

RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome

Qing Li^a, Jonathan I. Gent^b, Greg Zynda^c, Jawon Song^c, Irina Makarevitch^d, Cory D. Hirsch^a, Candice N. Hirsch^e, R. Kelly Dawe^b, Thelma F. Madzima^{f,1}, Karen M. McGinnis^f, Damon Lisch^g, Robert J. Schmitz^h, Matthew W. Vaughn^c, and Nathan M. Springer^{a,2}

^aMicrobial and Plant Genomics Institute, Department of Plant Biology, University of Minnesota, Saint Paul, MN 55108; ^bDepartment of Plant Biology, University of Georgia, Athens, GA 30602; ^cTexas Advanced Computing Center, University of Texas at Austin, Austin, TX 78758; ^dDepartment of Biology, Hamline University, Saint Paul, MN 55104; ^eDepartment of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108; ^fDepartment of Biological Science, Florida State University, Tallahassee, FL 32306; ⁹Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907; and ^hDepartment of Genetics, University of Georgia, Athens, GA 30602

Edited by Steven E. Jacobsen, University of California, Los Angeles, CA, and approved October 16, 2015 (received for review July 24, 2015)

The maize genome is relatively large (~2.3 Gb) and has a complex organization of interspersed genes and transposable elements, which necessitates frequent boundaries between different types of chromatin. The examination of maize genes and conserved noncoding sequences revealed that many of these are flanked by regions of elevated asymmetric CHH (where H is A, C, or T) methylation (termed mCHH islands). These mCHH islands are quite short (~100 bp), are enriched near active genes, and often occur at the edge of the transposon that is located nearest to genes. The analysis of DNA methylation in other sequence contexts and several chromatin modifications revealed that mCHH islands mark the transition from heterochromatin-associated modifications to euchromatin-associated modifications. The presence of an mCHH island is fairly consistent in several distinct tissues that were surveyed but shows some variation among different haplotypes. The presence of insertion/ deletions in promoters often influences the presence and position of an mCHH island. The mCHH islands are dependent upon RNA-directed DNA methylation activities and are lost in mop1 and mop3 mutants, but the nearby genes rarely exhibit altered expression levels. Instead, loss of an mCHH island is often accompanied by additional loss of DNA methylation in CG and CHG contexts associated with heterochromatin in nearby transposons. This suggests that mCHH islands and RNA-directed DNA methylation near maize genes may act to preserve the silencing of transposons from activity of nearby genes.

mCHH island | RdDM | chromatin boundary | maize | DNA methylation

he cytosine bases in a genome can be modified to 5-methylcytosine by adding a methyl group at the 5' position. This process, called DNA methylation, is conserved from algae to animals and plants (1, 2). DNA methylation can be separated into different types based on the local sequence context. In plants DNA methylation is found at the symmetric CG or CHG (where H = A, C, or T) sites or at nonsymmetric CHH sites. CG and CHG methylation are maintained at high fidelity following DNA replication due to activity of maintenance methyltransferases such as MET1 or chromomethylase (CMT) 3 (3, 4), whereas CHH methylation (mCHH) requires targeting by either domains rearranged methylase 2 (DRM2) or CMT2 (3-6). The DRM2 targeting occurs via RNAdirected DNA methylation (RdDM) and requires the activity of polymerase IV (PolIV) and polymerase V (PolV) complexes (3, 4). There is evidence that recruitment of PolIV and PolV may require the presence of dimethylation of lysine 9 of histone H3 (H3K9me2) or DNA methylation at the targeted genomic regions (7, 8). The specific mechanisms that recruit CMT2 are not well characterized but may require specific histone modifications (5, 6).

Much of our knowledge of DNA methylation in plants is derived from studies of the model plant *Arabidopsis thaliana*, which has a relatively small genome and relatively few examples of genes

with nearby transposons (36.3%; ref. 9). The maize genome is much more complex, with the majority (85.5%) of genes positioned within 1 kb of transposons. In both species, transposons tend to have quite high levels of CG and CHG methylation whereas genes have much lower levels (10). mCHH is often thought to provide an important component for silencing transposons, yet the maize genome has relatively low levels of mCHH despite the high transposon context (11). This is partially attributed to the lack of a CMT2 ortholog in maize (5), which may explain the reduced levels of mCHH in the middle of larger transposons. Although mCHH is low in maize, there are still genomic regions with elevated mCHH (12). Genomic profiles of mCHH in maize revealed that this modification is often found near genes (termed mCHH islands) and is dependent upon RdDM activity (12-14). This elevation of mCHH in regions surrounding genes is much less prevalent in Arabidopsis (10). A recent study showed that high mCHH can also be induced near genes that are up-regulated in plants subjected to phosphate starvation (15).

In this study we further probed the basis and function of these mCHH islands. We found that mCHH islands are short regions

Significance

RNA-directed DNA methylation (RdDM) provides a system for targeting DNA methylation to asymmetric CHH (H = A, C, or T) sites. This RdDM activity is often considered a mechanism for transcriptional silencing of transposons. However, many of the RdDM targets in the maize genome are located near genes or regulatory elements. We find that the regions of elevated CHH methylation, termed mCHH islands, are the boundaries between highly methylated (CG, CHG), silenced chromatin and more active chromatin. Analysis of RdDM mutants suggests that the function of the boundary is to promote and reinforce silencing of the transposable elements located near genes rather than to protect the euchromatic state of the genes.

Author contributions: R.J.S. and N.M.S. designed research; Q.L. performed research; J.I.G., I.M., C.D.H., C.N.H., R.K.D., T.F.M., K.M.M., D.L., and M.W.V. contributed new reagents/ analytic tools; Q.L., J.I.G., G.Z., J.S., C.D.H., R.J.S., M.W.V., and N.M.S. analyzed data; and Q.L. and N.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Sequence data have been deposited in the National Center for Biotechnology Information Sequence Read Archive and all accession numbers are provided in Table S2.

¹Present address: Division of Biological Sciences, School of STEM, University of Washington Bothell, Bothell, WA 98011.

²To whom correspondence should be addressed. Email: springer@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10. 1073/pnas.1514680112/-/DCSupplemental.

of elevated mCHH that flank nearly half of the genes in maize and many conserved noncoding sequences (CNSs). These mCHH islands mark a transition for CG and CHG DNA methylation, several histone modifications, and chromatin accessibility. The mCHH islands are relatively stable across different tissues but show some variation among haplotypes that are often associated with sequence insertions/deletions (InDels). The loss of mCHH islands does not strongly affect gene expression, but instead leads to an additional loss of CG and CHG methylation in some transposons flanking maize genes.

Results and Discussion

mCHH Islands Mark the Boundary Between Different Types of Chromatin in the Maize Genome. A metaprofile of context-specific DNA methylation surrounding maize genes reveals a gradual decline of CG and CHG methylation from flanking regions toward the genes (Fig. 1A). In contrast, there is an elevated level of mCHH in the regions flanking genes, and these regions have previously been termed CHH islands by Gent et al. (12); in this paper these will be referred to as mCHH islands to specify the regional accumulation of 5-methylcytosine in the CHH sequence context. The metaprofile of mCHH around genes has a fairly broad peak that spans ~200-800 bp upstream of the transcription start site (TSS). A similar region is observed downstream of the transcription termination site (TTS). We were interested in understanding whether this elevated mCHH in the metaprofile is due to elevated mCHH for all genes or whether this phenomenon was driven by a subset of genes. In addition, we wanted to define the actual size and location of mCHH islands relative to the TSS and TTS.

Whole-genome bisulfite sequencing (WGBS) data from the third leaf of B73 seedlings (16) were used to determine the coverage and level of DNA methylation in each sequence context for nonoverlapping 100-bp tiles across the entire maize genome. There are 29,922 maize genes with coverage for at least 50% of the tiles in the 2-kb region immediately upstream of maize genes and 25,973 genes with at least 50% coverage for the 2-kb region downstream of maize genes. An mCHH island was defined by the presence of a 100-bp tile with at least 25% mCHH within 2 kb. Genome-wide, only 1.2% of all 100-bp tiles have at least 25% mCHH but many genes contain an mCHH island within 2 kb of the 5' (51.3% of genes) or 3' (41.8% of genes) ends. Precise quantification of the number of genes with mCHH islands is hampered by the fact that the mCHH islands may exist in tiles with low coverage; thus, these numbers are likely underestimates. Indeed, the visualization of mCHH levels in flanking regions of all genes (Fig. 1B) also provides evidence that some genes classified as not having a strong mCHH island still contain a region of moderate mCHH within the flanking region. For each gene, there is a relatively small region (100–200 bp) with elevated mCHH levels, and these mCHH islands are most common in the first ~600 bp of the flanking regions but can occur anywhere throughout the 2-kb flanking regions (Fig. 1B). These results suggest that mCHH islands observed in the metaprofile are actually the result of a subset of maize genes that have elevated mCHH in sharply defined regions flanking the gene.

A second set of metaprofiles, only using genes with mCHH islands, were made to evaluate the context-specific profiles of DNA methylation and chromatin state relative to the mCHH islands (Fig. 1 *C-F*). These plots are centered on the 100-bp tiles identified as mCHH islands rather than on the TSS or TTS. The typical mCHH island has elevated levels of mCHH relative to the flanking regions (Fig. 1 *C* and *D*, red lines). The mCHH islands also clearly mark the transition from high levels of CG and CHG methylation that flanks genes to reduced CG and CHG methylation at the beginning and end of genes. The change in CG and CHG methylation is much sharper in plots centered on mCHH islands (Fig. 1 *C* and *D*) compared with plots centered

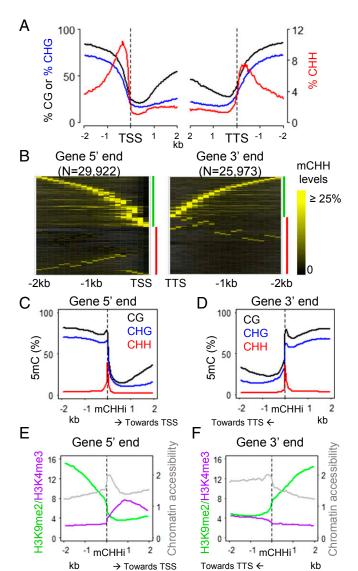


Fig. 1. Characterization of mCHH islands that are near genes. (A) Metaprofiles of context-specific DNA methylation for regions surrounding the gene TSS or gene TTS. CG (black) and CHG (blue) methylation are plotted using a different scale (Left) than mCHH (red) (Right) due to differences in abundance. (B) Genes with (green bar) or without (red bar) mCHH islands (100-bp tile with >25% mCHH anywhere within the 2-kb region) were used to create heat maps of mCHH in the 2-kb flanking regions at the 5' (Left) or 3' (Right) ends. (C and D) Context-specific DNA methylation profiles were generated for genes that contain an mCHH island. These plots are centered on mCHH islands (mCHHi) at the 5' end (C) or 3' end (D) rather than the TSS or TTS. (E and F) Similar profiles were generated to visualize the changes in chromatin at genes with mCHH islands. Different scales are used to visualize the different marks and the labels on the left or right of each plot indicate which scale is used for each mark using color coding of labels and lines. Read counts were used for the two histone marks and ratio of normalized read counts from libraries digested at 1 min and 16 min was used for chromatin accessibility (13).

on the TSS or TTS (Fig. 1A), suggesting that the mCHH island and not the TSS or TTS is the site of this change in methylation. Previous research suggested that mCHH islands themselves were not particularly enriched in the heterochromatin-associated H3K9me2 modification. Instead, these regions tended to have more accessible chromatin (13). This led us to evaluate whether mCHH islands might mark the transition zone between distinct types of chromatin by assessing the profile of chromatin on either side of

mCHH islands (Fig. 1 E and F). At the 5' end of genes containing mCHH islands, H3K9me2 exhibits a strong decrease at mCHH islands, whereas chromatin accessibility (13) is substantially increased for several hundred bases from mCHH islands toward the TSS. H3K4me3, a histone modification often associated with expressed genes, also shows a clear enrichment beginning in the region 3' of mCHH islands. The enrichment for accessible chromatin in the region flanking the mCHH island was observed both in highly expressed genes and silenced genes containing mCHH islands but is not as strong as the enrichment for these marks immediately upstream of the TSS (Fig. S1).

The mCHH islands flanking maize genes may provide boundaries between two distinct types of chromatin. We hypothesized that similar boundaries may also be required at regulatory regions to allow access to these regions for transcription factors. Turco et al. (17) identified a number of CNSs in the B73 genome and 11,680 of these are located >5 kb from the nearest gene. The profile of DNA methylation relative to these CNSs >5 kb away from genes reveals the presence of mCHH islands flanking CNSs and reduced DNA methylation at CNSs (Fig. S24). There are mCHH islands flanking 42.9% of the CNSs that are >5 kb from the nearest gene and the context-specific patterns of methylation and chromatin state at these mCHH islands flanking CNSs are very similar to those observed at mCHH islands flanking maize genes (Fig. S2 B and C).

mCHH Islands Are Present at Expressed Genes Located Near Terminal Inverted Repeat Elements. The mCHH islands found near genes and CNSs account for 49% of the regions with elevated (>25%) mCHH in the maize genome (Fig. 2A). Nearly half of the maize genes tested (with read coverage in at least half of the 2-kb region surrounding a gene) contain an mCHH island, and we were interested in understanding the factors that cause some genes to have mCHH islands whereas others do not. Previous research provides evidence that mCHH islands are enriched at more highly expressed genes and frequently occur at transposons

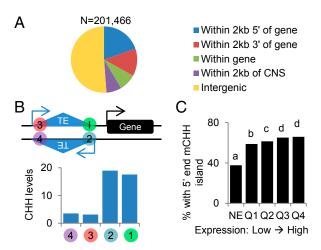


Fig. 2. mCHH islands are often located at transposon edges close to highly expressed genes. (A) All of the 100-bp tiles with >25% mCHH levels in B73 seedling were annotated based on their location relative to genes or CNSs. (B) The level of mCHH was determined for the distal and proximal edge (100 bp) of TEs greater than 1 kb in length. These TEs were also split according to whether they were located on the same or opposite strand as the gene. (C) The percentage of genes with a 5' end mCHH island is shown for genes that are not expressed (NE) or genes in each of the four expression quartiles (Q1-Q4) using RNA-seq data from the same tissue as the WGBS data, B73 seedling leaf. The small letter above each bar indicates whether there is a significant difference between the proportion of genes with mCHH island, with same letter indicating no difference and different letter indicating significant difference at P < 0.01 (prop.test in R) level.

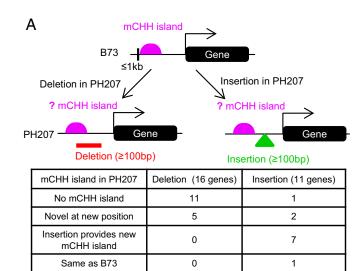
located near genes (12). Of the different classes of transposons, the mCHH islands identified in this study are enriched at terminal inverted repeat (TIR) DNA transposons located near genes (Fig. S3A, 38% vs. 6% for all tiles, P < 0.01). This enrichment is most apparent for transposons that are located closest to genes or CNSs and is only present at the edge of the transposon located closest to the gene and is not limited to DNA transposons (Fig. 2B and Fig. S3B).

The presence of a mCHH island is also associated with the expression level of the nearby gene. Genes that are expressed in the same tissue used for WGBS are much more likely to contain mCHH islands than silenced genes (P < 0.01, two-sample test for equality of proportions) and there is slight but steady increase in the frequency of genes with mCHH islands for more highly expressed genes (Fig. 2C and Fig. S3C). Genes with gene body CG methylation or genes that are syntenic between maize and sorghum are also more likely to have 5' mCHH islands (Fig. S3 D and E, P < 0.01). Although these analyses provide evidence that expression might be associated with mCHH islands, there are a number of expressed genes lacking mCHH islands (34%) and many silent genes have mCHH islands (38%). We further investigated these two subsets of genes to better understand the factors that might contribute to the presence of mCHH islands (Fig. 2C).

Highly expressed genes (the fourth expression quartile) in seedling tissue are enriched for having mCHH islands but some (34%) of these genes lack mCHH islands within 2 kb of the gene promoter. We hypothesized that these genes may lack the sequences or chromatin required for mCHH island formation. This could be due to a lack (or poor annotation) of TIR elements in the flanking regions or could be due to a lack of regions containing CG or CHG methylation located near the gene. The most highly expressed genes (fourth expression quartile) were divided into those with and without mCHH islands and then assessed for presence of transposons or highly methylated (>60% at CG/ CHG) tiles in the 2-kb flanking region. The genes without mCHH islands are less likely to contain TIR elements or tiles with high CG/ CHG methylation compared with genes with mCHH islands (Fig. S3 F and G, P < 0.01). However, many of the genes without mCHH islands do contain either TIR elements or elevated CG/CHG methylation, and it is not clear why these genes lack mCHH islands. Overall, these observations suggest that gene expression and presence of CG/CHG methylated DNA transposons near genes are important factors associated with the presence of an mCHH island, but these factors do not entirely explain the phenomenon.

Although genes that are not expressed in seedling leaf tissue are less likely to have mCHH islands, 38% of them do (Fig. 2C). The genes that are not expressed in seedling leaf tissue were further divided into two groups (expressed in other tissues or never expressed) based on their expression in 51 tissues or developmental stages of B73 (18). Genes that are expressed in other tissues are more likely to have mCHH islands than genes that are not detected in any of the tissues surveyed (Fig. S3H, P < 0.01). This suggests that mCHH islands are often present at expressed genes even in tissues without expression of the gene. This is further supported by the analysis of mCHH islands in four maize tissues in which the mCHH islands are fairly stable (Fig. S4A). Very few (<5%) mCHH islands have major differences among tissues (defined as having >25% mCHH in one tissue and <10% mCHH in another) and there was no evidence that the genes located near these rare mCHH islands that vary among tissues exhibit tissue-specific expression patterns that were related to the elevated mCHH levels (Fig. S4B).

Differences in mCHH Islands Among Genotypes Are Often Related to InDels. The stability of mCHH islands among diverse maize genotypes was assessed using WGBS from five maize inbreds (19). This analysis used 873 genes that have a 5' mCHH island in at least one genotype and have WGBS coverage for all tiles



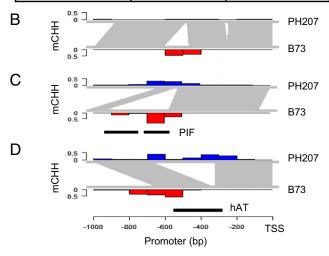


Fig. 3. Promoter InDels influence mCHH islands. (A) Genes with coverage for the first 1 kb of the promoter in both B73 and PH207 and an mCHH island in B73 were used to search for InDels > 100 bp located between the mCHH island and the TSS. Sixteen examples of PH207 deletions and 11 examples of PH207 insertions were identified. The effect on the PH207 mCHH islands was classified for each InDel and is shown in the table. Examples of the most common classes are shown in B-D. (B) At gene GRMZM5G871592_T01 the region containing the mCHH island (-400 to -500) in B73 is deleted in PH207 and no mCHH island is observed in PH207. (C) At gene GRMZM2G136178_T01 PH207 has a deletion that removes much of the B73 mCHH island but an mCHH island forms at a new location in PH207. (D) At gene AC194914.3_FGT002 PH207 has an insertion and an mCHH island forms at this insertion whereas the site of the B73 mCHH island has little mCHH in PH207. In C and D, the thick black lines represent B73 sequences annotated as DNA transposon and the subfamily (PIF or hAT) is indicated for each.

within 1 kb of the TSS in all five genotypes. In many cases the five genotypes have conserved mCHH islands (Fig. S4C). However, ~36% (312/873) of these mCHH islands have low (<10%) mCHH in at least one genotype. Many of the genes with genotype-specific mCHH islands were not differentially expressed (Fig. S4B). However, when the genes are differentially expressed, we observed that the haplotype with higher mCHH is more likely (66–73%) to exhibit elevated expression levels (Fig. S4D).

To further study the genetic factors leading to mCHH island variation we performed WGBS for PH207, a genotype with a de novo whole-genome assembly. This provided the opportunity to compare the full promoters in both genotypes. The analysis used a subset of 1,760 genes with a one-to-one match in B73 and

PH207 that also have WGBS coverage for the first 1 kb of the promoter in both genotypes. We identified 27 genes in this set that have an mCHH island in B73 and have an InDel >100 bp in PH207 that occurs between the location of the mCHH island and the TSS in B73. These include 16 examples with a deletion in the PH207 promoter that removes part or all of the sequence that forms the mCHH island in B73 (Fig. 3). Many (11/16) of the genes with a PH207 deletion lose the mCHH island from the 1-kb proximal promoter in PH207 (example in Fig. 3B). The other five loci containing a PH207 deletion form an mCHH island at a new site in PH207 (example in Fig. 3C). There are 11 loci with an insertion in the PH207 promoter located between the position containing the mCHH island and the TSS in B73 (Fig. 3). For 7 of these 11 insertions in PH207 a new mCHH island is formed at the insertion sequence itself (example in Fig. 3D). The majority of these InDels, including 14/16 PH207 deletions and 7/11 PH207 insertions, are annotated as transposons or have sequence similarity to transposons. These analyses suggest that the presence and location of mCHH islands can be influenced by transposon-derived InDels located in the promoter.

Loss of mCHH Islands Results in Additional Loss of Transposon CG and **CHG Methylation**. mCHH islands tend to form at TIR elements near expressed maize genes and mark a clear transition zone relative to CG/CHG methylation and several chromatin features that were assessed. The establishment/maintenance of mCHH islands requires mediator of paramutation 1 (MOP1) (GRMZM2G042443), a maize ortholog of the RDR2 gene in Arabidopsis (20, 21), MOP2 (GRMZM2G054225), and MOP3 (GRMZM2G007681), suggesting that mCHH islands are formed by RdDM activity (13). Sequence-capture bisulfite sequencing (22) data were used to assess the effects of three RdDM mutants mop1 (21), mop2 (23), and mop3 (24) on mCHH within and surrounding mCHH islands at 347 loci that were located in gene promoters and that were also included in a capture design that target a specific set of maize regions (Fig. 4A; ref. 25). All three mutations greatly reduce mCHH levels at mCHH islands. These materials provide a resource to further probe the function of mCHH islands. One hypothesis is that mCHH islands are important for gene expression levels and provide a border preventing the spread of heterochromatin toward the gene. This hypothesis would predict that CG and CHG methylation might be increased in regions 3' of the mCHH island and that genes containing mCHH islands would be more likely to be differentially expressed (at lower levels) in mop1 or mop3 relative to wild type. We do not see any evidence for increased CG or CHG methylation in regions 3' of the mCHH island (Fig. 4 B and C). Transcriptome profiling in mop1 and mop3 found that genes with mCHH islands are slightly enriched for differential expression but the majority of these differentially expressed genes were upregulated in the mutant (Table S1), which is not consistent with a function for mCHH islands in preventing silencing of the genes.

An alternative hypothesis is that mCHH islands act as a border to prevent the spread of euchromatin into transposons near active genes. This hypothesis would predict that the loss of mCHH islands might be accompanied by additional loss of CG and CHG methylation within the transposon and increased expression for some of the transposons near active genes. The sequence capture bisulfite sequencing experiments profiling mop1, mop2, and mop3 mutants reveal that mCHH methylation is greatly reduced at mCHH islands (Fig. 4A). CG and CHG methylation is also reduced at the region defined as the mCHH island with the strongest reduction in mop3 and minimal loss in mop2 (Fig. 4 B and C). The profiles also revealed a reduction in CHG (and to a lesser extent CG) methylation in the region 5' of the mCHH island. WGBS of mop1 supports these findings (Fig. S5A-C) with evidence for loss of CG and CHG methylation in the regions near the mCHH islands near both ends of genes and CNSs. The regions outside of the mCHH islands that exhibit loss of CG and CHG methylation have very low levels

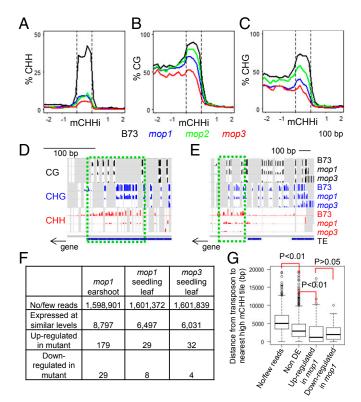


Fig. 4. Loss of mCHH islands in RdDM mutants and its potential consequences. (A-C) DNA methylation profiles in B73 and three RdDM mutants based on data from sequence capture bisulfite sequencing that includes 347 mCHH islands. The mCHH island is indicated between the two vertical dashed lines. The size of each mCHH island is normalized to 100 bp. (D and E) The context-specific levels of DNA methylation at the mCHH island and surrounding regions are shown for wild-type and two mutants using the Integrative Genomics Viewer (IGV) (26). In D the mutants show loss of methylation in all contexts within the mCHH island but CG and CHG methylation outside the mCHH island is largely maintained. In E there is evidence for loss of CHG methylation for several hundred base pairs outside of the mCHH island. The gray areas represent regions that do not have data coverage due to lack of reads or methylation sites (either CG or CHG or CHH). The mCHH island is indicated by the dashed green boxes. (F) Expression of transposons in mop1 and mop3 mutants relative to wild type. Uniquely mapping reads were used to classify each TE as having few or no reads, expression at similar levels or differential expression with up- or downregulation in the mutants. (G) A boxplot is used to show the distribution of distances for each type of transposon from the nearest high mCHH tile for the mop1 earshoot data. P value is based on Wilcoxon test.

of CHH methylation and do not seem to be active targets of RdDM in wild-type plants. We examined siRNA distributions around mCHH islands to test whether RdDM activity might actually cover a larger area than revealed just by mCHH. Consistent with mCHH, however, siRNAs were highly enriched specifically within mCHH islands and dramatically depleted in *mop1* (Fig. S5 D and E). It is possible that small RNAs and RdDM activity for these regions 5' of the mCHH islands is present at an earlier developmental stage and the loss of CG/CHG methylation in these regions reflects loss of RdDM activity at an earlier stage.

Visual examination of several loci confirms these trends and suggests variability in the locus-specific patterns in the mutants (Fig. 4 *D* and *E*). Several subtypes of mCHH islands were identified to better characterize the effects of the *mop* mutations on CG and CHG loss 5' of the mCHH islands. This analysis is restricted to 147 regions that have sequence capture probes and that have data in both the mCHH island itself and in the 400 bp 5' of the island. The analysis of CG and CHG methylation levels

at this region revealed that 48 of the loci only have CG/CHG methylation at the mCHH island itself (Fig. S6.4). The remaining 99 loci contain elevated CG and CHG methylation upstream of the mCHH islands. Clustering of the difference in CG and CHG methylation in *mop3* relative to wild type reveals that about one-third of these loci exhibit loss of CG and CHG methylation only at the mCHH island itself (Fig. S6 B and C). Another one-third of the loci exhibit strong loss of CG and CHG methylation for several hundred base pairs 5' of the mCHH island, whereas the final set shows partial loss of CG and CHG methylation 5' of the mCHH island (Fig. S6 B and C). Similar trends were observed for these loci in the *mop1* mutant (Fig. S6C) although the severity of the CG and CHG loss was reduced.

Earlier studies of the Mutator transposon in maize found evidence for progressive loss of DNA methylation and activation for these elements in the mop1 mutant (27) and analysis of shoot apical meristem found evidence for large-scale transposon and gene expression changes in mop1 (28). Therefore, we investigated whether loss of mCHH islands in mop1 or mop3 mutants is associated with activation of nearby transposons. There are several complications with attempts to study transposon expression. The repetitive nature of many transposons severely limits the ability to study specific insertions. In addition, the variation in specific transposon insertions in different haplotypes complicates analysis of transposon expression based on the reference genome. Uniquely mapping RNA sequencing (RNA-seq) reads were used to search for specific transposons that exhibit altered expression in mop1 (back-crossed into B73) and mop3 (in a non-B73 genetic background) mutants relative to wild-type siblings. The vast majority of transposons are not expressed (or not detected by uniquely mapping reads) or have low/similar levels of expression in mutant and wild-type genotypes in either leaves or immature ears (Fig. 4F). Depending on the mutant and tissue analyzed there are 36–208 individual transposons with altered expression (Fig. 4F). The majority (78–89%) of these exhibit increased expression in the mutant relative to wild type (Fig. 4F). There were fewer transposons with altered expression in mop3 mutants than in mop1 mutants. This is likely due to the fact that mop3 mutants are in a distinct genetic background and analysis of transposons in this background will be limited to those that are common in both genotypes. It is likely that there are additional transposons with altered regulation in mop3 that cannot be detected by alignments to the B73 reference. In some cases the up-regulated transposons are located near regions that were targeted for the sequence capture bisulfite sequencing and we could observe the coincident loss of the mCHH island and elevated expression of the transposon (TE) (Fig. S7 A and B). Genome-wide, the transposons that are up-regulated in mop1 or mop3 are significantly (P < 0.01, Wilcoxon test) closer to RdDM sites (>25% mCHH) than transposons that are silent (or equivalently expressed) in mutant and wild type, and a greater proportion of them are within 100 bp of RdDM sites (Fig. 4G and Fig. S7 C and D). This suggests that RdDM sites protect a subset of transposons from activation and these may be transposons that are located near active genes. Given the progressive loss of methylation at some *Mutator* elements over multiple generations in *mop1* (27) we might expect that the erosion of CG/CHG DNA methylation and activation of transposons near genes might affect even more loci in subsequent generations.

Conclusion

Plant genomes are often composites of genes and transposons with substantial variation in the abundance of transposons across different species. Although most *Arabidopsis* transposons have elevated mCHH (5), the CHH methylation in the maize genome is only found at some transposon regions. Recent studies have suggested that the recruitment of PolIV and RdDM activities requires DNA methylation and/or elevated H3K9me2 (7, 8). In addition, RdDM seems to target intergenic regions and plant genes located primarily in euchromatin (29). These requirements

suggest that PolIV and RdDM will primarily act at the borders between open chromatin and regions with elevated CG/CHG methylation. These borders typically occur at both edges of a transposon in the Arabidopsis genome (5). However, in the maize genome most transposons are present in large blocks of other transposons and therefore only the transposons that are at the edges near expressed genes might recruit RdDM activities. Indeed, much of the mCHH in the maize genome occurs near genes (13). This suggests that RdDM may not be a crucial requirement for silencing all transposons (5, 6, 30). Instead, this activity may be critical for maintaining the silencing of the transposons that are located near genes (31). Indeed, the analysis of changes in DNA methylation in the mop1 and mop3 mutants that perturb components of the maize RdDM system revealed that the loss of mCHH at the transposon edges near genes can often result in additional loss of CG/CHG methylation in the transposon. Only a subset of transposons exhibit transcriptional activation in mop1 and mop3 and these are often located near mCHH islands. This suggests that the mCHH islands and near-gene RdDM activity may be critical for creating a boundary that prevents the spread of open, active chromatin into adjacent transposons.

Materials and Methods

Sequencing Datasets. A full description of the biological samples, extraction of nucleic acids, library construction, and sequencing is available in *SI Materials and Methods*, and Table S2 lists all samples and accession numbers of raw

- 1. Feng S, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci USA* 107(19):8689–8694.
- Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919.
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204–220.
- Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat Rev Genet 15(6):394–408.
- Zemach A, et al. (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153(1):193–205.
- Stroud H, et al. (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol 21(1):64–72.
- Law JA, et al. (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498(7454):385–389.
- 8. Johnson LM, et al. (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* 507(7490):124–128.
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* 408(6814):796–815.
- West PT, et al. (2014) Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS One 9(8):e105267.
- Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. Science 326(5956):1112–1115.
- Gent JI, et al. (2013) CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. Genome Res 23(4):628–637.
- Gent JI, et al. (2014) Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation. Plant Cell 26(12):4903–4917.
- Regulski M, et al. (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res 23(10):1651–1662.
- Secco D, et al. (2015) Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. eLife 4, 10.7554/eLife.09343.
- Eichten SR, et al. (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. Plant Cell 25(8):2783–2797.
- Turco G, Schnable JC, Pedersen B, Freeling M (2013) Automated conserved noncoding sequence (CNS) discovery reveals differences in gene content and promoter evolution among grasses. Front Plant Sci 4:170.
- Sekhon RS, et al. (2013) Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS One 8(4):e61005.
- Li Q, et al. (2015) Examining the causes and consequences of context-specific differential DNA methylation in maize. Plant Physiol 168(4):1262–1274.
- Dorweiler JE, et al. (2000) mediator of paramutation 1 is required for establishment and maintenance of paramutation at multiple maize loci. Plant Cell 12(11): 2101–2118.
- Alleman M, et al. (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442(7100):295–298.

sequencing data. WGBS and RNA-seq data for B73 shoot apex, immature ear and anther tissue, H3K4me3 ChIP-seq for B73 seedling leaf tissue, RNA-seq data for mop1, mop3, and wild-type siblings earshoot/seedling leaf, and WGBS data for PH207 seedling leaf were generated for experiments described in this paper. In addition the analyses in this study used previously published WGBS data for B73 seedling leaf and RNA-seq data for B73, Mo17, Oh43, Tx303, and CML322 seedling leaf from Eichten et al. (16), H3K9me2 ChIP-seq data for B73 seedling from West et al. (10), WGBS data for Mo17, Oh43, Tx303, and CML322 seedling leaf from Li et al. (19), WGBS data for mop1 earshoot and targeted bisulfite sequencing data for mop1 earshoot, B73, mop2, and mop3 seedling leaf from Li et al. (22), and chromatin accessibility data for B73 earshoot from Gent et al. (13).

Data Analysis. All datasets were aligned to the AGPv2 B73 reference sequence (11) or the PH207 reference sequence. Annotations of genes and transposons were obtained from ftp://ftp.gramene.org/pub/gramene/maizesequence.org/release-5b/. DNA methylation, ChIP-seq, and chromatin accessibility data were calculated for 100-bp nonoverlapping sliding tiles in the maize genome at these tiles were annotated based on location relative to genes or transposons using BEDTools (32). Details for the analysis of gene expression, DNA methylation, and ChIP-seq data are available in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank University of Minnesota Genomics Center for library preparation and sequencing. Data analysis used tools and resources provided by the Texas Advanced Computing Center at the University of Texas at Austin. This work was supported by National Science Foundation Grants DBI-1237931 (to N.M.S., M.W.V., I.M., and D.L.) and 0607123 (to R.K.D.). C.D.H. was supported by National Science Foundation National Plant Genome Initiative Postdoctoral Fellowship in Biology Grant 1202724.

- Li Q, et al. (2014) Genetic perturbation of the maize methylome. Plant Cell 26(12): 4602–4616.
- Sidorenko L, et al. (2009) A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet 5(11):e1000725.
- Sloan AE, Sidorenko L, McGinnis KM (2014) Diverse gene-silencing mechanisms with distinct requirements for RNA polymerase subunits in Zea mays. Genetics 198(3):1031–1042.
- Li Q, et al. (2015) Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res 43(12):e81.
- 26. Robinson JT, et al. (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26.
- Woodhouse MR, Freeling M, Lisch D (2006) The mop1 (mediator of paramutation1)
 mutant progressively reactivates one of the two genes encoded by the MuDR
 transposon in maize. Genetics 172(1):579–592.
- Jia Y, et al. (2009) Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet 5(11):e1000737.
- Zhong X, et al. (2012) DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol 19(9):870–875.
- Li S, et al. (2015) Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in Arabidopsis reveals features and regulation of siRNA biogenesis. Genome Res 25(2):235–245.
- 31. Zheng Q, et al. (2013) RNA polymerase V targets transcriptional silencing components to promoters of protein-coding genes. *Plant J* 73(2):179–189.
- Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.
- Xi Y, Li W (2009) BSMAP: Whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10:232.
- Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11):1571–1572.
- Schultz MD, Schmitz RJ, Ecker JR (2012) 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. Trends Genet 28(12):583–585.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10–12.
- Kim D, et al. (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14(4):R36.
- 38. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with highthroughput sequencing data. Bioinformatics 31(2):166–169.
- 40. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106.
 41. Eichten SR, et al. (2012) Spreading of heterochromatin is limited to specific families
- of maize retrotransposons. *PLoS Genet* 8(12):e1003127.
 42. Larkin MA, et al. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):
- 42. Larkin MA, et al. (2007) Clustal W and clustal X version 2.0. Bioinformatics 23(21): 2947–2948.
- 43. Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23(6):657–663.