The maize gene *maternal derepression of r1* (*mdr1*) encodes a DNA glycosylase with maternal and paternal fertility functions

Jonathan I. Gent, Kyle W. Swentowsky, Kaitlin M. Higgins, Fang-Fang Fu, Yibing Zeng, Dong won Kim, R. Kelly Dawe, Nathan M. Springer, and Sarah N. Anderson

Corresponding author: gent@uga.edu

ABSTRACT

Demethylation of transposons can activate expression of nearby genes and cause imprinted gene expression in endosperm, and it is hypothesized to lead to expression of transposon siRNAs that reinforce silencing in the next generation through transfer either into egg or embryo. Here we describe *maternal derepression of R1* (*mdr1*), a DNA glycosylase with homology to Arabidopsis *DEMETER* that is partially responsible for demethylation of thousands of regions in endosperm. Maternally-expressed imprinted genes were enriched strongly enriched for overlap with demethylated regions, but the majority of genes that overlapped demethylated regions were not imprinted. Demethylated regions were depleted from the majority of repetitive DNA in the genome but enriched in a set of transposon families accounting for about a tenth of the total demethylated regions. Demethylated regions produced few siRNAs and were not associated with excess CHH methylation in endosperm or other tissues. *mdr1* and its close homolog *dng102* are essential factors in maternal and paternal fertility in maize, as neither double mutant microgametophytes nor megagametophytes gave rise to seeds. These data establish DNA demethylation by glycosylases as essential in maize endosperm and pollen and suggest that neither transposon regulation nor genomic imprinting are its main function.

INTRODUCTION

In 1970, Jerry Kermicle reported that the *red-color* (*R*) gene required for pigmentation in maize endosperm was preferentially expressed from maternal alleles (Kermicle 1970). Since then, other cases of genomically imprinted genes have been reported for hundreds of genes in maize, other plants, and animals as well [reviewed in (Batista and Köhler 2020)]. Research primarily on *Arabidopsis thaliana* endosperm has revealed that reduction of DNA methylation of the maternal genome can differentiate maternal and paternal alleles. In Arabidopsis, the key enzyme is a DNA glycosylase called DEMETER, which excises 5-methylcytosine from DNA and is primarily active in the central cell of the megagametophyte, which gives rise to the endosperm upon fertilization (Choi et al. 2002; Gehring et al. 2006; Park et al. 2017). In the pollen vegetative nucleus, where DEMETER is also highly expressed, many transposons are demethylated and expressed (He et al. 2019). DNA methylation not only represses transposons, but can also regulate developmental gene expression, presumably through inhibiting or promoting transcription factor binding (O'Malley et al. 2016; Borg et al. 2021; Khouider et al. 2021). Absence of maternal DEMETER causes 100% seed inviability (Choi et al. 2002). Pollen that lack DEMETER give rise to apparently normal seeds, but have reduced transmission frequent relative to wild-type and have pollen tube defects, severity dependent on the ecotype (Xiao et al. 2003; Schoft et al. 2011; Khouider et al. 2021). In rice, maternal glycosylase (ROS1a/OsDNG702) is required for seed development, and paternal glycosylase is required for pollen, such that null mutations cannot be transmitted maternally or paternally. The Arabidopsis genome encodes three other glycosylases whose mutants produce subtle phenotypes in a variety of tissues (Pillot et al. 2010; Yamamuro et al. 2014; Schumann et al. 2017; Schumann et al.

2019; Khouider et al. 2021). Endosperm derived from central cells that lack DEMETER have excessive DNA methylation at thousands of small regions (Ibarra et al. 2012; Park et al. 2016). Unifying features of these regions have not been identified, but they exhibit a preference for short transposons in genic areas and for the ends of long transposons. What guides DEMETER to these regions is unknown, but involves FACT, a versatile ATP-dependent chromatin remodeling complex (Frost et al. 2018). DEMETER depends on FACT for activity at about half of the ~10,000 identified target regions, which are distinguished by being more heterochromatic than the other half. The N-terminal half of DEMETER is likely important for targeting, as truncation of the protein decreases its activity at normal targets and leads to increased activity in gene bodies (Zhang et al. 2019).

In addition to maternal demethylation by DEMETER in specific regions, there is a pronounced and dynamic global reduction in DNA methylation of both parental genomes in endosperm relative to typical vegetative tissues in Arabidopsis (Ibarra et al. 2012; Moreno-Romero et al. 2016). A similar phenomenon has been observed in other plants, e.g., rice and in *Brassica napa* (Park et al. 2016; Chakraborty et al. 2021). This global reduction in methylation is not caused by DNA glycosylase activity, but instead by passive demethylation through DNA replication and reduced expression of multiple components of RNA-directed DNA methylation (RdDM) and the replication-coupled methyltransferase MET1 in early endosperm or central cell (Jullien et al. 2012; Belmonte et al. 2013; Kawakatsu et al. 2017). RdDM activity apparently increases later in endosperm development as evidenced by an increase in DNA methylation in all sequence contexts during after cellularization (Moreno-Romero et al. 2016). Polycomb Repressive Complex2 (PRC2) promotes non-CG methylation in endosperm, but even so, non-CG methylation in endosperm remains at greatly reduced levels relative to other tissues (Ibarra et

al. 2012). Perhaps related, PRC2 functions in imprinting of many genes independently of DEMETER [reviewed in (Batista and Köhler 2020)]. How much of endosperm demethylation is determined by the central cell is unclear. The central cell is unusual and distinct from endosperm in its lack of H3K9me2 and visible chromocenters (Garcia-Aguilar et al. 2010; Pillot et al. 2010; Yelagandula et al. 2014). It has apparently normal methylation outside of DEMETER target regions, however (Park et al. 2016).

Maternal demethylation of specific transposons can activate nearby gene expression [reviewed in (Anderson and Springer 2018; Batista and Köhler 2020)]. Studies in multiple species have revealed a high level of variation in which genes are imprinted, including within-species variation, and in some cases imprinting is associated with presence of transposons near genes (Gehring et al. 2009; Waters et al. 2013; Pignatta et al. 2014; Hatorangan et al. 2016; Pignatta et al. 2018; Rodrigues et al. 2021). Demethylation of transposons in the central cell is hypothesized to reinforce transposon silencing in the embryo (Ibarra et al. 2012; Bouyer et al. 2017). In this model, transposon activation in produces siRNAs in central cell or endosperm that move into the egg cell or embryo. Likewise, DEMETER activity in the pollen vegetative cell could silence transposons in the sperm cells. (Calarco et al. 2012; Ibarra et al. 2012; Martínez et al. 2016). Methylation patterns in rice *ros1a* mutant pollen also support this hypothesis (Kim et al. 2019).

While work on gene regulation and chromatin in Arabidopsis endosperm has provided many insights into maize as well, maize is different than Arabidopsis in several key aspects. Maize lacks a CMT2-type chromomethyltransferase, causing CHH methylation (mCHH, where H is A, T, or C) to be largely depleted from heterochromatin and instead associated specifically with RdDM (Zemach et al. 2013; Gent et al. 2014). It has a much larger set of transposons,

including ones in and near genes and currently active ones (McCarty et al. 2005; Springer et al. 2018; Dooner et al. 2019). Its endosperm persists past fertilization and has a distinct cellular morphology (Becraft and Gutierrez-Marcos 2012). Unlike both Arabidopsis and rice endosperm, maize endosperm in general is highly methylated both early and late in development (Zhang et al. 2014; Fu et al. 2018). Kermicle's demonstration that certain *R* haplotypes are activated specifically from the maternal genome in endosperm suggests that similar imprinting mechanisms function in maize as in rice and Arabidopsis. The phenomenon Kermicle studied is known as R mottling because of the pigmentation pattern. Maternal *R* is uniformly active and yields solidly pigmented endosperm. When maternal *R* is absent, non-uniform and weak expression of paternal *R* causes a mottling pattern. Kermicle also reported that mottling can occur in the presence of maternal *R* in a mutant he called *mdr* (for *maternal derepression of R*) (Kermicle 1995) (Figure 1A). He mapped the mutation to the long arm of chromosome 4, found it to be homozygous fertile, and demonstrated that maternal wild-type *Mdr* alone is required and sufficient for uniform *R* activation in endosperm (Kermicle 1995). In this study, we mapped *mdr* to a single gene that encodes one of two glycosylases with endosperm expression in maize and demonstrated that it is required for full demethylation of thousands of diverse regions in endosperm.

RESULTS

The *maternal derepression of R1* (*mdr1*) gene encodes a DNA glycosylase.

Figure 1

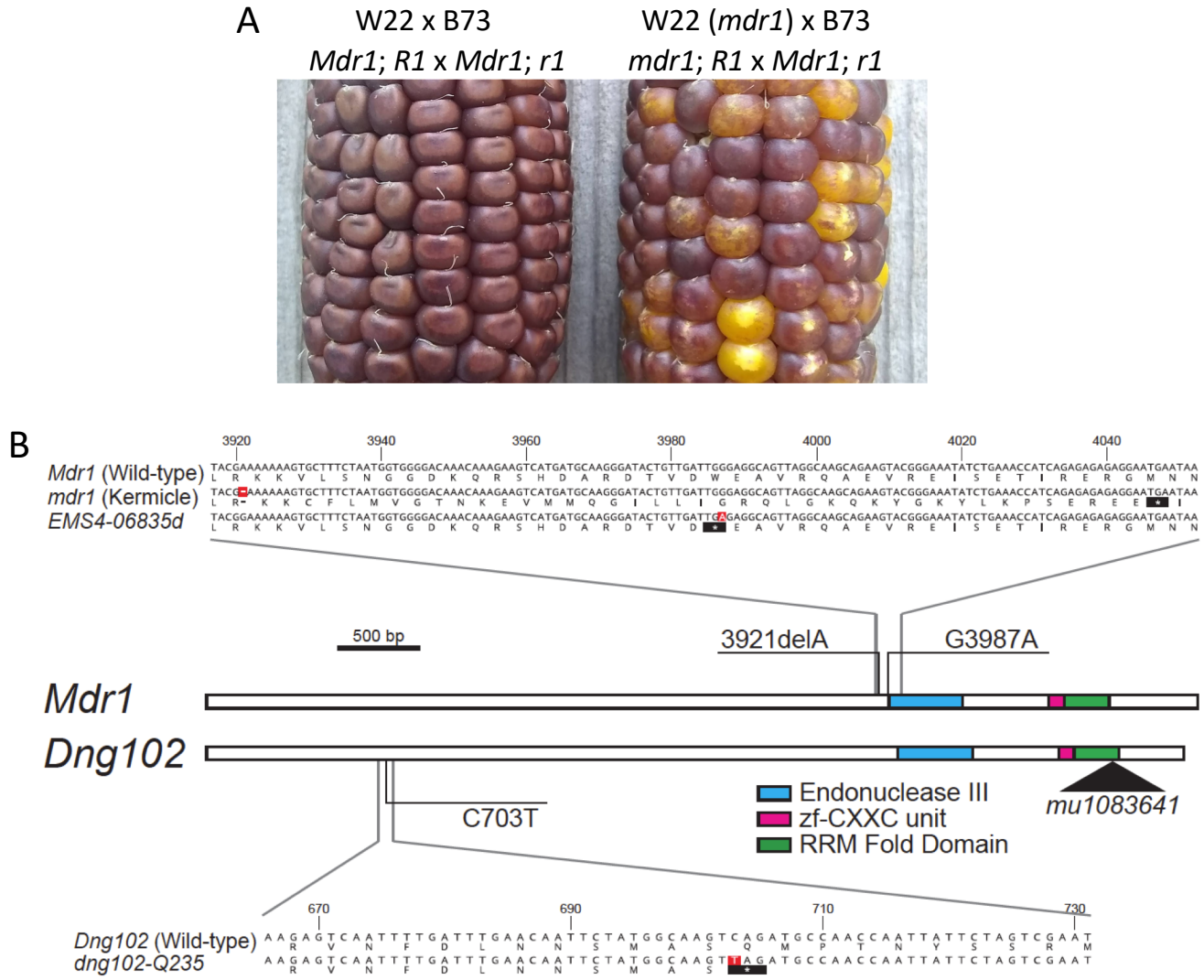

Figure 1 Maternal *R1* expression phenotype and endosperm DNA glycosylase alleles
A, Kermicle's *mdr1* allele produces a maternal R mottling phenotype. In crosses where functional *R1* is inherited maternally only, maternal *Mdr1* is required for full *R1* expression. W22 (*mdr1*) is the *mdr1* mutant in a full-color W22 background. All listed alleles are homozygous.
B, Sequence and structure of *mdr1* and *dng102*. Nucleotides highlighted in red indicate positions of mutations. Asterisks in black background indicate positions of stop codons. Black triangle indicates position of Mu insertion.

Consistent with current nomenclature in which the original *red-color* gene *R* was changed to *R1*, we refer to *mdr* as *mdr1* (https://www.maizegdb.org/nomenclature). To map *mdr1*, we took advantage of the fact that *mdr1* originated spontaneously in a W22 inbred stock (Kermicle, personal communication). We crossed an *mdr1* homozygote with the sequenced W22 stock (Springer et al. 2018) to generate *mdr1*/*Mdr1* heterozygotes. We then crossed heterozygotes as females with *mdr1* homozygotes as males to produce 50% homozygous and 50% heterozygous *mdr1*, which we expected to correspond to mottled and solid kernels. After DNA extraction in bulk from ~500 plants grown from mottled kernels, we prepared and sequenced Illumina libraries. For comparison, we also sequenced a single pool of five individuals of the W22 stock. A search for non-W22 reference mutations that were close to 100% homozygous from the mottled kernels revealed a single region on the long arm of chromosome 4 spanning approximately 20 MB (Supplemental Figure S1). Several regions of the genome were not shared with W22 and had abundant non-W22 variants. The long arm of chromosome 4, however, was nearly identical to W22 (Supplemental Figure S2). In fact, within the 20 MB region, there were only 45 SNPs or short indels that passed quality filters, and only one of them was in an exon (*GRMZM2G422464*). This was a deletion of a single T within a string of seven T's starting at position 229,617,698 and created a frameshift mutation upstream of predicted Endonuclease III and RRM-fold domains that are typical of DEMETER and ROS family glycosylases. (Figure 1B). To confirm that the mutation in *GRMZM2G422464* caused the maternal R mottling phenotype, we also obtained an EMS allele (EMS4-06835d) from the Lu et al (2018) Gene-Indexed Mutations in Maize collection (Lu et al. 2018), which caused a C to T nucleotide change in the same exon (at position 229,617,638), producing a premature stop codon. After backcrossing this allele four times into W22 to create an *EMS4-06835d* heterozygote in a W22

background, we tested for a mottling phenotype by crossing the heterozygote as the female parent with the *Mdr1* homozygous B73 inbred stock as male. Genotyping of *EMS4-06835d* in plants grown from these kernels revealed 20 of 20 plants from the strongest mottling kernels inherited the mutation while only 9 of 28 from solid kernels did (two tailed p-value < 0.0001, Fishers Exact test). These results provide independent evidence that mutations in *GRMZM2G422464* cause maternal R mottling. Hence we refer to *GRMZM2G422464* as *mdr1*.

Functional *mdr1* or its homolog *dng102* is required for male and female fertility

Gramene paralog analysis (http://www.gramene.org) revealed 4 paralogs of *mdr1,* but only one with close sequence homology (80% amino acid identity) (Supplemental Table S1), *GRMZM2G123587*, also known as *dng102* (for *DNA glycosylase102*). Of the five genes, only *mdr1* and *dng102* have strong expression in endosperm, with *mdr1* higher than *dng102,* including late in endosperm development [displayed on maizeGDB, using data from prior expression studies (Stelpflug et al. 2016; Doll et al. 2020)]. Both of these have ~32% amino acid identity to DEMETER (Supplemental Table S1) and cluster with DEMETER by phylogenetic analysis (Kapazoglou et al. 2013). The Gramene paralog analysis also revealed a similar pair of genes in the sorghum genome with one-to-one matches for *mdr1* (*SORBI_3004G149800*) and *dng102* (*SORBI_3008G085300*), indicating conservation of both genes. We obtained two mutant alleles of *dng102*, the UniformMu insertion *mu1083641* (McCarty et al. 2005) in the C-terminal RRM-fold domain and a TILLING allele (Till et al. 2004) with a premature stop codon upstream of both the Endonuclease III domain and the RRM-fold domain. Neither allele produced a maternal R mottling phenotype, nor any obvious phenotype of any kind. However, *mdr1 dng102* double heterozygotes failed to transmit either the UniformMu insertion or the TILLING allele of

dng102 simultaneously with the *mdr1* mutant allele, neither maternally or paternally. We took advantage of the Ds-GFP insertion collection to use GFP fluorescence to screen large numbers of kernels for mutation transmission frequency (Li et al. 2013; Warman et al. 2020). By creating a *mdr1 dng102* double heterozygote with Ds-GFP insertions closely linked to each wild-type allele and crossing to non-GFP wild-type plants, we were able to screen thousands of kernels for paternal Ds-GFP inheritance. All but six of 3,786 resulting kernels were florescent. PCR genotyping of four of the six revealed wild-type *Dng102*, *Mdr1*, or both. The remaining two kernels failed to germinate. These six exceptions are likely explained by pollen contamination or recombination between Ds-GFP insertions and mutations. Evenly filled rows of kernels from these crosses indicated a pre-fertilization pollen defect (Supplemental Figure S3). In a control cross with Ds-GFP insertions and wild-type *Dng102* and *Mdr1* alleles, 59 of 249 kernels lacked florescence. We also verified a requirement for maternal glycosylases by crossing the *mdr1 dng102* double heterozygotes with both Ds-GFP insertions by wild-type pollen. All of 574 resulting kernels in four crosses were florescent, indicating complete failure to transmit mutant *mdr1* and *dng102* alleles to the seed simultaneously. In all four crosses, a high frequency of gaps in the ears indicated an early defect in female gametophyte development or seed development (Supplemental Figure S3).

Identification of high-confidence sets of genomic regions that are demethylated in endosperm

To determine whether the *mdr1* mutant has detectable changes in DNA methylation relative to wild-type, we prepared and sequenced EM-seq libraries and searched for differentially methylated regions (DMRs) between homozygous wild-type and *mdr1* mutant mature

endosperm. In parallel, we also searched for DMRs between wild-type mature endosperm and wild-type developing tassel, and between wild-type mature endosperm and wild-type developing leaf. In order to identify regions with high-confidence differential methylation, we used a conservative approach that required differential methylation of both mCG and mCHG independently over non-overlapping 200-bp regions. First we identified the set of all regions that were eligible for differential methylation analysis based on their having at least 3X average coverage of at least five CGs and five CHGs in both tissues. From this set of eligible regions, we identified hyperDMR as having a greater than twofold relative increase in percent methylation (value2/value1 > 2) and a greater than 20% relative increase in methylation (value2 – value1 > 0.20) for both mCG and mCHG. HypoDMRs satisfied the opposite requirements. We used the entire set of eligible regions as controls for DMRs, since they were subject to the same biases in terms of ability to map EM-seq reads and containing informative CG and CHGs. In the case of the wild-type vs. *mdr1* endosperm analyses, these eligible region comprised about a third of the genome. For endosperm vs. tassel or endosperm vs. leaf, they made up about a fifth of the genome due to lower sequencing depth.

Consistent with expectations, in the wild-type vs. *mdr1* mutant endosperm analysis, nearly tenfold more DMRs had lower methylation (hypoDMRs) than higher methylation (hyperDMRs), comprising about three out of every thousand eligible regions (Figure 2A). The endosperm vs. tassel comparison yielded the strongest enrichment for hypoDMRs, seventyfold over hyperDMRs, making up nearly nine of every thousand eligible regions. As the hyperDMRs are likely indirect effects of endosperm demethylation or experimental noise, we excluded them from additional analysis, and for brevity we refer to hypoDMRs in endosperm simply as DMRs. We also compared wild-type tassel with *mdr1* mutant in tassel, which yielded few DMRs, of

Figure 2

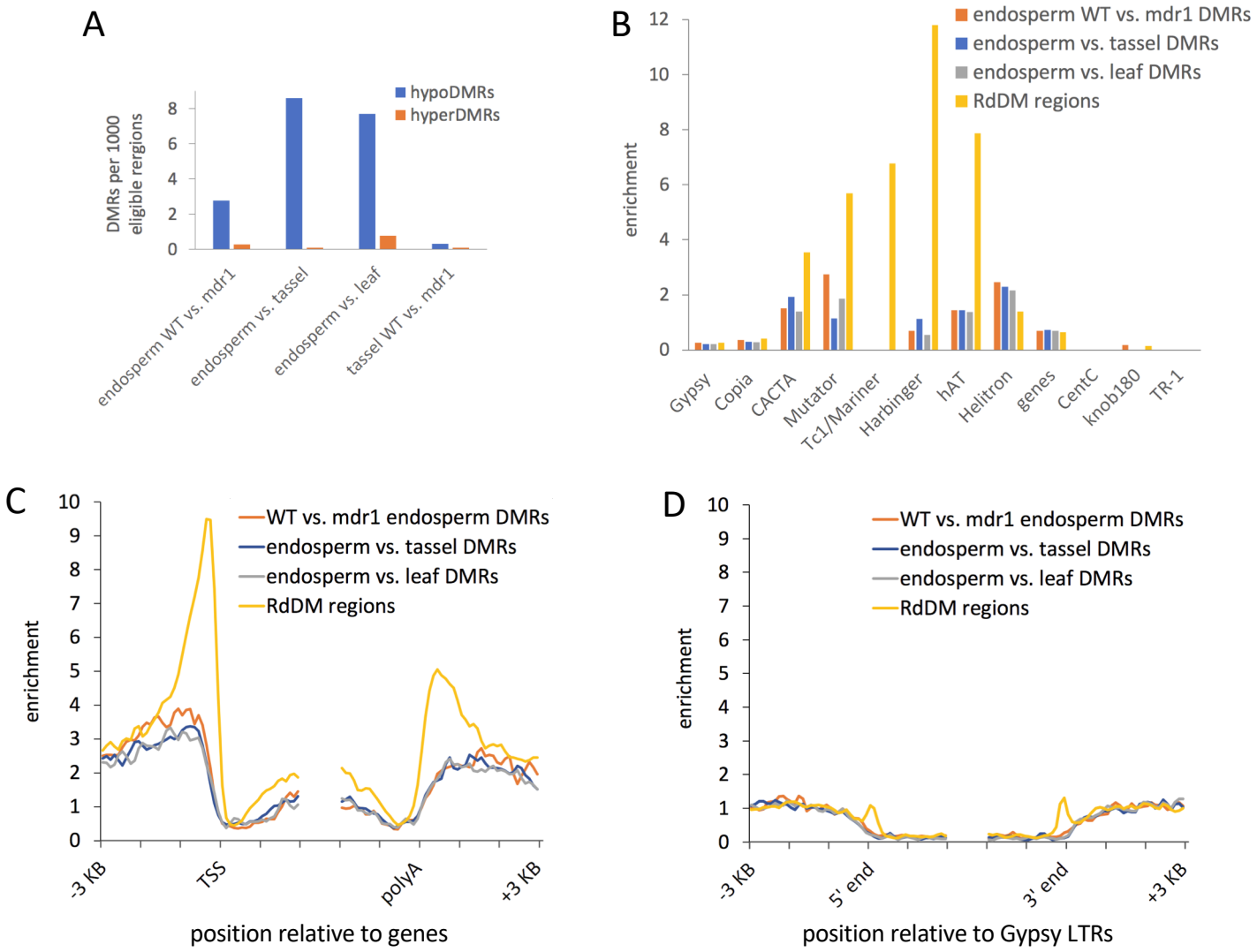

Figure 2 Genomic locations of endosperm DMRs

A, The majority of differentially methylated regions are hypomethylated in endosperm. Eligible regions are defined in the same way as DMRs in terms of read coverage and numbers of informative cytosines, but with no requirement for differential methylation. Endosperm is from mature kernels segregating homozygous wild-type and homozygous *mdr1* mutant. Tassel is premeiotic stage, ~1.5 cm in length, segregating homozygous wild-type and homozygous *mdr1* mutant. Leaf is 2nd leaf prior to emergence through 1st leaf. The first three comparisons use the same wild-type endosperm. The fourth uses the same wild-type tassel as the endosperm vs. tassel comparison, but compares homozygous wild-type tassel to sibling *mdr1* tassel.

B, Endosperm DMRs are depleted from Gypsy and Copia and weakly enriched in CACTA, Mutator, and Helitron transposon superfamilies. DMRs include only the hypoDMRs of **A**. Enrichment is defined as the number of overlaps between DMRs and different genetic normalized by the number of overlaps between eligible regions and the same genetic elements. RdDM regions are defined by average mCHH values of greater than 0.2 in tassel.

C, Endosperm DMRs are weakly enriched near genes. Enrichment values on the Y-axis are the sum of the DMR coverage or RdDM region coverage normalized by eligible region coverage of each 100-bp interval for each gene up to 3 Kb outside the genes and 2 Kb inside. TSS: transcription start site. polyA: polyadenylation site. Tick marks on the X-axis are on 1-Kb intervals.

D, Endosperm DMRs are strongly depleted in the Gypsy superfamily, as in C, except using edges of intact Gypsy retrotransposons rather than genes.

which only 0.3 of 1000 eligible regions were hypoDMRs (Figure 2A). While the constraints on depth of sequencing prevented us from defining an exhaustive list of DMRs even in unique portions of the genome, DMRs defined by the *mdr1* mutant endosperm should overlap at a high frequency with DMRs defined by comparison of endosperm with tassel or leaf. Indeed, 78.5% of the DMRs from the comparison with *mdr1* were within a kilobase of a DMR identified from the comparison with tassel or leaf. This was a 24-fold increase over the random expectation provided by eligible regions (3.3% of which were within a kilobase of these DMRs).

Endosperm demethylation is mildly enriched near euchromatin and is distinct from RdDM

To determine what types of genetic elements are demethylated in endosperm, we counted overlaps (at least 100bp) between different types of annotated features and the three sets of endosperm DMRs (vs. *mdr1* endosperm WT tassel, WT leaf). As controls we counted overlaps with all eligible regions. Over one-third of the DMRs overlapped with an annotated transposon, while approximately two-thirds of eligible regions did. The abundant LTR retrotransposons, both Gypsy and Copia superfamilies, were the most strongly depleted of DMRs (Figure 2B). Regions containing centromeric (*CentC*) and knob (*knob180* and *TR-1*) tandem repeats were also strongly depleted from endosperm DMRs (Figure 2B). This depletion was not because of inability to map reads to tandem repeat regions, as thousands of the control eligible regions overlapped with them. No transposon superfamilies were strongly enriched for DMRs, but Helitrons had the highest enrichment, more than twofold relative to eligible regions. To determine whether endosperm demethylation is linked to RdDM in other tissues, we identified 200-bp regions of the genome undergoing RdDM using mCHH in the tassel methylome as the defining feature. Specifically, we required a minimum of 3X coverage, at least 5 informative cytosines in the

CHH context, and an average mCHH value of greater than 20% over the region. Unlike endosperm DMRs, RdDM regions exhibited a clear pattern of stronger enrichment near TIR DNA transposons (Figure 2B). An analysis at the transposon family level identified 75 LTR, 11 Helitron, and 9 TIR families that were enriched (p-value < 0.01, binomial test) for DMRs out of 25,730 families that overlapped with at least one eligible region (Supplemental Figure S4). The majority of these families (63) were annotated in the RLX grouping, which means they were retrotransposons but could not be assigned to either Copia or Gypsy. Together the 95 families accounted for 11.6% of total endosperm vs. tassel DMRs and 1.8% of control eligible regions.

Gene bodies were slightly depleted of DMRs (Figure 2B), but regions flanking genes were enriched for multiple kilobases upstream and down (Figure 2C). The scarcity of DMRs from gene bodies, especially near their transcription start sites (TSSs) and polyadenylation sites, is not surprising given the normal absence of mCG and mCHG from these regions. When we compared DMRs with only the methylated subset of eligible regions, we found up to 7-fold enrichment for DMRs internal to both ends of genes (Supplemental Figure S5). We defined methylated eligible regions as ones with both mCG and mCHG values of at least 0.2. This indicates that gene ends that are methylated in tassel, in leaf, or in *mdr1* mutant endosperm were 7-fold more likely to overlap DMRs than the rest of the genome was. In contrast, the whole set of gene ends were about 50% less likely to overlap DMRs than the rest of the genome was. While endosperm DMRs and RdDM regions were both enriched near genes, RdDM regions were distinct in being strongly enriched in the first 500 bp upstream of genes (9.5 fold enrichment for RdDM regions vs. 3 fold for DMRs) (Figure 2C). While LTRs retrotransposons were strongly depleted of DMRs overall, the large number of LTR retrotransposons in the genome provided an opportunity to examine locations of DMRs within and near them. On average, DMR enrichment

reached levels expected for a random expectation relative to eligible regions within the first 2 Kb outside both Gypsy LTRs (Figure 2D) and Copia LTRs (Supplemental Figure S6). RdDM regions were also strongly depleted within both types of retrotransposons, but had small peaks in enrichment to approximately the random expectation immediately inside Gypsy LTRs and about twofold enrichment immediately inside Copia LTRs. Motif analysis revealed several that were significantly enriched but were only present in a small fraction of endosperm DMRs and none that were consistently enriched across the three sets of DMRs (Supplemental Figure S7). Endosperm DMRs were AT rich, on average 58.9% A + T in endosperm DMRs versus 53.9% A + T in control eligible regions.

Measuring methylation levels in and flanking endosperm DMRs confirmed their locations near euchromatin and revealed that kilobase-scale regions extending beyond the 200-bp regions defined by our method were detectably demethylated (Figure 3 and Supplemental Figure S8). For these analyses, DMRs were oriented relative to the closest euchromatin, using absence of mCHG as a proxy for euchromatin. Interestingly, *mdr1* mutant endosperm partially restored endosperm DMRs to a normal, non-endosperm state of mCG and mCHG. We also confirmed this by examining the distribution of methylation values for individual DMRs for the wild-type tissues and the *mdr1* mutant endosperm (Supplemental Figure S9). In terms of mCHH, *mdr1* caused an increase even beyond that of non-endosperm tissues, though still not much above background levels. Comparison with methylation of control eligible regions also indicated that methylation of mature endosperm is similar to somatic tissues at a genome-wide level. Overall, these data indicate that demethylation in endosperm is partially dependent on MDR1, distinct from RdDM, occurs in a limited but wide variety of genomic contexts, and has a mild preference for regions near euchromatin.

Figure 3

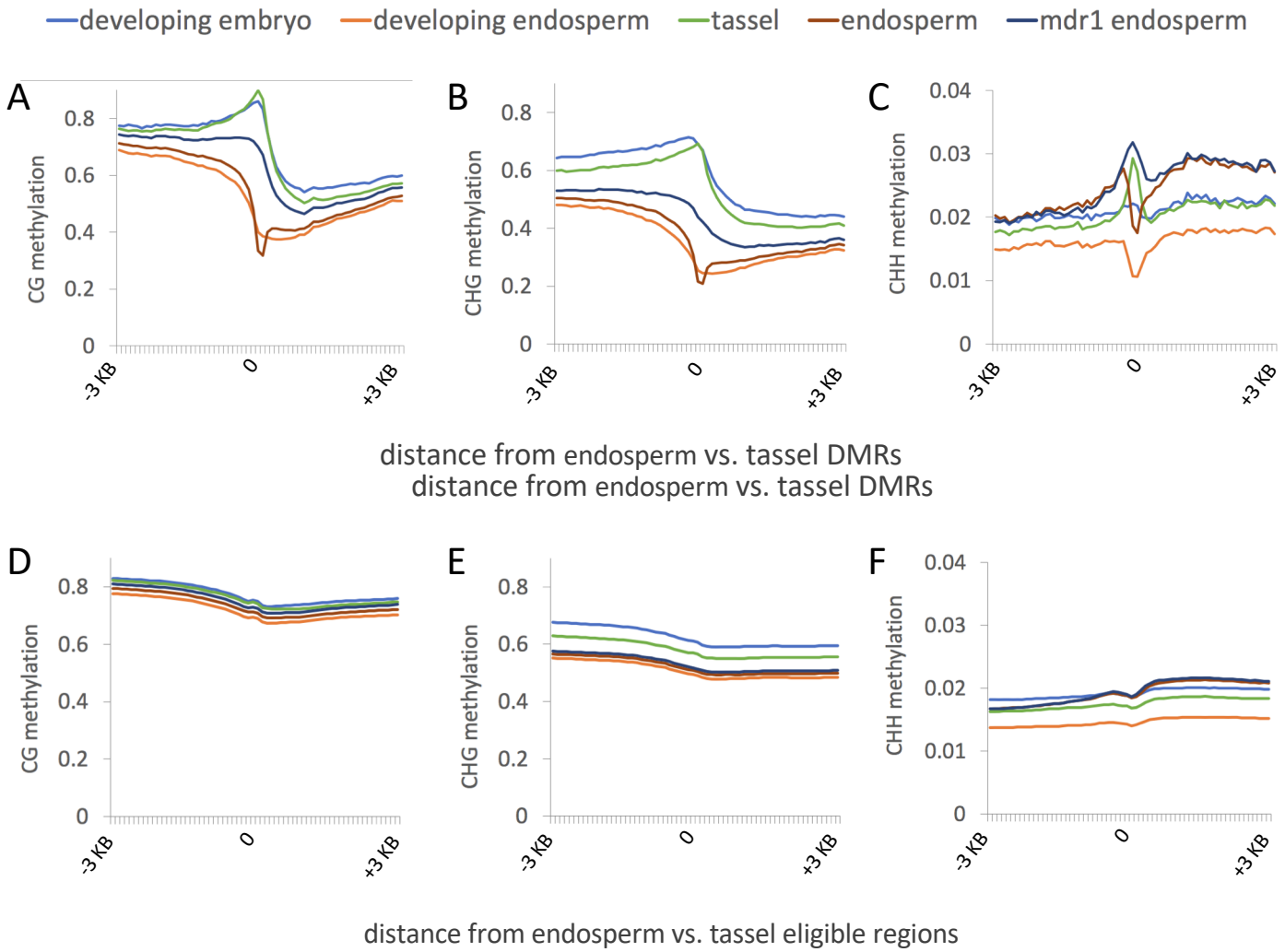

Figure 3 Mutation of *mdr1* partially reverts endosperm to typical methylation profiles
A, mCG profiles centered on endosperm vs. tassel DMRs. Methylation values are averages of 100-bp intervals within and up to 3 Kb on either side of DMRs. Orientation is determined by direction to the closest CHG unmethylated region in tassel (UMR, defined as mCHG < .2). Values for tassel and mature endosperm are from the same methylomes as described in Figure 2A. Values for developing endosperm and developing embryo are from previously published methylomes (Fu et al. 2018).
B, mCHG profiles centered on endosperm vs. tassel DMRs, as in A.
C, mCHH profiles centered on endosperm vs. tassel DMRs, as in A.
D, mCG profiles centered on endosperm vs. tassel eligible regions, as in A.
E, mCHG profiles centered on endosperm vs. tassel eligible regions, as in A.
F, mCHH profiles centered on endosperm vs. tassel eligible regions, as in A

Except near euchromatin, endosperm DMR siRNAs are in low abundance in endosperm and embryo

Even though endosperm DMRs have little mCHH, it is possible that siRNAs produced by endosperm demethylation could direct other chromatin modifications or function post-transcriptionally (Parent et al. 2021). In Arabidopsis, demethylation of heterochromatin in pollen vegetative cells corresponds to production of 21 and 22nt siRNAs from LTR retrotransposons (Creasey et al. 2014; McCue et al. 2015; Borg et al. 2021). Since the preference for DMRs is to be near euchromatin, where they have the potential to overlap RdDM regions, we separated endosperm DMRs into two groups, euchromatin DMRs, which were within 2 Kb of euchromatin (10142 regions), and heterochromatin DMRs, which were not (10014 regions). For this, we used DMRs from the endosperm vs. tassel comparison as it provided the largest number of hypoDMRs (Figure 2A). In theory, demethylation-dependent siRNAs would be more clearly evident in heterochromatin DMRs, where they would not be masked by siRNAs produced by canonical RdDM. We found, however, that 21 and 22nt siRNAs were rare in both euchromatin and heterochromatin DMRs, and 24nt siRNAs were rare in heterochromatin DMRs (Figure 4A). This was true of both endosperm siRNAs as well as non-endosperm siRNAs (embryo and tassel). To quantify the number of regions that associated with siRNAs we counted all 200-bp regions with at least 50 bp spanned by mapped siRNAs). 21, 22, and 24-nt siRNA regions were each identified independently (Figure 4B). The results do not rule out the possibility of demethylation promoting 21 or 22-nt siRNA expression, but they do indicate that 21 or 22-nt siRNA expression

Figure 4

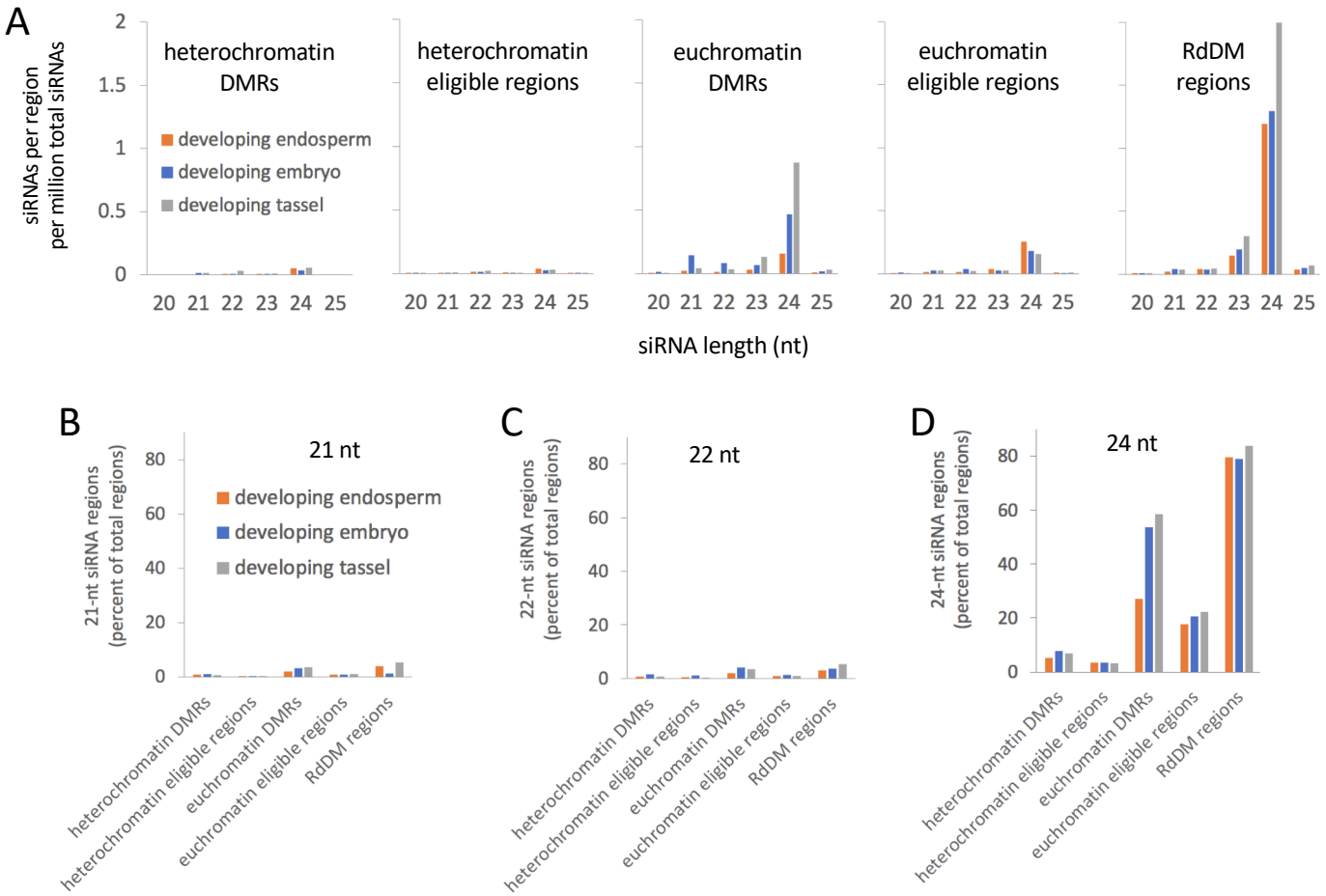

Figure 4 Endosperm DMRs are depleted of siRNAs relative to RdDM regions.
A, siRNAs length and abundance profiles. The normalized abundance of siRNAs from 20 to 25 nt in length are shown for each type of region. DMRs are from endosperm vs. tassel comparison. Heterochromatin DMRs (10014 total) are ones that are more than 2 Kb from the nearest CHG UMR. Euchromatin DMRs (10142 total) are ones that are 2 Kb or less from the nearest CHG UMR. These number include multimapping siRNAs, where one position is assigned at random among multiple equally valid positions. Developing tassel siRNAs are from the same tissue samples used for tassel methylomes. Developing embryo and developing endosperm siRNAs are from the same tissue types as their respective methylomes (Fu et al. 2018).
B, Frequency of 21-nt siRNA regions. Any region with at least 50 nt spanned by mapped siRNAs (25% of the region length) is defined as an siRNA region. These number include multimapping siRNAs, where one position is assigned at random among multiple equally valid positions. Developing tassel siRNAs are from the same tissue samples used for tassel methylomes. Developing embryo and developing endosperm siRNAs are from the same tissue types as their respective methylomes (Fu et al. 2018).
C, Frequency of 22-nt siRNA regions, as in **B**.
D, Frequency of 24-nt siRNA regions, as in **B**.

is not a typical feature of endosperm DMRs, nor is expression of 24-nt siRNAs except in the euchromatic subset of DMRs.

Endosperm DMRs imprint expression of a small subset of DMR-containing genes and transposons

Since endosperm inherits two copies of each maternal genome and one of each paternal genome, the expression of maternal transcripts would be expected to account for two thirds of the total transcript in endosperm. Reciprocal hybrid crosses, where maternal and paternal alleles can be differentiated by SNPs, provide a means to test this. The average reciprocal expression ratio (RER) of 0.67 in maize endosperm indeed corresponds to twice as high expression of alleles when inherited maternally vs. paternally (Waters et al. 2011; Zhang et al. 2011; Anderson et al. 2021). Reanalysis of published data in conjunction with endosperm DMRs reveals that the proportion of MEGs (genes with RER >.9) overlapping endosperm DMRs is more than 10-fold higher than the proportion of non-imprinted genes that overlap endosperm DMRs, and the propotion of PEGs (genes with RER < .1) overlapping DMRs is more 2-fold higher than non-imprinted genes. (Figure 5A, B). An example MEG is the transcription factor *myb155 (GRMZM2G150680)*, with a 3-Kb region overlapping its promoter that is demethylated in endosperm (Figure 5D). DMRs in MEGs were preferentially located towards their 5' ends (Figure 5E) while DMRs in genes as a whole had a more balanced representation at both ends (Figure 5F). These analyses revealed that while some genes with DMRs and expression in endosperm like *myb155* are imprinted, the majority of genes with DMRs and expression in endosperm do not appear to be imprinted. The majority of genes with DMRs clustered around

Figure 5

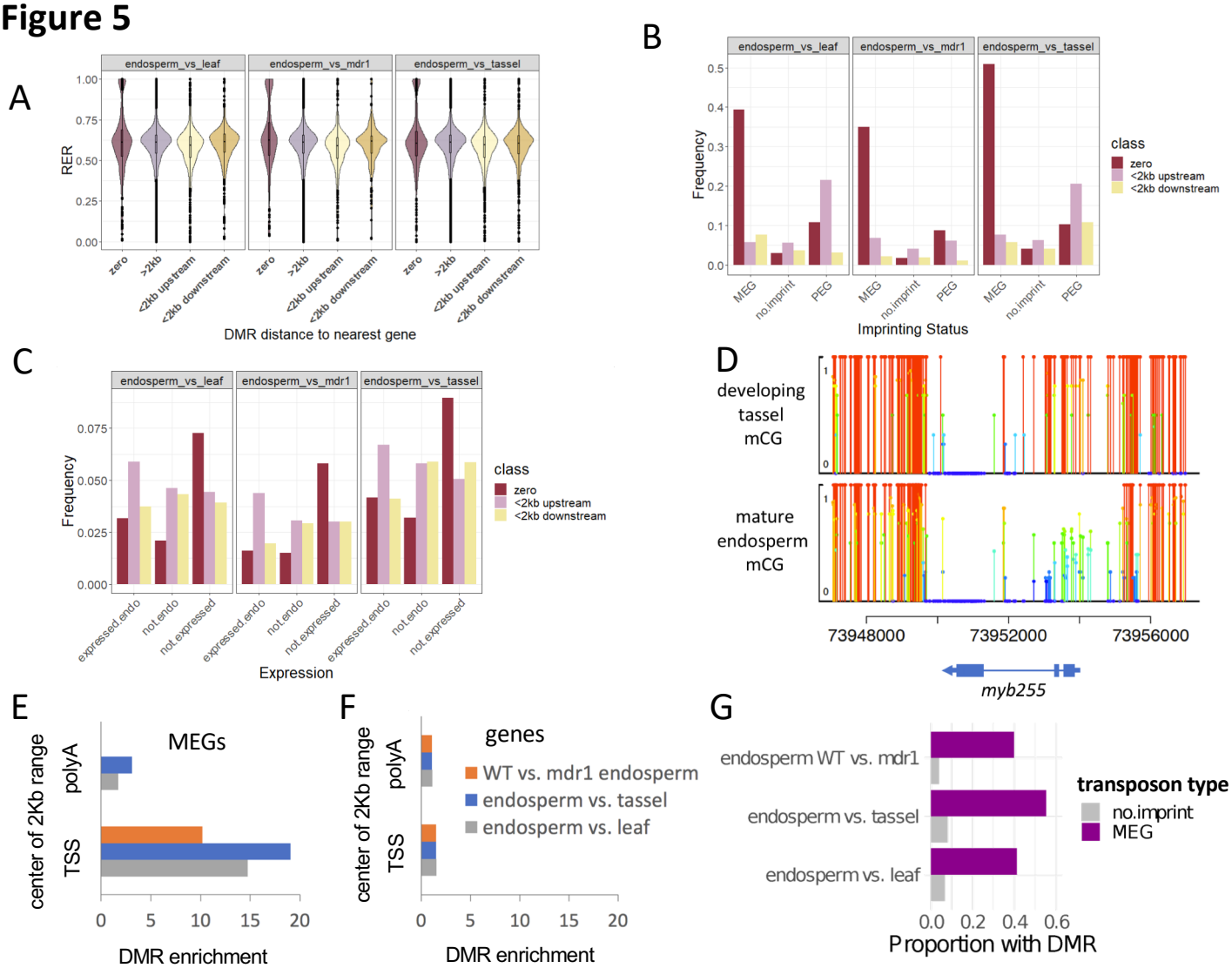

Figure 5 Endosperm DMRs are enriched in maternally-imprinted genes and transposons, but the majority of endosperm DMR genes are biparentally expressed

A, Genes that overlap endosperm DMRs are skewed toward maternal expression but still center on RER values of 0.67. RER is reciprocal expression ratio, where a value of 1 means completely maternal and 0 means completely paternal. Expression data include all genes, not just syntenic ones, from (Anderson et al. 2021). Genes are classified according to distance from endosperm DMRs. Endosperm DMRs from each of the three comparisons are shown separately. Boxes are interquartile ranges.

B, MEGs have a high frequency of overlap with endosperm DMRs. Genes are classified according distance from endosperm DMRs (by color) and according to imprinting status (X-axis labels). Y-axis indicates the proportion of all genes in each set. MEGs are defined as an RER value greater than 0.9 (117 genes), PEGs as less than 0.1 (69 genes), and not-imprinted as all intermediate values (14,450 genes).

C, DMRs do not predict endosperm gene expression. Genes are classified according distance from endosperm DMRs (by color) and by expression in endosperm (X-axis labels). Expression categories as follows: “expressed.endo” means expressed in endosperm, without regard to expression in other tissues (12,416 genes); “not.endo” means low or no expression in endosperm, but expressed in other tissues (5,019 genes); and “not expressed” means low or no expression in endosperm and other tissues (1,615 genes).

D, An example methylation profile for a MEG. mCG values are shown at single-base level resolution over a 10-Kb region of chromosome 10 including the gene *myb255* (*GRMZM2G150680*). Both height and color of lollypops indicates mCG values (darkest blue = zero mCG).

E, MEGs are enriched for endosperm DMRs near their 5’ ends. DMR enrichment is the number of endosperm DMRs that overlap with a 2-Kb range either centered on transcription start sites (TSS) or polyadenylation sites (polyA) of MEGs normalized by the number of eligible regions that overlap the same range. Endosperm DMRs from each of the three comparisons are shown separately. The absence of an orange bar in WT vs. *mdr1* is because there were no overlaps in the polyA-centered range.

F, Genes are not enriched for endosperm DMRs near their 5’ ends. Same as in **D**.

G, Maternal imprinted transposons have a high frequency of overlap with endosperm DMRs. MEG-type transposons (95 total) have an RER value greater than 0.9, PEG as less than 0.1, and non-imprinted (1773 total) as all intermediate values. The Y-axis indicates the proportion of transposons that overlap endosperm DMRs from each of the three comparisons. In all three, MEG-type transposons are significantly enriched for DMRs over non-imprinted ones (p-value < 0.01, binomial test). Only one PEG-type transposon overlapped with DMRs and is not shown.

RER scores of 0.67 even if they overlapped DMRs (Figure 5A). Genes with DMRs in either their 2-Kb flanking regions, either 5' or 3', were also enriched for MEGs and PEGs, but these flanking effects were weak and were less reproducible between the three DMR comparisons (Figure 5A, B). Endosperm expression in general did not predict overlap with DMRs (Figure 5C).

The majority of transposon copies with sufficient expression in endosperm and with variants to allow distinguishing parent of origin also have RER values near 0.67 (Anderson et al. 2021). However, we found that the 5% of transposons with MEG-type expression (RER > .9 and passing significance thresholds) are ~fivefold more likely to overlap endosperm DMRs than the endosperm-expressed but non-imprinted set (Figure 5G). Only three transposon copies (0.16% of the endosperm-expressed set) had PEG-type expression (RER <.1), too few for a meaningful analysis. These results indicate that demethylation in endosperm-expressed genes usually has weak or no effects on genomic imprinting, but in specific situations demethylation can cause maternal-specific expression of both genes and transposons.

Maternally mutant *mdr1* causes reduced maternal *r1* gene expression. In theory, full maternal expression of the *r1* gene and other MEGs could depend on maternal *mdr1* expression either because of a requirement for demethylation initiating before paternal genome activation or because *mdr1* is itself a MEG. To test this, we looked at prior imprinting data in W22 and B73 (Anderson et al. 2021). This revealed that *mdr1* was biparentally expressed, with a reciprocal expression ratio of 0.73 and 0.68 calculated from over ~1,000 reads for each library. *dng102* was also biparentally expressed, but with a paternal preference, with ratios of 0.36 and 0.33 calculated from fewer than 100 reads in most libraries. These numbers indicate that neither *mdr1*

nor *dng102* are MEGs and raise the possibility that the *r1* gene and other MEGs are exceptional in their requirement for *mdr1* prior to paternal genome activation.

DISCUSSION

In Arabidopsis, DEMETER is required for normal development not just in endosperm, but in the pollen vegetative cell and non-reproductive cell types as well (Borg et al. 2021; Khouider et al. 2021; Kim et al. 2021). In the case of pollen vegetative cell, developmental phenotypes are clearly determined by gene regulation rather than transposon activation. Here we have demonstrated in maize that it is essential to have a functional copy of at least one of the DNA glycosylases *mdr1* or *dng102* present prior to fertilization in both the maternal and paternal gametophytes, based on inability to transmit two mutant alleles simultaneously. Since imprinted gene expression in maize is also clearly linked to demethylation of maternal alleles (Zhang et al. 2014), we expect that gametophyte defects are a consequence of mis-regulated gene expression rather than genome instability due to transposon activation. Such developmentally dynamic mCG and mCHG in and near genes would be exceptions to the general rule of such methylation being stable (Oka et al. 2017; Crisp et al. 2020). In multiple species, demethylation of transposons near MEGs is clearly connected to their maternal expression (Gehring et al. 2009; Hatorangan et al. 2016; Pignatta et al. 2018; Rodrigues et al. 2021). A prediction from these observations is that a genome like maize, with more frequent transposons near genes, would have larger numbers of imprinted genes. This, however, is not the case. In addition, we found that endosperm DMRs were only weakly associated with transposons, mainly unclassified retrotransposons. These transposons that were enriched for endosperm DMRs accounted for about a tenth of total

endosperm DMRs. It is clear both from our findings and prior work on other species that DNA glycosylases demethylate repetitive elements and that their demethylation can cause imprinted gene expression in endosperm [reviewed in (Anderson and Springer 2018; Batista and Köhler 2020)]. Our results are consistent with this, but further indicate that demethylation of transposons is not the major function of DNA glycosylases in maize endosperm. Counter-intuitively, transposon demethylation would make sense in the context of demethylation being part of a transposon silencing mechanism, where demethylation promotes siRNA expression that transfer into egg or embryo. While our siRNA expression data does not rule out this hypothesis, it does indicate that 21 and 22-nt siRNAs are not a typical feature of endosperm DMRs, and 24-nt siRNAs are only typical of a subset of DMRs near euchromatin. In addition, recent work in *Brassica napa* also found no evidence to support the idea of endosperm siRNAs being transferred to embryo (Chakraborty et al. 2021). These observations suggest transposon defense, and transposon-mediated gene regulation, are not the main function of DNA glycosylases in maize endosperm.

Consistent with what is known about glycosylase activity in other plants, MEGs were strongly enriched for overlapping endosperm DMRs, especially in the 5' end of genes. However, the majority of genes that were expressed in endosperm and contained endosperm DMRs were not imprinted (Figure 5A). This indicates that either the majority of endosperm DMRs do not have a major effect on gene expression or that the effect is not limited to one parent. This latter possibility, in which glycosylases act post-fertilization on both parental genomes is consistent with the biparental expression of both *mdr1* and *dng102* in endosperm long after the two parental genomes have merged. Testing this will require analysis of hybrid methylomes, where SNPs allow distinguishing one parental allele from another. If glycosylases act on both genomes in

endosperm, both maternal and paternal alleles would have reduced methylation relative to non-endosperm control tissue. Future experiments with higher read coverage and more complete identification of DMRs may also reveal whether DMRs in specific genomic contexts, i.e. overlapping transcription start sites, may have stronger effects on imprinting. It is also possible that endosperm demethylation has an additional function independent of gene regulation. Endosperm is unusual in its rapid rate of cell division, which partly takes place in a syncytium [reviewed in (Becraft and Gutierrez-Marcos 2012)]. One could imagine, for example, that removal of DNA methylation from select regions of the genome has a function in facilitating rapid DNA replication. While *mdr1* null mutants effectively behave as partial-loss-of-function due to complementation by its homolog *dng102*, the viability of *mdr1* endosperm and pollen will continue to provide an experimental resource for answering these questions.

METHODS

Candidate *mdr1* mutation mapping

Homozygous *mdr1* plants derived from stock X336J (from the Maize Genetics Cooperation Stock Center, http://maizecoop.cropsci.uiuc.edu/) were crossed with the W22 sequenced stock (Springer et al. 2018) to create *mdr1/Mdr1* heterozygotes. These were crossed as females with *mdr1* homozygous males derived from stock X336J to produce progeny segregating homozygous and heterozygous *mdr1*. All plants were homozygous for the *R1-r:standard* haplotype which is capable of mottling . Approximately 500 putative *mdr1* homozygotes were selected based on kernel mottling and planted in five batches of 100 each.

Single leaf tips from each seedling were combined for each of the five batches, and a single DNA extraction performed on each batch, followed by Illumina library preparation with a KAPA HyperPrep Kit (KK8502). Three batches were sequenced single-end and two sequenced paired-end, all with 150-nt reads. A single batch of five W22 leaf tips was also processed in the same manner as a homozygous *Mdr1* control and sequenced single-end. Adapter sequences were trimmed from all reads using cutadapt 1.9.1 (Martin 2011) with adapter sequence AGATCGGAAGAGC for both forward and reverse reads and parameters -q 20 -O 1 -m 100. Trimmed reads were aligned to the W22 reference genome (Springer et al. 2018) using BWA mem version 0.7.15 (Li and Durbin 2009) with the -M parameter.

To generate a set of high-confidence variants, duplicates were marked using Picard MarkDuplicates (v2.16.0 http://broadinstitute.github.io/picard). Downstream variant calling was performed using software in the GATK package (v3.8.1) and the GATK best practices (Van der Auwera et al. 2013). In particular, indels were realigned using RealignerTargetCreator and IndelRealigner, and an initial round of SNP calling was performed using HaplotypeCaller and GenotypeGVCFs. The resulting vcf files for the W22 bulk and the combined *mdr1* bulks were filtered to remove variants that were present in the W22 bulk using the BEDTools subtract tool version 2.26.0 (Quinlan and Hall 2010) and stringently filtered to remove low quality score variants using VCFtools 0.1.15 (Danecek et al. 2011) with the -minQ 200 parameter. The resulting vcf was used as a training set for base recalibration of the W22 bulk and *mdr1* combined bulks using BaseRecalibrator, and HaplotypeCaller/GenotypeGVCFs was used to call variants in W22 and *mdr1*. VariantsToTable was used to convert vcf data into a readable table for analysis in RStudio.

The mapping scheme predicted that variants across the genome would have a frequency of 0.75 non-W22 variants (alternative alleles). Since we were looking specifically for a region was near homozygous for alternative alleles in the *mdr1* bulks, the following additional criteria were imposed to identify putative variants linked to *mdr1* in the vcf table: 1) An alternative allele frequency in the W22 bulk of zero, with a read depth of at least two but no more than 15. 2) A read depth in the *mdr1* combined bulks of at least five but no more than 50. 3) A normalized Phred-scaled likelihood for homozygous reference alleles of at least 400 and for homozygous alternative alleles of zero in the *mdr1* combined bulks. Filtering the vcf table in this way selected for near-homozygous alternative alleles across the genome in the *mdr1* bulks, with an average of 96% alternative allele frequency. A single ~20 MB region (chr4:216965546-237367197) stood out, with a ~98.5% alternative allele frequency. A second method of analyzing the GATK variants was also used to identify the *mdr1* candidate region. The vcf table was filtered to exclude all variants but those with a GenotypeQuality of at least 20 in both genotypes, depth of coverage between 5-14 in the W22 bulk, a depth between 6-49 in the combined *mdr1* bulks, and variant frequencies above 0.75 in the *mdr1* bulks. Based on the mapping scheme, this is expected to remove the majority of technical artifacts. As with the first method, this singled out the same region on chromosome 4 as the most homozygous in the genome, with variant frequencies near 1, as shown in Supplemental Figure 1. A comparison of variant density by genomic position revealed that a large region of chromosome 5 as well as a few small regions elsewhere in the genome of the *mdr1* mutant stock X336J were not shared with the sequenced W22 genome (Supplemental Figure S2). The tip of chromosome 4, however, had few variants and was clearly shared between both genomes.

Mutant genotyping

An 867-bp region of *mdr1* was amplified by PCR with primers W22-Ros1-F1 and W22-Ros1-R1 and digested with TaqI-v2 (NEB #R0149S). The different *mdr1* alleles could be distinguished by the pattern of restriction fragments due to the presence of SNPs that created or removed TaqI restriction sites in different genetic backgrounds (Kermicle's *mdr1* allele or its progenitor in W22, the *EMS4-06835d* allele or its progenitor in B73, and the allele linked to the Ds-GFP insertion *R128F11* in A188). A single TaqI restriction site in B73 yields two fragments of 382 and 485 bp. A second site in A188 yields three fragments, 107, 275, and 485 bp. Absence of any sites in W22 yields an intact amplicon of 867 bp. In cases where the wild-type and mutant were in the same genetic background or when there was a possibility of recombination between *R128F11* and *mdr1*, the *mdr1* mutation was genotyped by Sanger sequencing of the amplicon. Kermicle's *mdr1* allele reduced the string of seven A's to six in the W22 amplicon subsequence AGTTTACGAAAAAAAGTGCTTTC. Primers are as follows:

>W22-Ros1-F1

TTAGAACAAGGCGACTTTTCACC

>W22-Ros1-R1

AGGAACTCGGTAGCTCTTCT

To genotype the *dng102-Q235 allele*, A 569-bp region of *dng102* was amplified by PCR with JIG-339 and JIG-340 primers and digested with Hpy188I (NEB # R0617S). This amplicon has two Hpy188I sites in wild-type, but the mutation that converts Q to a stop codon at amino acid position 235 removes one of them such that digestion of the wild-type amplicon yields three

fragments of 317, 156, and 96 bp, but the *dng102-Q235* allele yields two fragments of 473 and 96 bp. Primers are as follows:

>JIG-339

TCCAAAGCCTCCCAAAGAAA

>JIG-340

GCATATGATCGACCGAGCTG

To genotype the *dng102 mu1083641* allele, short regions were amplified using EoMumix3 primers along with either of two *dng102* primers, JIG-337 or JIG-338, which were also used together to amplify the wild-type locus. Primers are as follows:

>EoMumix3

GCCTCYATTTCGTCGAATCCS

>JIG-337

CTGGCAGGAAAATGCAACCT

>JIG-338

GCAATTGCGCAAGCTTAATG

To genotype the Ds-GFP insertion *R153F01* linked to *Dng102*, two primer combinations were used: the Ds-GFP primer JSR01dsR with the flanking primer JIG-368, and the Ds-GFP primer JGp3dsL with the flanking primer JIG-369. JIG-368 and JIG-369 were also used as a pair to amplify the wild-type locus. Primers are as follows:

>JSR01dsR

GGGTTCGAAATCGATCGGGATA

>JGp3dsL

ACCCGACCGGATCGTATCGG

>JIG-368

CGGAGTCATTCTTGCTCTTCT

>JIG-369

AGGACGAGACCTTAGTCCTTAT

Amino acid sequence comparison of maize DNA glycosylases and DEMETER.

To analyze the relationship between DEMETER (AT5G04560.2) and its maize homologs, the longest putative isoform of each protein was selected. Pairwise alignments were performed between DEMETER and the five maize homologs and between MDR1 and the remaining four maize homologs. Alignments were performed using the MUSCLE algorithm (Edgar 2004) in Geneious Prime v2020.2.4 (http://www.geneious.com). The percentage of identical sites (%Id) and alignment length are reported. The five maize glycosylase amino acid sequences are from the B73 v5 reference annotations: Zm00001eb202980_P001, Zm00001eb241240_P002, Zm00001eb289030_P003, Zm00001eb241310_P001, and Zm00001eb279940_P003 (Hufford et al. 2021).

Ds-GFP screening for mutation transmission

Stocks containing Ds-GFP alleles (tdsgR128F11) and (tdsgR153F01) were obtained from the Maize Genetics Cooperation Stock Center. Double heterozygous plants (*R128F11/mdr1 R153F01/dng102-Q235*) were reciprocally crossed with wild-type Oh43 inbred plants that did

not carry Ds-GFP. Presence of Ds-GFP insertions in kernels was determined using a blue light and orange filter glasses similar to previous methods, but on a smaller scale without automated kernel counting (Warman et al. 2020). The *dng102-Q235* allele was derived from the TILLING stock 04IA-B73PS-055_C4 in B73, which was then backcrossed into non-mutagenized B73, selfed twice, then crossed three times into either W22 or W22-related stocks before crossing with Ds-GFP stocks.

Preparation of EM-seq libraries

EM-seq libraries (NEBNext® Enzymatic Methyl-seq Kit #E7120S) were prepared from three tissues. 1) Early developing 2nd leaves were collected before the unfurling of the first leaves. Three replicates each of the W22-related *mdr1* homozygous stock derived from X336J and the W22-related wild-type stock X336K were included. Each replicate included leaves from ~5 to 9 plants. 2) Premeiotic tassels were collected from plants at approximately the 15-leaf stage, when tassels were between 0.8 and 2.3 cm in length. Three replicates each of *mdr1* homozygous and *Mdr1* homozygous siblings were included. The heterozygous parents that were crossed to produce these plants were the product of a cross between the sequenced W22 inbred stock and the W22-related *mdr1* homozygous stock derived from X336J. Each replicate included two or three individual tassels. 3) Mature endosperm was separated from pericarp with forceps after soaking in water for 20 minutes and separated from embryo by dissection with a razor blade. Three replicates each of *mdr1* homozygous and *Mdr1* homozygous siblings were included. The dissected kernels were siblings of the ones that were planted and grown for tassel collection. Each replicate consisted of endosperm from a single kernel. DNA was extracted using an IBI plant DNA extraction kit (#IB47231) after grinding to a fine powder (in liquid nitrogen for tassel

and leaf). The input for each EM-seq library consisted of 200 ng of genomic DNA that had been combined with 1 pg of control pUC19 DNA and 20 pg of control lambda DNA and sonicated to fragments averaging ~700 bp in length using a Diagenode Bioruptor. The protocol for large insert libraries was followed. Libraries were amplified with 4 or 5 PCR cycles. Libraries were Illumina sequenced using paired-end 150 nt reads. Read counts and accession numbers are listed in Supplemental Table 2.

Preparation of small RNA sequencing libraries

The same tassel tissues used for preparation of EM-seq libraries were also used for preparation of small RNA sequencing libraries. RNA was extracted using a mirVana miRNA isolation kit (Thermo Fisher Scientific #AM1560) using the small RNA enrichment method. Plant RNA Isolation Aid (Thermo Fisher Scientific #AM9690) was added at the lysis step. Small RNA sequencing libraries were prepared using the NEXTflex Small RNA-Seq Kit v3 (PerkinElmer #5132-05) using the gel-free size selection & cleanup method. ~500 ng small RNA enriched RNA was used as input for each library, and libraries were amplified with 13 cycles of PCR. Read counts and accession numbers are listed in Supplemental Table 3.

Gene, transposon, and tandem repeat annotations

Gene annotations (Zm-W22-REFERENCE-NRGENE-2.0_Zm00004b.1.gff3.gz) and transposon annotations (W22.allTE.withSolo_10Jan18.gff3.gz) were downloaded from https://ftp.maizegdb.org/MaizeGDB/FTP/Zm-W22-REFERENCE-NRGENE-2.0/. The transposon annotations are based on a conservative identification method using intact known structural features (Springer et al. 2018; Anderson et al. 2019a). Retrotransposons having solo

LTRs were distinguished from those having both LTRs in these annotations by “Zm00004bS” in their names. Superfamilies were identified by their three-letter codes, e.g., RLG for Gypsy. A subset of genic analysis (Figure 1B and C, Figure 5E and F, and Supplemental Figure S5) made use of a more stringently filtered set of gene annotations where genes were annotated in both the W22 reference and B73 refgen_v4 (Jiao et al. 2017). These constituted 24,133 of the total 40,650 W22 genes. MEG and PEG status of genes in W22 were determined as previously described (Anderson et al. 2021). To define the endosperm gene expression categories used in Figure 5C, a B73 developmental expression atlas data (Stelpflug et al. 2016) was analyzed using similar methods as previously (Anderson et al. 2019b). Genes that are constitutively lowly or not expressed (not.expressed) genes have an RPM > 1 in fewer than 3 of the 225 total libraries representing different tissues, developmental stages, and biological replicates. Endosperm expressed (endo.expressed) and not endosperm expressed (not.endo) were identified from the rest of the genes based on the following criteria: The "endo.expressed" genes have an RPM > 1 in at least 18 of 21 endosperm libraries. The "not.endo" genes have an RPM value of < 0.5 in at least 20 of the 21 endosperm libraries. Finally, the W22 genes corresponding to the B73 genes meeting these criteria were determined with a gene key, which can be found at https://github.com/SNAnderson/Expression/blob/main/genes.expression.development.categories.20Sept21.txt. Locations of tandem repeats in the genome were determined by mapping dimerized consensus sequences for *CentC*, *knob180*, and *TR-1* (Gent et al. 2017; Liu et al. 2020) to the genome using Bowtie2 (Langmead and Salzberg 2012) (version 2.4.1) with the --local parameter. All repeat loci that were within 50 bp of each other were merged with the BEDTools merge tool with the -d 50 parameter. Following this, all loci less than 500 bp in length were removed.

Methylome read processing, definition of regions by methylation status

EM-seq reads were trimmed of adapter sequence using cutadapt, parameters -q 20 -a AGATCGGAAGAGC -A AGATCGGAAGAGC -O. Reads were aligned to each genome and methylation values called using BS-Seeker2 (version 2.1.5), parameters -m 1 --aligner=bowtie2 -X 1000 (Guo et al. 2013). To identify DMRs, CGmaps from each group of three individual replicates were merged into single CGmaps using the merge2 tool of CGmapTools version 0.1.2 (Guo et al. 2018). Intersected CGmaps for each pair of merged CGmaps were produced using the CGmapTools intersect tool in each sequence context separately (CG, CHG, and CHH). Single-bp level methylation measurements in Figure 5C were produced using the CGmapTools lollipop tool on the endosperm-tassel intersected CGmaps. CG and CHG DMRs and control eligible regions were then identified from the separate CG and CHG intersected maps using DMR_Finder.py (see Supplemental scripts), parameters as follows: binSize = 200, minCov = 3 minCountC = 5, minAbsDif = .20, minRelDif = .5. The set of CG DMRs that overlapped CHG DMRs were identified using the BEDTools (version 2.29.2) intersect tool. To finalize the set of DMRs, all DMRs located on scaffolds rather than chromosomes 1 through 10 were discarded. RdDM regions were identified from the merged tassel CGmap using MR_Finder.py (see Supplemental scripts), parameters as follows: binSize = 200, minCov = 3, minCountC = 5, minMeth = .2, context1 = 'CHH', context2 = 'CHH'. CHG unmethylated regions (UMRs) were identified from the merged tassel CGmap using UMR_Finder.py (see Supplemental scripts), parameters as follows: binSize = 200, minCov = 3, minCountC = 5, maxMeth = .2, context1 = 'CHG', context2 = 'CHG'.

Characterization of DMRs in terms of methylation levels, overlap with specific genetic elements, and sequence motifs

Methylation averages on 100-bp intervals centered on DMRs and control eligible regions were produced using the CGmapTools mfg tool, requiring a minimum coverage of two with the -c 2 parameter. Prior to the mfg analysis, regions were first divided into two groups based on distance to closest CHG UMR: greater than two Kb as heterochromatin, less than or equal to two Kb as euchromatin. These distances were measured using the BEDTools closest tool with -t first -d parameters. Regions were oriented relative to CHG UMRs such that the nearest CHG UMRs was always in the downstream direction using region_UMR_orienter.py (see Supplemental scripts). Orientation for regions that were equidistant from two CHG UMRs was randomized. DMR methylation averages used for violin plots were obtained using the CGmapTools mtr tool. Only DMRs with an effective coverage of at least 15 were included. Effective coverage is the number of cytosines covered by reads in the region times the number of reads that cover each cytosine. The numbers of overlaps between regions and different genetic elements, requiring at least 100 bp of overlap to count, were obtained using the BEDTools intersect tool with parameters -u -f .5. To determine whether DMRs contained motifs associated with differential methylation, fasta sequences were derived from bed files using the BEDTools getfasta tool. Three subsamples of 5000 sequences were input into STREME, each with a different seed number (Bailey 2021). All motifs identified by STREME were input into FIMO to identify matches throughout the complete sets of sequences. SpaMo (Whitington et al. 2011) was then used to search for combinations of motifs that may work together to cause the differential methylation, but was also inconclusive. AT richness was measured using the EMBOSS program infoseq (Rice et al. 2000).

Characterization of siRNA expression in methylated regions

Small RNA-seq reads were quality filtered, trimmed of adapters, and filtered for lengths of 20-25 nt using cutadapt (Martin, 2011), parameters -q 20 -a TGGAATTCTCGGGTGCCAAGG -e .05 -O 5 --discard-untrimmed -m 28 -M 33. The four random nucleotides at each end were then removed using cutadapt with the -u 4 parameter followed by cutadapt with the -u -4 parameter. Reads were aligned to the genome with BWA-backtrack (version 0.7.15), parameters aln -t 8 -l 10. A set of miRNA loci were identified using ShortStack software (Axtell 2013) (grep 'MIRNA=Y' from the ShortStack_D.gff3 output file). All wild-type small RNA bam files from embryo, endosperm, and tassel were combined and used as a single input file for ShortStack to identify miRNA loci. Also included in this combination were *zmet2* and *zmet5* heterozygous samples SRR6425186, SRR6425229, SRR6425204, and SRR6425207, which were not included in subsequent DMR analyses. To produce the final sets of siRNAs, the BEDTools intersect tool with parameters -v -f .9 was used to filter out reads that overlapped at least 90% of their length with ShortStack miRNA loci. Multi-mapping reads were included in all analyses. siRNAs that overlapped by at least 90% of their lengths with DMRs were identified using the BEDTools intersect tool with the -u -wa -f .9 parameters. The number of overlaps with DMRs was obtained using the BEDTools coverage tool.

Accession numbers

All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA759188.

Individual accession numbers for small RNAs are listed in Supplemental Table S2 and for DNA methylation in Supplemental Table S3.

ACKNOWLEDGMENTS

We thank the Maize Genetics Cooperation Stock Center for providing maize stocks; Akshay Nair, Ankush Sangra, and Mary Tindall Smith for genotyping stocks; and John Fowler for advice on use of Ds-GFP insertions for measuring frequency of mutation transmission. This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

FUNDING

NSF 1744001 and NSF 2114797

REFERENCES CITED

Anderson SN, Springer NM. 2018. Potential roles for transposable elements in creating imprinted expression. *Curr Opin Genet Dev* **49**: 8-14.

Anderson SN, Stitzer MC, Brohammer AB, Zhou P, Noshay JM, O'Connor CH, Hirsch CD, Ross-Ibarra J, Hirsch CN, Springer NM. 2019a. Transposable elements contribute to dynamic genome content in maize. *Plant J* **100**: 1052-1065.

Anderson SN, Stitzer MC, Zhou P, Ross-Ibarra J, Hirsch CD, Springer NM. 2019b. Dynamic Patterns of Transcript Abundance of Transposable Element Families in Maize. *G3 (Bethesda)* **9**: 3673-3682.

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**: e1009491.

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19**: 740-751.

Bailey TL. 2021. STREME: Accurate and versatile sequence motif discovery. *Bioinformatics* doi:10.1093/bioinformatics/btab203.

Batista RA, Köhler C. 2020. Genomic imprinting in plants-revisiting existing models. *Genes Dev* **34**: 24-36.

Becraft PW, Gutierrez-Marcos J. 2012. Endosperm development: dynamic processes and cellular innovations underlying sibling altruism. *Wiley Interdiscip Rev Dev Biol* **1**: 579-593.

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD et al. 2013. Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* **110**: E435-444.

Borg M, Papareddy RK, Dombey R, Axelsson E, Nodine MD, Twell D, Berger F. 2021. Epigenetic reprogramming rewires transcription during the alternation of generations in Arabidopsis. *Elife* **10**.

Bouyer D, Kramdi A, Kassam M, Heese M, Schnittger A, Roudier F, Colot V. 2017. DNA methylation dynamics during early plant life. *Genome Biol* **18**: 179.

Calarco JP, Borges F, Donoghue MT, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD et al. 2012. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151**: 194-205.

Chakraborty T, Kendall T, Grover JW, Mosher RA. 2021. Embryo CHH hypermethylation is mediated by RdDM and is autonomously directed in Brassica rapa. *Genome Biol* **22**: 140.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* **110**: 33-42.

Creasey KM, Zhai J, Borges F, Van Ex F, Regulski M, Meyers BC, Martienssen RA. 2014. miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature* **508**: 411-415.

Crisp PA, Marand AP, Noshay JM, Zhou P, Lu Z, Schmitz RJ, Springer NM. 2020. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc Natl Acad Sci U S A* **117**: 23991-24000.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.

Doll NM, Just J, Brunaud V, Caïus J, Grimault A, Depège-Fargeix N, Esteban E, Pasha A, Provart NJ, Ingram GC et al. 2020. Transcriptomics at Maize Embryo/Endosperm Interfaces Identifies a Transcriptionally Distinct Endosperm Subdomain Adjacent to the Embryo Scutellum. *Plant Cell* **32**: 833-852.

Dooner HK, Wang Q, Huang JT, Li Y, He L, Xiong W, Du C. 2019. Spontaneous mutations in maize pollen are frequent in some lines and arise mainly from retrotranspositions and deletions. *Proc Natl Acad Sci U S A* **116**: 10734-10743.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

Frost JM, Kim MY, Park GT, Hsieh PH, Nakamura M, Lin SJH, Yoo H, Choi J, Ikeda Y, Kinoshita T et al. 2018. FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc Natl Acad Sci U S A* **115**: E4720-E4729.

Fu FF, Dawe RK, Gent JI. 2018. Loss of RNA-Directed DNA Methylation in Maize Chromomethylase and DDM1-Type Nucleosome Remodeler Mutants. *Plant Cell* **30**: 1617-1627.

Garcia-Aguilar M, Michaud C, Leblanc O, Grimanelli D. 2010. Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. *Plant Cell* **22**: 3249-3267.

Gehring M, Bubb KL, Henikoff S. 2009. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**: 1447-1451.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495-506.

Gent JI, Madzima TF, Bader R, Kent MR, Zhang X, Stam M, McGinnis KM, Dawe RK. 2014. Accessible DNA and Relative Depletion of H3K9me2 at Maize Loci Undergoing RNA-Directed DNA Methylation. *Plant Cell* **26**: 4903-4917.

Gent JI, Wang N, Dawe RK. 2017. Stable centromere positioning in diverse sequence contexts of complex and satellite centromeres of maize and wild relatives. *Genome Biol* **18**: 121.

Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M. 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**: 774.

Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. 2018. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**: 381-387.

Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C. 2016. Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. *Plant Cell* **28**: 1815-1827.

He S, Vickers M, Zhang J, Feng X. 2019. Natural depletion of histone H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation. *Elife* **8**.

Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y et al. 2021. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655-662.

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T et al. 2012. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360-1364.

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS et al. 2017. Improved maize reference genome with single-molecule technologies. *Nature* **546**: 524-527.

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. 2012. DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. *Curr Biol* **22**: 1825-1830.

Kapazoglou A, Drosou V, Argiriou A, Tsaftaris AS. 2013. The study of a barley epigenetic regulator, HvDME, in seed development and under drought. *BMC Plant Biol* **13**: 172.

Kawakatsu T, Nery JR, Castanon R, Ecker JR. 2017. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biol* **18**: 171.

Kermicle J. 1995. Location, time of action, and dominance of an imprintor of *R*-mottled maize. In *Modification of gene expression and non-mendelian inheritance*, (ed. Oono, Takiawa). NIAR, Japan.

Kermicle JL. 1970. Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* **66**: 69-85.

Khouider S, Borges F, LeBlanc C, Ungru A, Schnittger A, Martienssen R, Colot V, Bouyer D. 2021. Male fertility in Arabidopsis requires active DNA demethylation of genes that control pollen tube function. *Nat Commun* **12**: 410.

Kim MY, Ono A, Scholten S, Kinoshita T, Zilberman D, Okamoto T, Fischer RL. 2019. DNA demethylation by ROS1a in rice vegetative cells promotes methylation in sperm. *Proc Natl Acad Sci U S A* doi:10.1073/pnas.1821435116.

Kim S, Park JS, Lee J, Lee KK, Park OS, Choi HS, Seo PJ, Cho HT, Frost JM, Fischer RL et al. 2021. The DME demethylase regulates sporophyte gene expression, cell proliferation, differentiation, and meristem resurrection. *Proc Natl Acad Sci U S A* **118**.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.

Li Y, Segal G, Wang Q, Dooner HK. 2013. Gene tagging with engineered Ds elements in maize. *Methods Mol Biol* **1057**: 83-99.

Liu J, Seetharam AS, Chougule K, Ou S, Swentowsky KW, Gent JI, Llaca V, Woodhouse MR, Manchanda N, Presting GG et al. 2020. Gapless assembly of maize chromosomes using long-read technologies. *Genome Biol* **21**: 121.

Lu X, Liu J, Ren W, Yang Q, Chai Z, Chen R, Wang L, Zhao J, Lang Z, Wang H et al. 2018. Gene-Indexed Mutations in Maize. *Mol Plant* **11**: 496-504.

Martin M. 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnetjournal* **17**: 3.

Martínez G, Panda K, Köhler C, Slotkin RK. 2016. Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nat Plants* **2**: 16030.

McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J et al. 2005. Steady-state transposon mutagenesis in inbred maize. *Plant J* **44**: 52-61.

McCue AD, Panda K, Nuthikattu S, Choudury SG, Thomas EN, Slotkin RK. 2015. ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. *EMBO J* **34**: 20-35.

Moreno-Romero J, Jiang H, Santos-González J, Köhler C. 2016. Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *EMBO J* **35**: 1298-1311.

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR. 2016. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**: 1280-1292.

Oka R, Zicola J, Weber B, Anderson SN, Hodgman C, Gent JI, Wesselink JJ, Springer NM, Hoefsloot HCJ, Turck F et al. 2017. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol* **18**: 137.

Parent JS, Cahn J, Herridge RP, Grimanelli D, Martienssen RA. 2021. Small RNAs guide histone methylation in. *Genes Dev* **35**: 841-846.

Park JS, Frost JM, Park K, Ohr H, Park GT, Kim S, Eom H, Lee I, Brooks JS, Fischer RL et al. 2017. Control of DEMETER DNA demethylase gene transcription in male and female gamete companion cells in. *Proc Natl Acad Sci U S A* **114**: 2078-2083.

Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y et al. 2016. DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proc Natl Acad Sci U S A* **113**: 15138-15143.

Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife* **3**: e03198.

Pignatta D, Novitzky K, Satyaki PRV, Gehring M. 2018. A variably imprinted epiallele impacts seed development. *PLoS Genet* **14**: e1007469.

Pillot M, Baroux C, Vazquez MA, Autran D, Leblanc O, Vielle-Calzada JP, Grossniklaus U, Grimanelli D. 2010. Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis. *Plant Cell* **22**: 307-320.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.

Rodrigues JA, Hsieh PH, Ruan D, Nishimura T, Sharma MK, Sharma R, Ye X, Nguyen ND, Nijjar S, Ronald PC et al. 2021. Divergence among rice cultivars reveals roles for transposition and epimutation in ongoing evolution of genomic imprinting. *Proc Natl Acad Sci U S A* **118**.

Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarz L, Mosiolek M, Park JS, Park GT et al. 2011. Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proc Natl Acad Sci U S A* **108**: 8042-8047.

Schumann U, Lee J, Kazan K, Ayliffe M, Wang MB. 2017. DNA-Demethylase Regulated Genes Show Methylation-Independent Spatiotemporal Expression Patterns. *Front Plant Sci* **8**: 1449.

Schumann U, Lee JM, Smith NA, Zhong C, Zhu JK, Dennis ES, Millar AA, Wang MB. 2019. DEMETER plays a role in DNA demethylation and disease response in somatic tissues of Arabidopsis. *Epigenetics* **14**: 1074-1087.

Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F, Barad O, Barbazuk WB, Bass HW, Baruch K, Ben-Zvi G et al. 2018. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat Genet* **50**: 1282-1288.

Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Buell CR, de Leon N, Kaeppler SM. 2016. An Expanded Maize Gene Expression Atlas based on RNA Sequencing and its Use to Explore Root Development. *Plant Genome* **9**.

Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo CA, Enns LC, Odden AR et al. 2004. Discovery of induced point mutations in maize genes by TILLING. *BMC Plant Biol* **4**: 12.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**: 11.10.11-11.10.33.

Warman C, Panda K, Vejlupkova Z, Hokin S, Unger-Wallace E, Cole RA, Chettoor AM, Jiang D, Vollbrecht E, Evans MMS et al. 2020. High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements. *PLoS Genet* **16**: e1008462.

Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM. 2013. 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* **110**: 19639-19644.

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh CT, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM. 2011. Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**: 4221-4233.

Whitington T, Frith MC, Johnson J, Bailey TL. 2011. Inferring transcription factor complexes from ChIP-seq data. *Nucleic Acids Res* **39**: e98.

Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL. 2003. Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev Cell* **5**: 891-901.

Yamamuro C, Miki D, Zheng Z, Ma J, Wang J, Yang Z, Dong J, Zhu JK. 2014. Overproduction of stomatal lineage cells in Arabidopsis mutants defective in active DNA demethylation. *Nat Commun* **5**: 4062.

Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M et al. 2014. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell* **158**: 98-109.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**: 193-205.

Zhang C, Hung YH, Rim HJ, Zhang D, Frost JM, Shin H, Jang H, Liu F, Xiao W, Iyer LM et al. 2019. The catalytic core of DEMETER guides active DNA demethylation in Arabidopsis. *Proc Natl Acad Sci U S A* **116**: 17563-17571.

Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, Li H, Yang W, Zhao H, Wang G et al. 2014. Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res* **24**: 167-176.

Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Guan H, Hu X, Jiao Y, Song W et al. 2011. Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* **108**: 20042-20047.