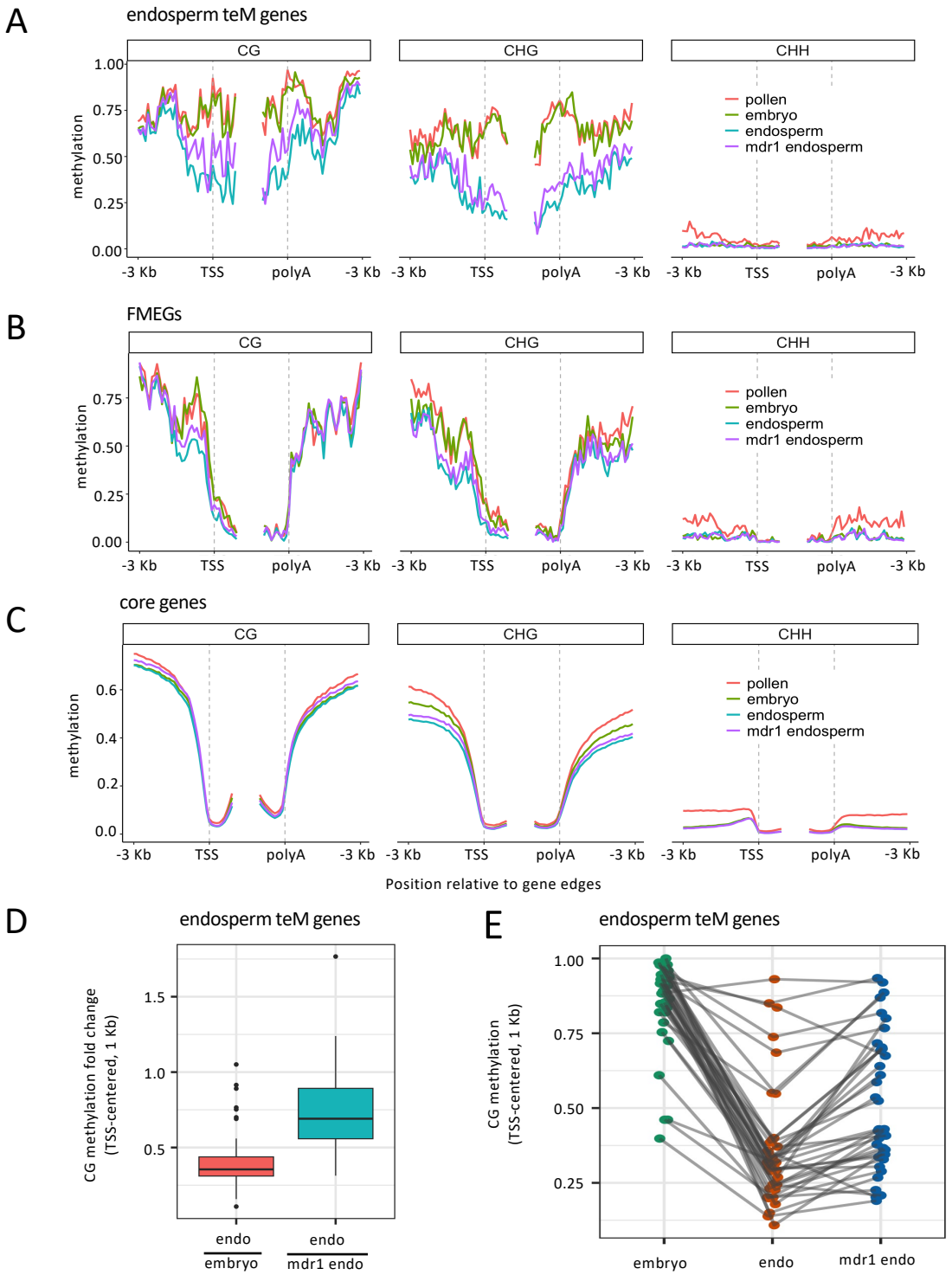

Fig. 2: Endosperm teM genes have reduced CG and CHG methylation in endosperm

a. Metagene methylation profile for endosperm teM genes in the W22 genome. These are the 42 of 67 genes present in both the W22 and B73 genome annotations. Profiles are centered on transcription start sites (TSSs) and polyadenylation sites (polyA). Methylation was measured on 100-bp intervals, from 3-Kb external to 1-Kb internal. **b.** Metagene methylation profile for the 54 FMEGs (flank-methylated endosperm genes) in the W22 genome. These represent classical imprinted genes that are maternally demethylated and expressed in endosperm. **c.** Metagene methylation profile for 28291 core genes. Core genes are annotated at syntenic positions in W22, B73, and the 25 other NAM founder genomes. **d.** Methylation differences measured as fold changes. Boxplots show distributions of CG methylation for TSS-centered 1-Kb regions for each gene as fold changes between endosperm and embryo (both wildtype) or between wildtype endosperm and mdr1 mutant endosperm. **e.** CG methylation values underlying differences in (**d**).

to work with [15]. Instead we used the *mdr1* single mutant which is fully viable but has a detectable increase in methylation at DNG target loci in endosperm. In spite of the noise introduced by random sampling of demethylated maternal reads and methylated paternal reads, 34 of 37 the 1-Kb regions had decreased methylation in wild-type relative to *mdr1* mutant, with a median wild-type value being 69.1% of mutant endosperm (Fig. 2E F). Together these data reveal that like many imprinted genes, endosperm teM genes are demethylated by DNGs. Unlike most imprinted genes, however, they are demethylated not just in flanking cis-regulatory elements, but also throughout their coding DNA.

Endosperm specific and high expression of methylated endosperm genes

By definition, endosperm teM genes had a minimum of fivefold more expression in endosperm than in the other nine tissues examined. However, the median expression difference was much higher, 258-fold, and the mean of 661-fold, indicating these genes were exceptionally endosperm specific (Fig. 3A,B). As a comparison we defined a set of 1091 endosperm genes using the same criteria as endosperm teM genes but without regard to methylation. Of these 1091 endosperm genes, which include the 67 teM genes, the median difference in expression was 12-fold, and the mean 132-fold. FMEGs had a slight preference for endosperm expression, median 1-fold, and a mean of 69-fold. Like their counterparts in pollen, the majority (53 of 67) of endosperm teM genes were encoded in two or fewer exons, 28 of these in a single exon (Fig. 3C). They also encoded short proteins with a median CDS length of 465 bp, compared to 1272.5 for FMEGs and 1052 for core genes (Fig. 3D). Endosperm teM genes were also among the highest expressing genes in endosperm. 67 genes made up more than 40% of the transcriptome at 16 DAP (Fig. 3E). At every stage of development from 6-DAP to 34-DAP, at least 20 were in the top 1%, and at least 26 in the top 10% (Supplemental Fig. S1). At 8 DAP, 46 of 67 were in the top 10%.

TE-like methylation of early-expressed secreted proteins and late-expressed zeins

Endosperm consists of diverse cell types with different functions, and it undergoes major changes as it progresses from the central cell gamete, to the multinucleate coenocyte, to a predominantly starchy structure with few living cells [2]. In rice and Arabiopsis, demethylation initiates in the central cell [30]. In maize, the fact that the *mdr1* mutant has an *r1* pigmentation phenotype in *mdr1* heterozygote endosperm when *mdr1* is maternally inherited but not paternally inherited also suggests key activity in the central cell [22,23]. The fact that the *r1* pigmentation gene that is demethylated by DNGs is expressed in the outermost cell layer late in endosperm development suggests that the unmethylated state from the central cells persist at least partially through endosperm development. To test whether teM genes show any characteristic expression patterns, we first clustered them using the same time course expression data that we used to identify them [46]. This revealed two major clusters of genes, cluster 1 with 39 genes that peaked within the first eight days after pollination (DAP) and declined after, and cluster 2 with 28 genes that started low, increased rapidly at 8 DAP, and stayed high until 30 DAP (Fig. 3G,F and Supplemental Fig. S2). FMEGs produced a more variable set of expression patterns than teM genes and did not produce clear clusters (Supplemental Fig. S2).

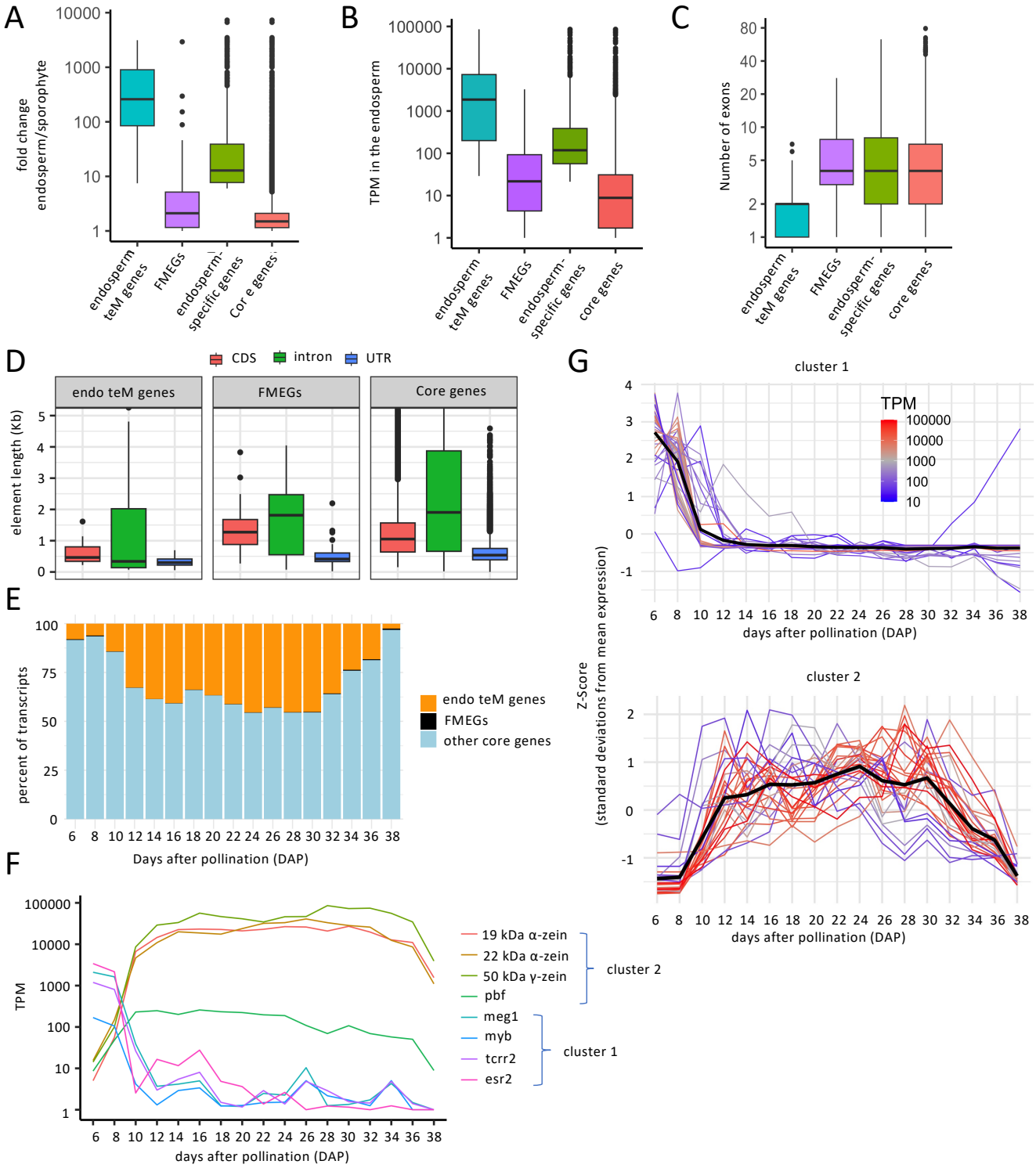

Fig. 3: Expression patterns and structural features of endosperm teM genes

a. Endosperm teM genes are exceptionally endosperm specific. The Y-axis values indicate the ratio of the highest expression in endosperm development divided by the highest expression in any of the nine sporophyte transcriptomes from the NAM founder study. Box plots show the distribution for the 67 endosperm teM genes, the 54 FMEGs, and the 28291 core genes. Also included is a set of 1091 endosperm-specific genes that were defined by having at least fivefold more expression in at least one of the endosperm stages than any of the nine sporophyte transcriptomes from the NAM founder study. **b.** Endosperm teM genes are exceptionally highly expressed. Box plots show the distribution of maximum transcript per million (TPM) values for each gene over the endosperm development time course. **c.** Endosperm teM genes have few exons. **d.** Endosperm teM genes encode short proteins. **e.** Endosperm teM genes make up a large fraction of the endosperm transcriptome. **f.** Endosperm teM genes fall into two major clusters based on developmental expression profiles. The thick black lines indicate the average of all individual genes. **g**. Expression of example endosperm teM genes during development.

Cluster 1 of the endosperm teM genes had wide-ranging spatial expression patterns and gene functions. Some genes, such as the carboxypepsidase Zm00001eb240770 or the ubiquitin paralogs Zm00001eb240770 and Zm00001eb275020 are expressed across endosperm compartments [48]. There was a strong trend among these genes, however, for expression in the peripheral cell layers of the endosperm adjoining the embryo, the maternal chalazal cells, or the maternal pericarp cells. These layers, called the embryo surrounding region (ESR), basal endosperm transfer layer (BETL), and aleurone, are all critical for signaling between endosperm and embryo or between endosperm and the maternal sporophyte [2]. Consistent with this, 15 of the cluster 1 genes are known to encode proteins that are transferred between cells. These are the defensin-like BETL1/BET1 [49], two tandem duplicates of the antifungal protein BETL2/BAP1 [50]; three lipid transfer proteins, BETL9/NPTD6, AL9, and PLT32/NPTD16 [48,51,52]; five tandem duplicates of the BETL developmental regulator MEG1 [53]; the protein of unknown function AE1/ANE1 [54]; the endopeptidase EBE2 [55]; and the CLAVATA3/ESR homolog ESR2 [56]. Two two-component response regulators, TCRR1 and TCRR2, are also transferred out of the BETL, but into adjacent endosperm, not maternal sporophyte nor embryo [57]. The MYB-related transcription factor MYBR19 that activates BETL genes was also in cluster 1 [58]. Two other MYB transcription factors (MYBR9 and Zm00001eb378960) and the AP2/ERF transcription factor ABI26) from cluster 1 that were included in a large-scale endosperm transcription atlas study are also strongly enriched in one or more of the three boundary cell layers [48].

Cluster 2 was made up mainly of zein family genes, 14 19-kDA α-zeins, 8 22-kDA α-zeins, and one 50-kDa γ-zein. Zeins are prolamins, the major seed storage proteins of maize. Cluster 2 also encodes PBF1, a transcription factor that promotes zein expression [59]. Unlike cluster 1 genes with predominant early expression in peripheral cell layers, zeins express mid to late in endosperm development in the starchy endosperm. In midstage endosperm development, α-zeins make up about 50% of the total transcripts [46]. Maternal demethylation of α-zein genes was detected previously using methylation-sensitive enzyme digestion [60]. I Zein genes are encoded in families of highly-similar and short duplicate genes, which can complicate mapping of RNA-seq reads to any single gene. Independently of expression measurements, we counted how many annotated zein genes were categorized as teM genes [40]. All 18 of the 19-kDa α-zeins, 11 of the 22-kDa α-zein genes, and the single 50-kDa γ-zein genes were teM genes (Supplemental Table S2). The single 16-kDa and 27-kDa γ-zeins, the single 10-kDa δ-zein, and the single 15-kDa β-zein genes were unmethylated; and the single 18-kDa δ-zein had intermediate methylation. Demethylation of seed storage proteins and activation by DNGs has also been reported in rice [61,62].

Imprinting of endosperm teM genes is associated with methylation in promoters

Some alleles of α-zein genes were reported to have maternal preferred expression [47,60], as was a *meg1* gene [63]. Quantifying imprinting relies on use of hybrid endosperm with genetic variants to distinguish between parental alleles. This becomes problematic for short genes, especially ones with multiple copies, where RNA-seq reads do not clearly distinguish between gene copies. Thus absence of evidence for imprinting from existing studies may not indicate absence of imprinting. To more thoroughly investigate the extent of imprinting among the endosperm teM genes, we examined the source data from a recent study that included eight different genetic backgrounds and multiple timepoints [47]. Because of the variability of imprinting

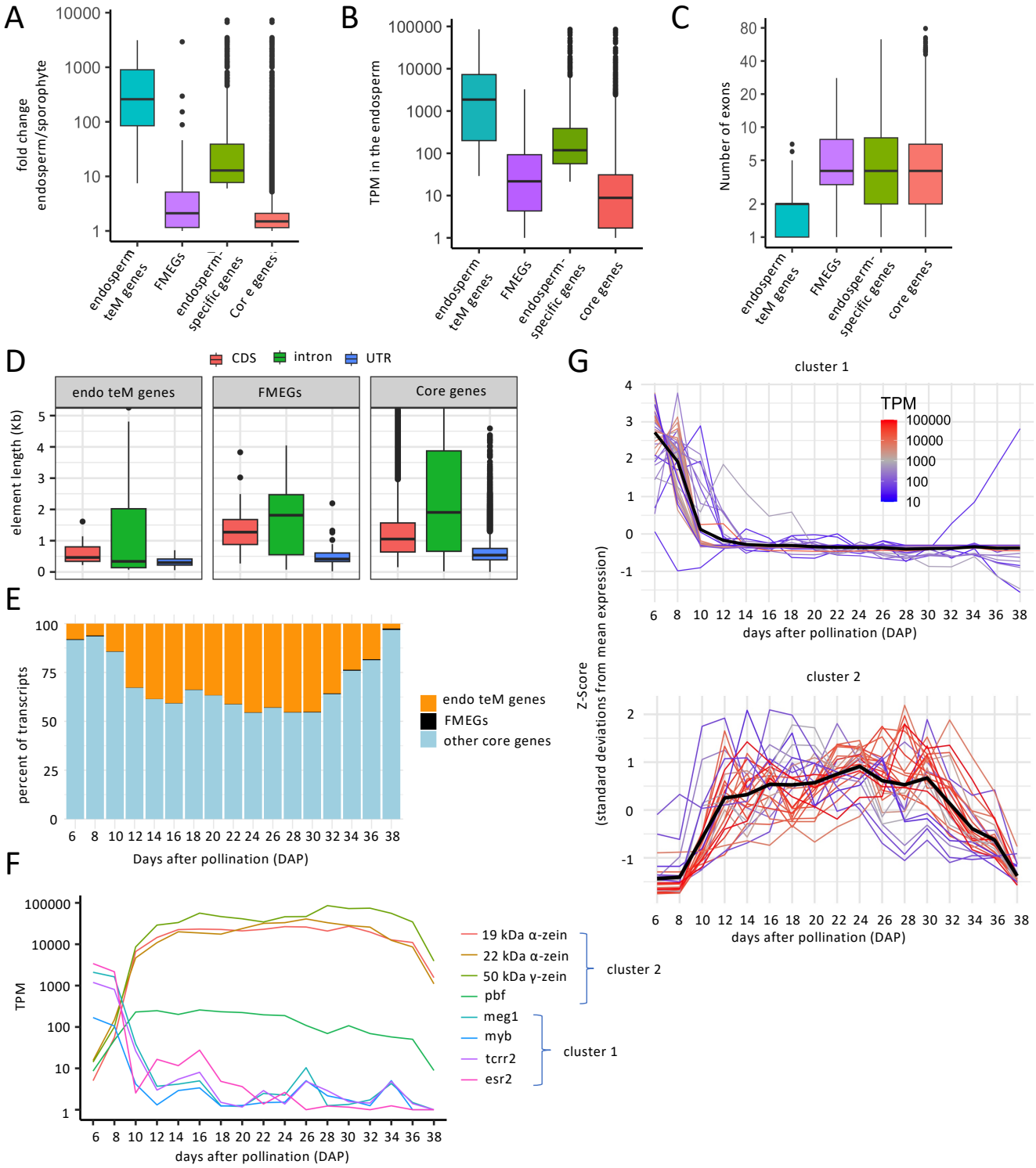

Fig. 3: Expression patterns and structural features of endosperm teM genes

a. Endosperm teM genes are exceptionally endosperm specific. The Y-axis values indicate the ratio of the highest expression in endosperm development divided by the highest expression in any of the nine sporophyte transcriptomes from the NAM founder study. Box plots show the distribution for the 67 endosperm teM genes, the 54 FMEGs, and the 28291 core genes. Also included is a set of 1091 endosperm-specific genes that were defined by having at least fivefold more expression in at least one of the endosperm stages than any of the nine sporophyte transcriptomes from the NAM founder study. **b.** Endosperm teM genes are exceptionally highly expressed. Box plots show the distribution of maximum transcript per million (TPM) values for each gene over the endosperm development time course. **c.** Endosperm teM genes have few exons. **d.** Endosperm teM genes encode short proteins. **e.** Endosperm teM genes make up a large fraction of the endosperm transcriptome. **f.** Endosperm teM genes fall into two major clusters based on developmental expression profiles. The thick black lines indicate the average of all individual genes. **g**. Expression of example endosperm teM genes during development.

between stages of endosperm development and between alleles of the same gene, we only considered genes with consistent maternal preference across multiple genotypes 13 of 67 endosperm teM genes passed our quality control standards (Supplemental Table S1). Without imprinting, the expected maternal ratio of expression is ⅔ becuase endosperm carries two maternal genomes for each paternal genome. With threefold maternal preference, the maternal ratio would be 6/7 (~0.86). Six of the 13 exceeded that threshold (Fig. 4A). In comparison 51 of 10,700 qualifying core genes did. This amounts to a 97-fold enrichment for maternal preference among endosperm teM genes.

The lack of maternal preference for seven of the measured genes indicates little or no impact of DNA methylation on their paternal expression in endosperm. A plausible explanation is that the presence of TFs such as O2 and PBF1 that drive unusually high levels of transcription in endosperm[2] could override paternal DNA methylation, particularly in the unusual chromatin environment of endosperm[64–66]. Since CHG methylation is globally reduced in maize endosperm (Fig. 2B), we looked at parent-specific methylation in data from B73/W22 hybrid endosperm [67] to check whether the paternal DNA of teM genes also has reduced CHG methylation in endosperm. Relative to developing leaf, paternal mCHG was reduced in 5' regions of endosperm teM genes, from 74% in leaf to 51% in mature endosperm (Supplemental Fig. S3). Reduced CHG methylation in maize endosperm and other changes in heterochromatin could facilitate expression of some sperm teM genes from the paternal alleles, but it does not explain why some have strong maternal preference (e.g., *meg1* genes, some zeins, and the five of 13 identified here).

Given the stronger connections between methylation in cis-regulator elements and gene silencing than between between methylation in gene bodies and gene silencing [18,40]), we hypothesized that the amount of methylation in promoters in endosperm teM genes would inversely correlate with paternal expression. CG methylation in particular should have a strong effect since methylation in this context often reaches 100% (Supplementary Fig. S3). To test this, we took into account not only methylation as the ratio of methylated cytosines to total cytosines at each position, but also the number of cytosines in each promoter. The rationale behind this is that if there were ten cytosines in a 100 bp region and five were methylated (50% methylation), the effect on transcription would likely be more severe than if there was only one cytosine and it was 100% methylated (Fig. 4B). Normalizing by the total number of nucleotides rather than just the total number of cytosines would give values of 5% for the first case but 1% for the second. Consistent with our hypothesis, the six genes with strong maternal preference had a mean N-normalized CG methylation value of 5.06%, while the seven without maternal preference had a mean N-normalized CG methylation value of 0.44% (Fig. 4C and (Supplemental Fig. S4)). This difference was statistically significant (p-value = 0.0017, one-tailed Wilcoxon rank-sum test). There was also a significant difference in promoter CHG methylation even though CHG methylation is globally reduced in endosperm (2.1% vs 0.58%, p-value 0.019). Methylation levels in gene bodies did not correlate with imprinting (Supplemental Fig. S4). Extrapolating from the 13 tested endosperm genes, we expect based on their promoter methylation levels in B73 leaf that about 30 of 67 would also be strongly maternally preferred (Fig 4C). 8 of these were zeins (Supplemental Table S1).

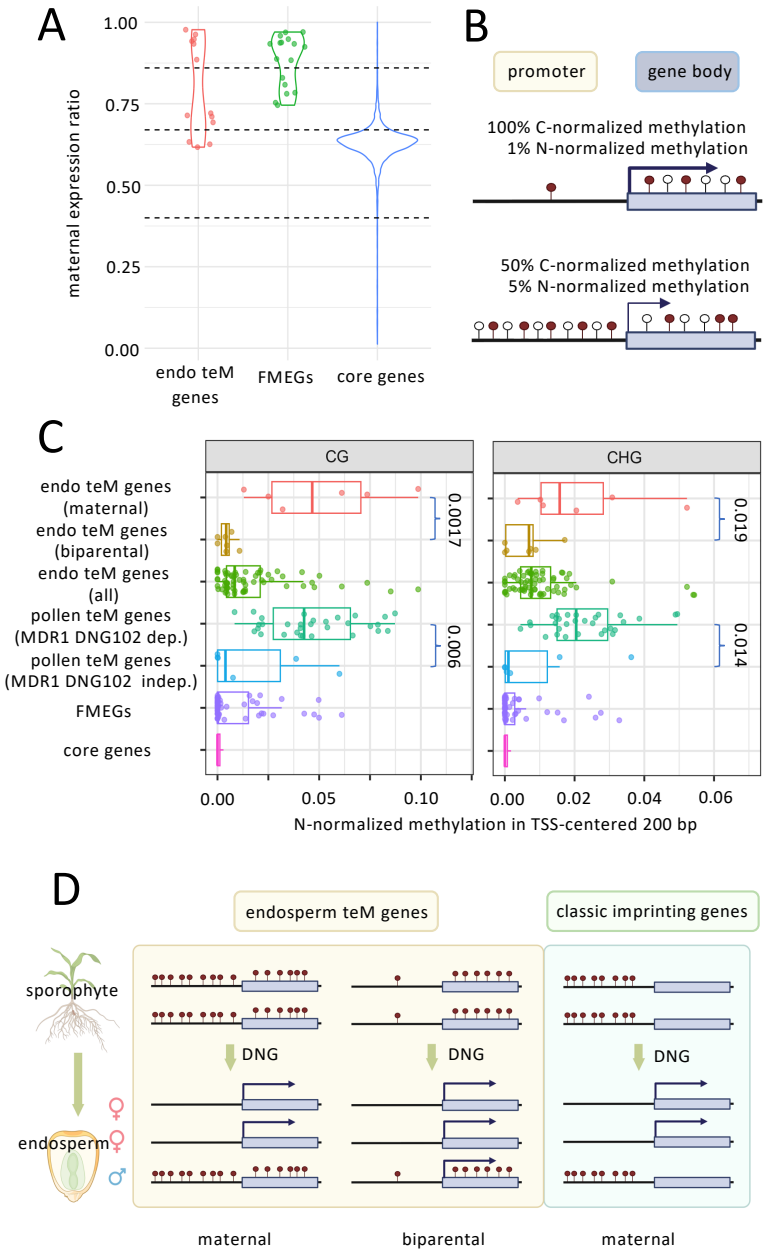

Fig. 4: Imprinting of endosperm teM genes correlates with promoter methylation.
a. Some endosperm teM genes have imprinted expression, some do not. Maternal expression ratio is the proportion of maternally-derived transcripts as measured by parent-specific sequence variants in RNA-seq data from hybrids. Upper dashed line indicates a maternal expression ratio of 6/7 (threefold more than expected), the middle line indicates a ratio of 2/3 (expected), and the lower line indicates a value of 2/5 (threefold less than expected). **b.** A schematic of nucleotide-normalized (N-normalized) methylation vs cytosine-normalized methylation (C-normalized). Open lollipops represent unmethylated cytosines, solid lollipops by methylated cytosines. **c.** N-normalized methylation in the 200-bp region centered on transcription start sites (TSSs) of endosperm teM genes correlates with dependence on demethylation for expression. For endosperm teM genes, "maternal" are the six genes in (**a**) with maternal expression more than threefold greater than expected, "biparental" are the seven with expected maternal expression, and "all" is the complete set of 67. For pollen teM genes, "MDR1 DNG102 dep" are the set of 31 whose pollen expression depends on these two DNGs, and "MDR1 DNG102 indep" are the set of six whose pollen expression is independent of them. **d.** A model for imprinting of endosperm teM genes based on methylation in promoters.

Dependence on MDR1 and DNG102 of pollen endosperm teM genes is associated with methylation in promoters

While the majority of pollen teM genes (called methylated pollen genes in the prior study) are strongly downregulated in *mdr1 dng102* double mutant pollen, some retain high expression [45]. Of 56 pollen teM genes identified, 37 had high confidence expression measurements in the double mutants, with 31 being called as differentially expressed and six retaining high expression. We had hypothesized that one or both of the other DNGs encoded in maize (DNG103 and DNG105) might demethylate these six genes in pollen. However, our finding that methylation in promoters of endosperm teM genes associated with paternal silencing led us to ask whether the six pollen genes that appeared to be unaffected in *mdr1 dng102* double mutants might simply have fewer cytosines in their promoters. Supporting this hypothesis, the differentially expressed pollen teM genes had a mean N-normalized CG methylation value of 4.7% in promoters while the unchanged genes had a value of 1.8% (with p-value 0.006) (Fig. 4C). CHG methylation in their promoters also strongly correlated with differential expression, with mean N-normalized CHG methylation value of 2.3% for differential expressed genes and 0.9% for unchanged genes (with p-value 0.014). Together with the results from the 13 endosperm teM genes that have high-confidence measures of imprinted expression in endosperm, these results suggest that methylation near transcription start sites rather than in the gene body is key to silencing teM genes, at least in endosperm and pollen cells which are primed to express these genes.

DISCUSSION

While endosperm teM genes and classical imprinted genes share a common biochemical feature, targeting by DNGs for demethylation, they have notable differences. First, teM genes are more tissue specific and more highly expressed. Second, their methylation is throughout their gene bodies. Third, they have few introns and encode shorter proteins. The connection between endosperm teM genes and imprinting also raise a related question of why imprinting exists and what benefits it provides. One theory is that it could allow for haploid selection in diploid cells (or triploid in the case of endosperm)[68]. Another is that maternally preferred gene expression favors equal distribution of maternal resources among seeds, though at the cost of smaller or less nutrient-dense endosperm [69,70]. Maternally preferred zein expression to be an exception, as is maternally preferred expression of *meg1* genes [53] and of DNG-mediated demethylation of glutelins and starch synthase genes in rice [61]. Another explanation for imprinting is as a mechanism for adjusting gene dosage from what was optimal in the central cell and sperm to what is optimal in endosperm [71]. Each parental gamete contributes not just different amounts of chromatin, but also different amounts of diverse gene products that must function together after fertilization at the correct stoichiometric ratios. Differential gene regulation between parental genomes could aid in producing optimal gene dosage for some genes. Related to this, endosperm development is preferentially influenced by the maternal nuclear and cytoplasmic genomes in multiple ways other than through genomic imprinting; for example, development of the female gametophyte itself strongly influences development of the endosperm, particularly the BETL [72]. Thus the logic of having stronger

maternal control of some genes post-fertilization may not be easy to experimentally separate from pre-fertilization functions.

The causes and consequences of TE-like methylation in gene bodies may, like CG gene body methylation, also turn out to be difficult to pin down. Unlike CG gene body methylation, however, TE-like methylation in gene bodies is clearly associated with strong repression, at least in sporophyte cells, and its removal in endosperm and pollen with strong expression. How TE-like DNA methylation functions in gene regulation has been interpreted largely through the lens of genes like FWA in Arabidopsis, where methylation of TE fragments in its promoter prevents ectopic expression outside of endosperm [73]. Many other related examples have also been reported where methylation of TEs and other repetitive DNA in cis regulatory elements sensitize gene expression to DNA methylation, e.g, *ROS1* and *SDC* in Arabidopsis; and *b1*, *r1*, *pl1*, and *ufo1* in maize [9,74–79]. In contrast to these unusual situations, gene regulation in plants does not normally involve DNA methylation because cis-regulatory elements are constitutively unmethylated [16]. For some cis-regulatory elements that may require continuous surveillance by DNGs to actively remove methylation [32,36]. teM genes represent something different (ignoring the many dubious gene annotations with TE-like methylation because they are in fact TEs, pseudogenes, or misannotated genes [40]). TE-like methylation also needs to be distinguished from the CG-only methylation that is conventionally called gene body methylation [43]. These two forms of gene methylation are on opposite extremes of expression patterns, while unmethylated genes are somewhere in the middle (Fig. 5).

In pollen, highly cell-type-specific and strong expression of most but not all teM genes is directly dependent on demethylation by MDR1 or DNG102 [45]. Our findings here indicate some endosperm teM genes behave similarly to pollen teM genes in that their highly tissue-specific and strong expression is limited to the demethylated maternal genome alleles. Some, however, have highly tissue-specific and strong expression from both parental genomes in spite of the paternal alleles retaining CG methylation at sporophyte levels. We found that the difference between the maternal preferred teM genes and biparental ones could be explained by methylation in promoters rather than in gene bodies, an explanation that is supported by known repressive effects of methylation in promoters and by a similar pattern of methylation in promoters correlating with dependence of pollen teM genes on MDR1 and DNG102. This raises the question then of why some genes would have TE-like methylation in just the gene body and not in promoters if it does not repress expression in endosperm. One possibility is that the methylation in the gene bodies is useful because it contributes to repression of genes in the sporophyte. The other is that the methylation in gene bodies is not functional and can arise as a consequence of the genes' never being expressed in the sporophyte lineage. In that case, TE-like methylation in gene bodies would be a signature of high expression in endosperm and repression in sporophyte rather than a cause of that expression pattern. It is clear, however, that removal of TE-like methylation, most crucially near transcription start sites, is essential for expression of some of these genes, as evidenced both by strong maternal expression preference in endosperm and by the requirement for DNGs for expression in pollen [40]. Whether or not it is causal, gene body TE-like methylation in the sporophyte is a signature of high specificity and high expression in endosperm and pollen.

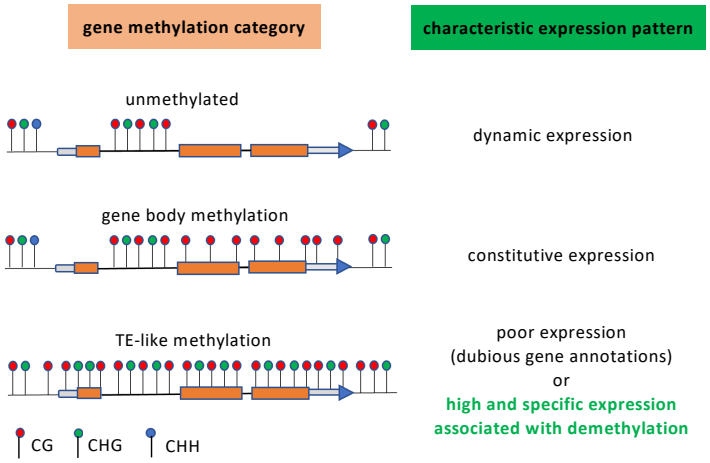

Fig. 5: Schematic of gene methylation categories.
While the vast majority of genes with TE-like methylation may reflect misannotated TEs or pseudogenes, a few are highly expressed and highly specific and function in pollen and endosperm. Figure adapted from Zeng et al. 2023 [40].

METHODS

Identification of endosperm teM genes and other gene sets from expression and methylation data

Using the Zm-B73-REFERENCE-NAM-5.0 annotation file, core genes, present in all 26 NAM genome lines and B73, were defined based on the pan_gene_matrix_v3_cyverse.csv downloaded from https://de.cyverse.org/anon-files//iplant/home/shared/NAM/NAM_genome_and_annotation_Jan2021_release/SUPPLEMENTAL_DATA/pangene-files/pan_gene_matrix_v3_cyverse.csv as described previously described [18] resulting in a total of 28,291 core genes. To identify endosperm teM genes, we collected the B73 RNA-seq data from a previous study [46] spanning 6 to 38 days after pollination (DAP). All 21 samples were downloaded directly from Sequence Read Archive (SRA) under accession number SRP037559 using SRA-Toolkit v3.0.3 and parallel-fastq-dump v0.6.7. The raw sequences were trimmed by Trim Galore v0.6.7 with the parameter --paired (https://github.com/FelixKrueger/TrimGalore) . Then the trimmed reads were aligned to B73 v5 genome using STAR v2.7.10b [80]. Prior to alignment, a STAR index was generated with the parameter --sjdbGTFtagExonParentTranscript Parent to format the GFF3 annotation file. STAR mapping was performed with default parameters. To count the mapping reads of each gene, we used Subread v2.0.6 [81] to count the paired-end reads of each transcript within the exons using -p --countReadPairs -t exon -O -g Parent. The -O -g Parent options ensured that the reads mapped to multiple transcript isoforms were properly assigned. Maize reference genome sequence and annotation data were downloaded at https://download.maizegdb.org/Zm-B73-REFERENCE-NAM-5.0/. We filtered the resulting raw read counts to retain only the canonical transcripts, then averaged the expression values of biological replicates. Gene expression was normalized using the TPM (Transcripts Per Million) method to account for sequencing depth and transcript lengths. To identify endosperm teM genes, we used similar criteria used in the previous study of pollen teM genes, considering only methylation within CDSs. Methylation data from developing second leaves and TPM from nine sporophyte tissue samples was obtained from prior work[40] which was derived from the NAM founder genome assembly project[18] (https://raw.githubusercontent.com/dawelab/Natural-methylation-epialleles-correlate-with-gene-expression-in-maize/refs/heads/main/Data/B73.all.csv). The nine tissue samples were embryo, tassel, anther, shoot, ear, leaf tip, leaf middle, leaf base, and root. Genes were classified as teM genes if they had at least ten informative CHGs (ones that were spanned by EM-seq reads) and methylation values of at least 40% for CG and for CHG in leaf.

To select for teM genes that are enriched for endosperm expression, we required a ≥ 5-fold increase in TPM in at least one stage of endosperm development endosperm compared to all other nine sporophyte samples. In addition, a minimum TPM of 20 in at least one endosperm stage was required to filter out low-expressed genes with unreliable TPM fold change measurements. This produced a set of 119 genes. After removing ones that were not in the core

gene set, this analysis identified 67 endosperm teM genes. To define a control set of genes, all the same criteria were used except without regard to DNA methylation. This produced a set of 1,091 genes that we refer to as endosperm genes.

To generate a representative set of DNG-dependent classical imprinted genes, we used allele-specific expression data from a study where RNA-seq from reciprocally crossed W22 and B73 endosperm at 11, 14, 17, and 21 DAP was used to calculate the maternal allele expression ratio based on sequence variants associated with each genome [47]. We filtered out the genes with RPM < 1 at all four time points to avoid misleading maternal preference caused by extremely low expression. Genes with maternal ratios of ≥ 0.9 were classified as MEGs; genes with maternal ratios of ≤ 0.3 were PEGs. Genes that met criteria for both MEG and PEG at different time points were excluded. Gene ID conversion between B73 and W22 annotations were conducted based on MaizeGDB_maize_pangene_2020_08.tsv downloaded from https://download.maizegdb.org/Pan-genes/archive/pan-zea-2020/. This identified 295 core genes. To identify the subset of imprinted genes regulated by demethylation in embryo, we selected all that overlapped with endosperm demethylated methylated regions (DMRs) anywhere in their gene bodies or 1-Kb flanking regions. These DMRs were identified previously through a comparison between endosperm and embryo [15]. After removing seven genes from this list that were in our endosperm teM genes, this produced a list of 54 maternally expressed, imprinted genes that we call flank-methylated endosperm genes (FMEGs).

K-means clustering of expression patterns

To characterize the dynamic expression pattern, we calculated Z-score for the TPM values of the genes across 6 to 38 DAP using the formula $z = (x - \mu) / \sigma$, where μ and σ are the mean and standard deviation of TPM for each gene. We used factoextra 1.0.7 R package (https://github.com/kassambara/factoextra) to determine the optimal number of clusters using both Elbow Method and Silhouette Method, both of which indicated k=2 is the optimal (Supplemental Fig. S2). Then we used kmeans R function to cluster the 67 endosperm teM genes into 2 groups. Gene names, protein functions, and families were manually curated using data from MaizeGDB and by blasting protein sequences to identify conserved domains. The figures were generated by ggplot2 v3.5.1. [82].

Imprinted expression analysis

Since the teM endosperm genes tend to have fewer exons and are enriched in gene families, making them more likely to have multimapping issues, we applied a stringent criterion to assess their imprinting status. We obtained mat_pref_NAM.txt from a published research in which B73 was reciprocally crossed with 8 inbred lines, including W22, B97, Oh43, Ky21, NC358, CML333, M162W, and Ki11 [47]. Gene IDs were converted to B73 annotation using MaizeGDB_maize_pangene_2020_08.tsv downloaded from MaizeGDB at https://download.maizegdb.org/Pan-genes/archive/pan-zea-2020/. We selected the genes for which at least one allele had ≥ 5 RPM in at least one cross. To get conserved imprinting status, we further filtered the genes with maternal_preference variance ≤ 0.02 across all crosses that

meet the ≥ 5 RPM threshold. This stringent method found 32,365 genes with high-confidence imprinting data. The mean maternal preference values were used to represent imprinting status.

Identification of zein genes

Zeins genes were identified in the maize genome by searching for specific keywords in MaizeGDB [85] : Z1 for α zeins, bz15 for beta zeins, gz for gamma zeins, and dzs for delta zeins. from B73v5 with pan-genome IDs, a reference table linking gene IDs to pan IDs was retrieved from (https://raw.githubusercontent.com/dawelab/Natural-methylation-epialleles-correlate-with-gene-expression-in-maize/refs/heads/main/Data/B73.all.csv). Genes sharing the same pan ID as known zeins were incorporated into the analysis. Methylation information for these genes was obtained from the same data.

DNA methylation analyses

Whole genome methylation CGmaps produced by previous studies were used as the data sources for measuring methylation values, developing 2nd leaf from B73 [18], mature pollen from W22 [45], 15-DAP endosperm (wild-type and *mdr1* mutant) and embryo from W22 [15], and mature endosperm and developing 2nd leaf from B73 X W22 hybrids [67].
The CGmapTools [86] v1.3 MTR tool was used to measure methylation over individual loci using the “by site” method and reporting methylation using the conventional mC/total C ratio. Metagene methylation measurements were obtained using the CGmapTools MFG tool on 100 bp intervals, also using the conventional mC/total C ratio. For N-normalized methylation values, the mC/total C ratio for each region was multiplied by the number of cytosines (context specific) and divided by the total number of nucleotides. To quantify cytosines in specific sequence contexts, we implemented a custom function in R. For each gene sequence, both the forward strand and reverse complement were considered. Counts were obtained by sliding a window of 2 or 3 nucleotides across the sequence to identify CG, CHG, and CHH sites. Cytosine contexts were counted if the cytosine itself resided within the analyzed region. Parent specific methylation values were calculated using the CGmapTools SNV tool followed by the ASM tool with “allele-specific methylated site” output format, with an additional SNV quality control filtering step in between them with a custom python script “BS_VCF_filter.py”. ASM output was then used to generate parent-specific CGmaps with another custom python script, “AlleleSpecificCGmapper.py”.

REFERENCES

1. Schmid, M. W., Schmidt, A. & Grossniklaus, U. The female gametophyte: an emerging model for cell type-specific systems biology in plant development. *Front. Plant Sci.* **6**, 907 (2015).
2. Wu, H., Becraft, P. W. & Dannenhoffer, J. M. Maize endosperm development: Tissues, cells, molecular regulation and grain quality improvement. *Front. Plant Sci.* **13**, 852082 (2022).
3. Del Toro-De León, G. *et al.* Epigenetic and transcriptional consequences in the endosperm of chemically induced transposon mobilization in Arabidopsis. *Nucleic Acids Res.* **52**, 8833–8848 (2024).
4. Rodrigues, J. A. *et al.* Divergence among rice cultivars reveals roles for transposition and epimutation in ongoing evolution of genomic imprinting. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2104445118 (2021).
5. Li, T., Yin, L., Stoll, C. E., Lisch, D. & Zhao, M. Conserved noncoding sequences and de novo Mutator insertion alleles are imprinted in maize. *Plant Physiol.* **191**, 299–316 (2023).
6. Pignatta, D. *et al.* Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife* **3**, e03198 (2014).
7. Anderson, S. N., Zhou, P., Higgins, K., Brandvain, Y. & Springer, N. M. Widespread imprinting of transposable elements and variable genes in the maize endosperm. *PLoS Genet.* **17**, e1009491 (2021).
8. Kermicle, J. L. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* **66**, 69–85 (1970).
9. Walker, E. L. Paramutation of the r1 locus of maize is associated with increased cytosine methylation. *Genetics* **148**, 1973–1981 (1998).
10. Batista, R. A. & Köhler, C. Genomic imprinting in plants-revisiting existing models. *Genes*

Dev. **34**, 24–36 (2020).

11. Zemach, A. *et al.* The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
12. Gouil, Q. & Baulcombe, D. C. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet.* **12**, e1006526 (2016).
13. Stroud, H. *et al.* Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
14. Fu, F.-F., Dawe, R. K. & Gent, J. I. Loss of RNA-directed DNA methylation in maize chromomethylase and DDM1-type nucleosome remodeler mutants. *Plant Cell* **30**, 1617–1627 (2018).
15. Gent, J. I. *et al.* The maize gene maternal derepression of r1 encodes a DNA glycosylase that demethylates DNA and reduces siRNA expression in the endosperm. *Plant Cell* **34**, 3685–3701 (2022).
16. Crisp, P. A. *et al.* Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23991–24000 (2020).
17. Oka, R. *et al.* Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**, (2017).
18. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
19. Du, X. *et al.* Molecular basis of the plant ROS1-mediated active DNA demethylation. *Nat. Plants* **9**, 271–279 (2023).
20. Ono, A. *et al.* A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *Plant J.* **71**, 564–574 (2012).
21. Choi, Y. *et al.* DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* **110**, 33–42 (2002).

22. Kermicle, J. L. Imprinting of gene of action in maize endosperm. in *Maize Breeding and Genetics* (ed. Walden, D. B.) 357–371 (John Wiley & Sons, Inc, 1978).
23. Kermicle, J. Location, time of action, and dominance relations of an imprintor of R-mottled maize. in *Modification of gene expression and non-mendelian inheritance* (ed. Oono & Takaiwa) 120–134 (National Institute of Agrobiological Resources (Japan), 1995).
24. Xu, Q. *et al.* DNA demethylation affects imprinted gene expression in maize endosperm. *Genome Biol.* **23**, 77 (2022).
25. Ando, A., Kirkbride, R. C., Qiao, H. & Chen, Z. J. Endosperm and Maternal-specific expression of EIN2 in the endosperm affects endosperm cellularization and seed size in Arabidopsis. *Genetics* **223**, (2023).
26. Haun, W. J. *et al.* Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. *Plant J.* **49**, 325–337 (2007).
27. Tonosaki, K. *et al.* Mutation of the imprinted gene OsEMF2a induces autonomous endosperm development and delayed cellularization in rice. *Plant Cell* **33**, 85–103 (2021).
28. Gehring, M. *et al.* DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495–506 (2006).
29. Hermon, P., Srilunchang, K.-O., Zou, J., Dresselhaus, T. & Danilevskaya, O. N. Activation of the imprinted Polycomb Group Fie1 gene in maize endosperm requires demethylation of the maternal allele. *Plant Mol. Biol.* **64**, 387–395 (2007).
30. Park, K. *et al.* DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 15138–15143 (2016).
31. Williams, B. P., Bechen, L. L., Pohlmann, D. A. & Gehring, M. Somatic DNA demethylation generates tissue-specific methylation states and impacts flowering time. *Plant Cell* **34**, 1189–1206 (2022).
32. Halter, T. *et al.* The Arabidopsis active demethylase ROS1 cis-regulates defence genes by erasing DNA methylation at promoter-regulatory regions. *Elife* **10**, (2021).

33. Hemenway, E. A. & Gehring, M. The 5-methylcytosine DNA glycosylase ROS1 prevents paternal genome hypermethylation in Arabidopsis endosperm. *Genome Biol.* **26**, 1–27 (2025).
34. Borg, M. *et al.* Epigenetic reprogramming rewires transcription during the alternation of generations in Arabidopsis. *Elife* **10**, (2021).
35. Khouider, S. *et al.* Male fertility in Arabidopsis requires active DNA demethylation of genes that control pollen tube function. *Nat. Commun.* **12**, 410 (2021).
36. Wang, M. *et al.* Histone H3 lysine 4 methylation recruits DNA demethylases to enforce gene expression in Arabidopsis. *Nat. Plants* **11**, 206–217 (2025).
37. Tang, K., Lang, Z., Zhang, H. & Zhu, J.-K. The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat. Plants* **2**, 16169 (2016).
38. Frost, J. M. *et al.* FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4720–E4729 (2018).
39. Zhang, C. *et al.* The catalytic core of DEMETER guides active DNA demethylation in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 17563–17571 (2019).
40. Zeng, Y., Dawe, R. K. & Gent, J. I. Natural methylation epialleles correlate with gene expression in maize. *Genetics* **225**, (2023).
41. Swentowsky, K. W. *et al.* Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev* **34**, 1239–1251 (2020).
42. Kawakatsu, T. *et al.* Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. *Cell* **166**, 492–505 (2016).
43. Muyle, A. M., Seymour, D. K., Lv, Y., Huettel, B. & Gaut, B. S. Gene Body Methylation in Plants: Mechanisms, Functions, and Important Implications for Understanding Evolutionary Processes. *Genome Biol Evol* **14**, (2022).
44. Schmitz, R. J. *et al.* Patterns of population epigenomic diversity. *Nature* **495**, 193–198

(2013).

45. Zeng, Y. *et al.* Potent pollen gene regulation by DNA glycosylases in maize. *Nat. Commun.* **15**, 8352 (2024).
46. Chen, J. *et al.* Dynamic Transcriptome Landscape of Maize Embryo and Endosperm Development. *Plant Physiology* **166**, 252 (2014).
47. Higgins, K., Nyabashi, V. & Anderson, S. Conservation of imprinted expression across genotypes is correlated with consistency of imprinting across endosperm development in maize. *G3 (Bethesda)* **15**, (2025).
48. Zhan, J. *et al.* RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* **27**, 513–531 (2015).
49. Gómez, E. *et al.* The maize transcription factor myb-related protein-1 is a key regulator of the differentiation of transfer cells. *Plant Cell* **21**, 2022–2035 (2009).
50. Serna, A. *et al.* Maize endosperm secretes a novel antifungal protein into adjacent maternal tissue. *Plant J.* **25**, 687–698 (2001).
51. Royo, J. *et al.* Two maize END-1 orthologs, BETL9 and BETL9like, are transcribed in a non-overlapping spatial pattern on the outer surface of the developing endosperm. *Front. Plant Sci.* **5**, 180 (2014).
52. Fang, C. *et al.* A systematic investigation of lipid transfer proteins involved in male fertility and other biological processes in maize. *Int. J. Mol. Sci.* **24**, 1660 (2023).
53. Costa, L. M. *et al.* Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr. Biol.* **22**, 1980 (2012).
54. Magnard, J. L. *et al.* Genes normally expressed in the endosperm are expressed at early stages of microspore embryogenesis in maize. *Plant Mol. Biol.* **44**, 559–574 (2000).
55. Magnard, J.-L. *et al.* ZmEBE genes show a novel, continuous expression pattern in the central cell before fertilization and in specific domains of the resulting endosperm after

fertilization. *Plant Mol Biol* **53**, 821–836 (2003).

56. Bonello, J.-F. *et al.* Esr proteins are secreted by the cells of the embryo surrounding region. *J. Exp. Bot.* **53**, 1559–1568 (2002).
57. Muñiz, L. M. *et al.* Atypical response regulators expressed in the maize endosperm transfer cells link canonical two component systems and seed biology. *BMC Plant Biol* **10**, 84 (2010).
58. Yuan, Y. *et al.* Decoding the gene regulatory network of endosperm differentiation in maize. *Nat Commun* **15**, 34 (2024).
59. Zhang, Z., Zheng, X., Yang, J., Messing, J. & Wu, Y. Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. *Proc Natl Acad Sci U S A* **113**, 10842–10847 (2016).
60. Lund, G., Ciceri, P. & Viotti, A. Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of Zea mays L. *Plant J* **8**, 571–581 (1995).
61. Zemach, A. *et al.* Local DNA hypomethylation activates genes in rice endosperm. *Proceedings of the National Academy of Sciences* **107**, 18729–18734 (2010).
62. Yan, Y. *et al.* A new demethylase gene, OsDML4, is involved in high temperature-increased grain chalkiness in rice. *J Exp Bot* **73**, 7273–7284 (2022).
63. Gutiérrez-Marcos, J. F. *et al.* maternally expressed gene1 Is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* **16**, 1288–1301 (2004).
64. Baroux, C., Pecinka, A., Fuchs, J., Schubert, I. & Grossniklaus, U. The triploid endosperm genome of Arabidopsis adopts a peculiar, parental-dosage-dependent chromatin organization. *Plant Cell* **19**, 1782–1794 (2007).
65. Hsieh, T.-F. *et al.* Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**, 1451–1454 (2009).
66. Moreno-Romero, J., Jiang, H., Santos-González, J. & Köhler, C. Parental epigenetic

asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *The EMBO Journal* (2016) doi:10.15252/embj.201593534.

67. Higgins, K., Gent, J. & Anderson, S. MDR1 DNA glycosylase regulates the expression of genomically imprinted genes and Helitrons. *Plant J* **123**, e70304 (2025).
68. Immler, S. Haploid selection in 'diploid' organisms. *Annu. Rev. Ecol. Evol. Syst.* **50**, 219–236 (2019).
69. Khouider, S. & Gehring, M. Parental dialectic: Epigenetic conversations in endosperm. *Curr. Opin. Plant Biol.* **81**, 102591 (2024).
70. Wilkins, J. F. & Haig, D. What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* **4**, 359–368 (2003).
71. Dilkes, B. P. & Comai, L. A differential dosage hypothesis for parental effects in seed development. *Plant Cell* **16**, 3174–3180 (2004).
72. Chettoor, A. M., Phillips, A. R., Coker, C. T., Dilkes, B. & Evans, M. M. S. Maternal gametophyte effects on seed development in maize. *Genetics* **204**, 233–248 (2016).
73. Kinoshita, Y. *et al.* Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* **49**, 38–45 (2007).
74. Sidorenko, L. V., Chandler, V. L., Wang, X. & Peterson, T. Transcribed enhancer sequences are required for maize p1 paramutation. *Genetics* **226**, (2024).
75. Williams, B. P., Pignatta, D., Henikoff, S. & Gehring, M. Methylation-sensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. *PLoS Genet* **11**, e1005142 (2015).
76. Henderson, I. R. & Jacobsen, S. E. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* **22**, 1597–1606 (2008).
77. Belele, C. L. *et al.* Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet* **9**, e1003773 (2013).

78. Deans, N. C. *et al.* Paramutation at the maize pl1 locus is associated with RdDM activity at distal tandem repeats. *PLoS Genet* **20**, e1011296 (2024).

79. Wittmeyer, K. *et al.* The Dominant and Poorly Penetrant Phenotypes of Maize Are Caused by DNA Methylation Changes at a Linked Transposon. *Plant Cell* **30**, 3006–3023 (2018).

80. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

81. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).

82. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).

83. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

84. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

85. Portwood, J. L., 2nd *et al.* MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res* **47**, D1146–D1154 (2019).

86. Guo, W. *et al.* CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**, 381–387 (2018).