

# The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA

Michael Regulski,<sup>1,5</sup> Zhenyuan Lu,<sup>1,5</sup> Jude Kendall,<sup>1,5</sup> Mark T.A. Donoghue,<sup>1</sup> Jon Reinders,<sup>2</sup> Victor Llaca,<sup>2</sup> Stephane Deschamps,<sup>2</sup> Andrew Smith,<sup>3</sup> Dan Levy,<sup>1</sup> W. Richard McCombie,<sup>1</sup> Scott Tingey,<sup>2</sup> Antoni Rafalski,<sup>2</sup> James Hicks,<sup>1</sup> Doreen Ware,<sup>1,4</sup> and Robert A. Martienssen<sup>1,6</sup>

<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; <sup>2</sup>DuPont Crop Genetics Research, Experimental Station, Wilmington, Delaware 19880, USA; <sup>3</sup>Molecular and Computational Biology, University of Southern California, Los Angeles, California 90089, USA; <sup>4</sup>USDA-ARS NAA, Robert W. Holley Center for Agriculture and Health, Ithaca, New York 14853, USA

The maize genome, with its large complement of transposons and repeats, is a paradigm for the study of epigenetic mechanisms such as paramutation and imprinting. Here, we present the genome-wide map of cytosine methylation for two maize inbred lines, B73 and Mo17. CG (65%) and CHG (50%) methylation (where H = A, C, or T) is highest in transposons, while CHH (5%) methylation is likely guided by 24-nt, but not 21-nt, small interfering RNAs (siRNAs). Correlations with methylation patterns suggest that CG methylation in exons (8%) may deter insertion of *Mutator* transposon insertion, while CHG methylation at splice acceptor sites may inhibit RNA splicing. Using the methylation map as a guide, we used low-coverage sequencing to show that parental methylation differences are inherited by recombinant inbred lines. However, frequent methylation switches, guided by siRNA, persist for up to eight generations, suggesting that epigenetic inheritance resembling paramutation is much more common than previously supposed. The methylation map will provide an invaluable resource for epigenetic studies in maize.

[Supplemental material is available for this article.]

Maize exhibits a wealth of epigenetic phenomena, from transposon silencing, cycling, and presetting, to gene imprinting and paramutation. Furthermore, despite the complexity and sophistication of maize breeding, there is a large degree of “hidden” variation for many traits that is difficult to explain by allelic variation alone (Gottlieb et al. 2002). At least some of this unexplained variation might be due to epigenetic rather than genetic changes in the maize genome (Richards 2011). A recent study using anti-methylcytosine antibodies and microarray hybridization to detect DNA methylation demonstrated clear differences between maize inbred lines, lending support to this hypothesis (Eichten et al. 2011). Similarly, studies using genome-wide sequencing of methylation-dependent restriction fragments have revealed that most methylation is found in transposable elements, prime sources of such variation (Palmer et al. 2003; Wang et al. 2009). However, neither method had the capability to detect individual cytosines in their sequence context.

In plants, cytosine methylation occurs in symmetric (CG and CHG, where H is A, C, or T) as well as asymmetric (CHH) contexts. Methylation in each context is associated with DNA replication, histone modification, and RNA interference, respectively, although these mechanisms overlap (Law and Jacobsen 2010). The maize genome comprises roughly 50,000 genes and more than 1

million transposons and related repeats (Schnable et al. 2009). Approximately 29% of the cytosine residues are methylated as 5-methylcytosine (Montero et al. 1992), mostly in transposons (Palmer et al. 2003). We have used very-high-coverage whole-genome bisulfite sequencing to explore DNA methylation at nucleotide resolution in genes, transposons, and other features of the maize genome, as well as its heritability and potential contribution to traits. We have found that methylation in different sequence contexts is guided differentially by small RNA and is correlated with transposon insertion and mRNA splicing. Heritable and predictable switches in DNA methylation were detected in recombinant inbred lines. These shifts were apparently triggered by small RNA, resembling paramutation, but then maintained by replication-dependent symmetric methylation.

We present a high-resolution and high-coverage map comprising the methylation status of individual cytosines throughout the inbred maize genome. Using this resource, we demonstrate that methylome sequencing of recombinant inbred lines at much lower coverage is sufficient to detect widespread paramutation in the maize genome. Future studies using low-coverage methylome sequencing can take advantage of this resource to determine the impact of differentially methylated regions on gene expression, chromosome biology, and transgenerational inheritance. For example, this will allow breeders to determine the contribution of cytosine methylation to phenotypic variation among elite inbreds and hybrids, artificially induced chromosomal variants (such as doubled haploids), and clonally micropropagated strains, which are subject to such epigenetic variation (Richards 2011). Thus, breeders could deploy a form of “epigenomic selection,” analogous to genomic selection, by which individuals

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>Corresponding author  
E-mail [martiens@cshl.edu](mailto:martiens@cshl.edu)

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.153510.112>. Freely available online through the *Genome Research* Open Access option.

with desired epigenomic patterns could be retained in breeding programs.

## Results and Discussion

### Bisulfite sequencing strategy

Treatment of DNA with sodium bisulfite converts unmethylated cytosines into thymines and allows generation of “methylomes” at single-base resolution (Harris et al. 2010). We used Illumina next-generation instruments to sequence bisulfite-converted genomic DNA from two reference maize strains, B73 and Mo17, at a coverage that allowed us to establish methylation states of individual cytosines with high confidence (all libraries are listed in Supplemental Table S1). We also sequenced bisulfite-converted libraries from nine recombinant inbred lines (RILs) at lower coverage to investigate epigenetic heritability (Table 1). Finally, we sequenced cDNA libraries from mRNA and small RNA from the same tissue as the methylation libraries, namely, shoots at the coleoptilar stage 5 d after germination (Methods). Reads were aligned to the reference genome of maize B73, taking into account conversion as well as polymorphisms with Mo17 (Methods).

### Genome-wide patterns of DNA methylation

The percentage of methylated cytosines varied depending on the local sequence context (CG, CHG, and CHH) (Supplemental Fig. S1) and on the surrounding genomic features. This can be visualized on a whole-genome scale with a chromosomal plot (Fig. 1A) or as a browsable graphical interface (Fig. 1B). Methylation patterns derived from bisulfite sequencing were validated by comparison with those generated by propagation of sequencing clones in McrBC<sup>+</sup> *Escherichia coli* strains (a method known as methyl filtration) (Fig. 1A; Palmer et al. 2003). In excellent agreement with our bisulfite sequencing results, the mapping results from methyl filtration confirm that regions of dense methylation correspond to pericentromeric regions of each chromosome, which have the lowest numbers of genes and the highest number of repeats (Schnable et al. 2009). These regions, presumably heterochromatin, are most heavily methylated in symmetric CG and CHG rather than asymmetric CHH contexts (Fig. 1B). Reported results from methyl filtration had previously indicated that some centromeric repeats were largely devoid of methylation (Palmer et al. 2003). Our bisulfite sequencing shows that several domains within the pericentromeric heterochromatin, including the centromeric repeats, are hypomethylated in asymmetric (CHH) contexts (Fig. 1A). Recently, chromatin immunoprecipitation with

antibodies to the centromeric histone CenH3 also recovered DNA depleted for CHH methylation (Gent et al. 2012). However, coverage was insufficient in these studies to detect the depletion of CHH methylation in the pericentromeric region as a whole.

### Methylation of transposable elements and repeats

The vast majority of transposons in the maize genome are found in intergenic and nongenic regions and are highly variable between different inbred strains (Baucom et al. 2009; Schnable et al. 2009; Wei et al. 2009). We used a reference database (<http://www.maizetdb.org>) to map reads onto “exemplar” elements of each of the major transposon classes found in B73 (Baucom et al. 2009). We compared the methylation levels of all transposon classes in B73 and Mo17 as scatterplots in each methylation context (Supplemental Fig. S2). Similar to previous results in inbred accessions of *Arabidopsis* (Vaughn et al. 2007; Becker et al. 2011; Schmitz et al. 2011), transposon methylation was highly conserved between the two inbreds, with fewer than 50 of 1526 transposon exemplars showing differential methylation (Supplemental Tables S2–S4). We note, however, that quantitative variation in expression of individual copies of high-copy repeats remains difficult to detect. CHG and CHH levels were more uniformly if less densely methylated than CG levels and varied mostly in DNA transposons, especially in miniature inverted-repeat transposable elements (MITES) and their progenitors, which have relatively low copy numbers (Baucom et al. 2009).

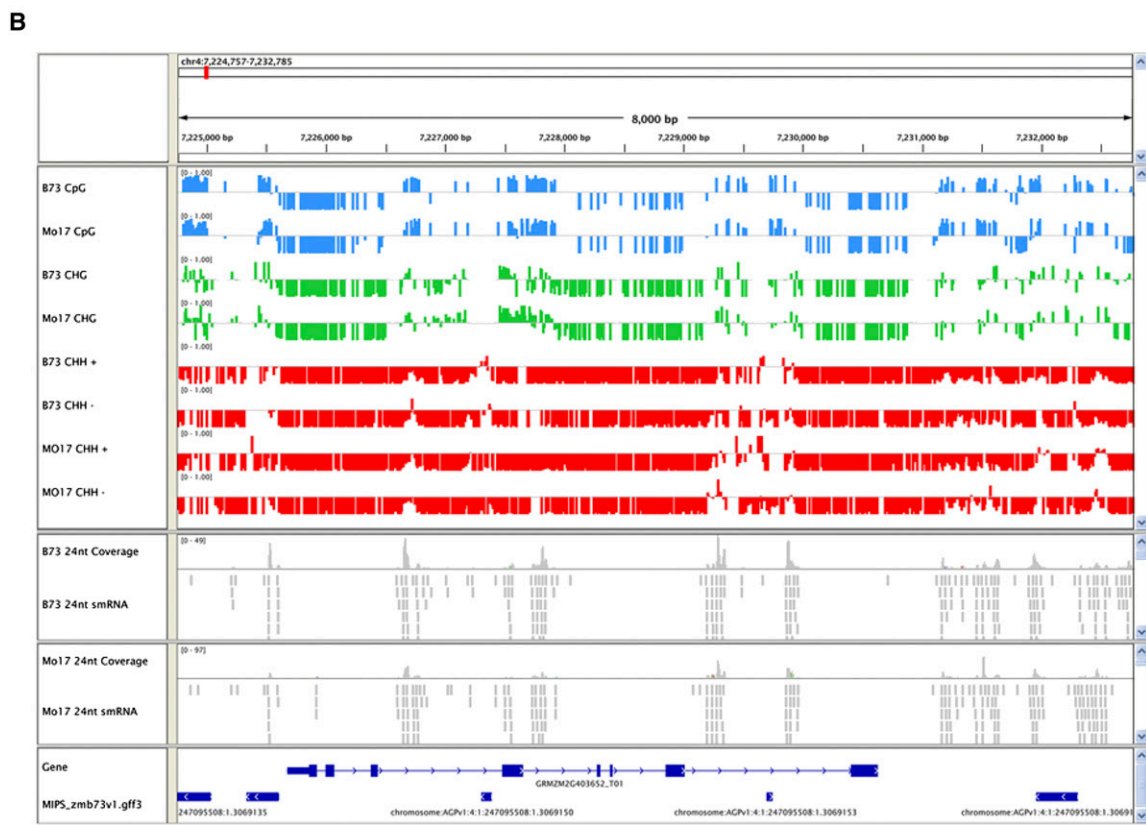
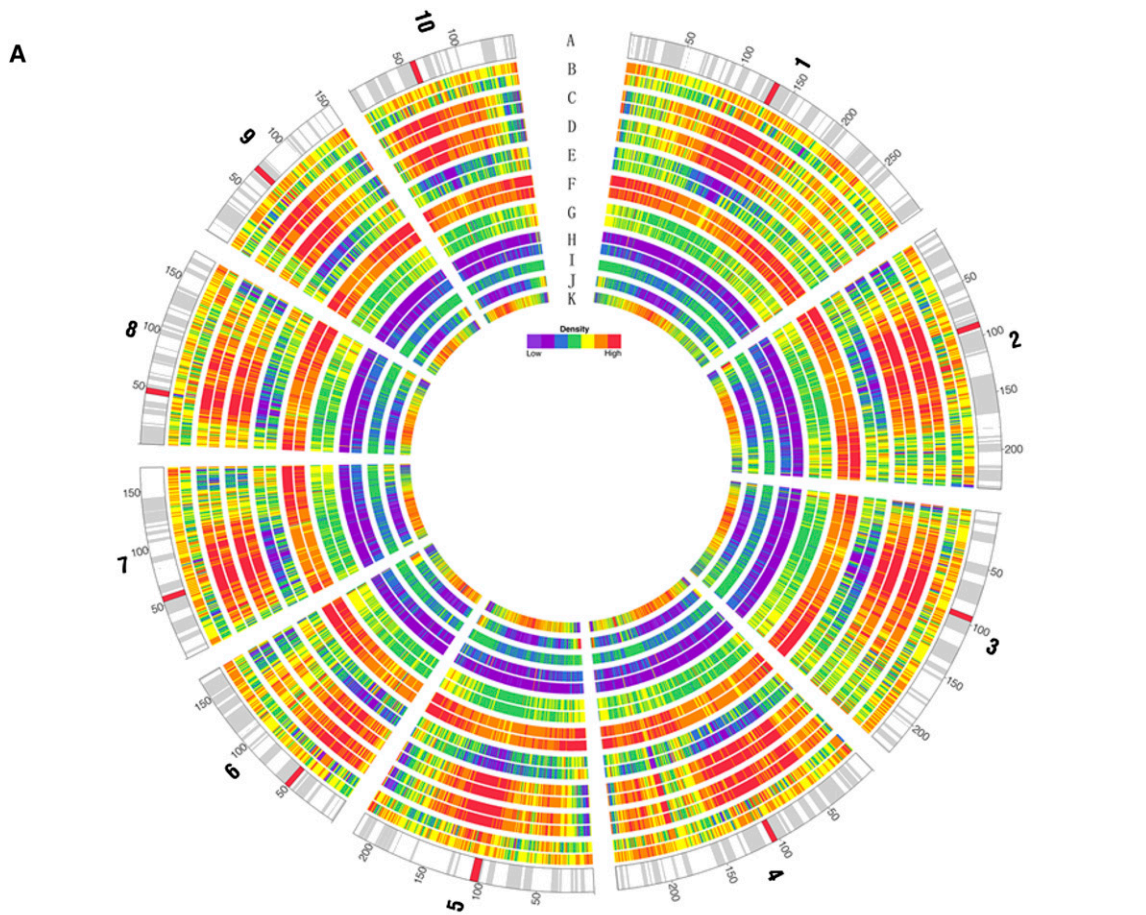
Next, we compared methylation levels in each context with transposon copy number, which we estimated from the number of transposon-matching reads in each library (Supplemental Fig. S3). We found that TE copy number is not correlated with CHH methylation (B73:  $\rho = 0.0342$ , Mo17:  $\rho = 0.0253$ ) as both high- and low-copy TEs display high levels of methylation. In contrast, CHG and, in particular, CG methylation is positively correlated with TE copy number, even though copy number varied between the accessions (B73: CHG  $\rho = 0.1102$ , CG  $\rho = 0.2144$ , Mo17: CHG  $\rho = 0.2792$ , CG  $\rho = 0.4184$ ). We then compared transposon methylation with mRNA levels in RNA-seq libraries prepared from the same tissues. We found most transposons to be expressed at very low levels, and mRNA from methylated and unmethylated transposons accumulated to similar levels, with only a moderate correlation between CG methylation and TE expression (B73:  $\rho = 0.1072$ , Mo17:  $\rho = 0.2108$ ) (Supplemental Fig. S4), suggesting that higher copy number is largely balanced by higher levels of repression.

Finally, we compared methylation levels with matching small RNA isolated from the same tissues (Supplemental Fig. S5). Asymmetric CHH methylation in *Arabidopsis* reflects persistent de novo methylation guided by small RNA (Law and Jacobsen 2010), and thus we anticipated a strong correlation between small RNA and CHH DNA methylation levels in maize. For pericentromeric and intergenic transposons, 24-nt siRNA was, indeed, positively correlated with CHH methylation (Fig. 1A). In contrast, transposons with high levels of 21-nt and 22-nt siRNAs had very low levels of CHH methylation (Supplemental Fig. S5). We next examined small RNA and genome methylation in repetitive windows across the entire genome, regardless of transposon annotation (Fig. 2). We found that the fraction of CHH methylation in each window had a strong positive correlation with levels of 24-nt siRNA and a moderate negative correlation with levels of 21-nt siRNA (Fig. 2). In *Arabidopsis*, transposons generate 21-nt siRNA when they become unmethylated in rapidly dividing “immortalized” cell culture (Tanurdzic et al. 2008), and in mutants defective in

**Table 1. Bisulfite sequencing summary**

Strain	Total bases (Gb)	Coverage depth	Coverage width	Bisulfite conversion rate
B73	140	28.8	95.50%	98.20%
Mo17	134	19.7	88.20%	98.50%
RIL 1-9	170	3.7	74.40%	99.01%

Coverage depth is the average number of reads per base within the genomic space uniquely mapped with 50-bp reads. Coverage width is the number of bases covered within the space mapped with 50-bp reads. Bisulfite conversion rate was based on the percentage of the essentially unmethylated chloroplast genome converted. For RILs, total bases covered by sequencing reads were from all nine RILs, but depth, width, and bisulfite conversion rate are the average per RIL.



**Figure 1.** (Legend on next page)

DNA methylation (Slotkin et al. 2009). It is likely therefore that 21-nt siRNA in maize corresponds to active unmethylated transposons in each inbred line.

The centromeric regions of maize chromosomes comprise thousands of copies of the tandem repeat CentC, as well as hundreds of copies of the conserved cereal retrotransposon *CRM* (Schnable et al. 2009). Both 22-nt and 24-nt siRNAs could be detected in both types of repeat but were rare in *CRM*, and CHH methylation was very low (2%–5%) compared with CG and CHG methylation (Fig. 1A; Supplemental Table S5). These results are in agreement with previous results indicating that transcripts from these repeats are abundant, but not necessarily processed into small RNA (Topp et al. 2004).

### Gene body methylation may deter transposon insertions

Gene body methylation has been widely observed in eukaryotes, but its function is still unclear. Focusing on the genic regions and surrounding promoters, we constructed a “metagene” methylation profile covering the exons of “gene bodies” and extending from 2 kb upstream of to 2 kb downstream from the transcript (Fig. 3A). We found that CHG methylation was suppressed in gene bodies in each inbred, while CG methylation accumulated at low levels, relative to the surrounding DNA. CHH methylation remained at low levels, but accumulated at the 5′ and 3′ ends (Gent et al. 2012). To study the relationship between gene methylation level and gene expression level, RNA-seq libraries were generated in triplicate for both B73 and Mo17 samples. The RNA-seq reads were processed using the working gene set (Methods) as the gene model. We divided the genes into five levels and built methylation profiles for each level (Fig. 3B–D). Genes with higher expression levels had slightly lower CG methylation around the start and end of the gene and elevated CG methylation in the middle of the gene body (Fig. 3B), while the genes with a low expression level had elevated levels of CHG methylation overall (Fig. 3C). A similarly mixed correlation has been observed in *Arabidopsis* (Cokus et al. 2008; Lister et al. 2008), and gene body methylation does not greatly impact gene expression (Hollister et al. 2011; Yang et al. 2011). However, compared with *Arabidopsis*, we found that intergenic DNA in maize was much more highly methylated than gene bodies, reflecting the high density of intergenic retrotransposons (Baucom et al. 2009).

In a previous study, insertion sites of the extremely aggressive transposon Robertson’s *Mutator* (*Mu*) were mapped genome-wide and found to insert preferentially within unmethylated regions of the B73 genome, although the methylation context could not be determined (Liu et al. 2009). *Mu* transposons are well known to insert preferentially within genes, and so we compared methylation in the sequences surrounding each insertion site with methylation in exons and entire gene bodies. We found that *Mutator* insertion sites within genes were strongly depleted of CG methylation, but were not depleted of CHG and CHH methylation relative to gene body methylation overall (Supplemental Table S6). As in *Arabidopsis* (Hollister et al. 2011; Yang et al. 2011), we found

that CG methylation in maize was more prevalent in exons than introns (Fig. 4A,B). It is therefore likely that CG methylation of exons, such as those in highly expressed genes, deters transposon insertions that might otherwise disrupt gene function, while allowing insertions into introns and other noncoding regions.

### Splice-site methylation inhibits alternative splicing

Like in *Arabidopsis*, CHH and CHG methylation was largely excluded from gene bodies (Hollister et al. 2011; Yang et al. 2011), but a notable exception lay at the intron–exon junctions. The donor and acceptor site consensus sequences in plants are AAG<sup>^</sup>GTAAG and TTTGCAG<sup>^</sup>GT, respectively (Reddy 2007), with rare non-consensus U12-splice donor sites CAG<sup>^</sup>GCAAG (Sheth et al. 2006). Cytosines at these sites were sometimes methylated on the sense and the antisense strands (Fig. 4A,B, respectively). We hypothesized that this methylation might influence splicing efficiency and/or alternate splicing.

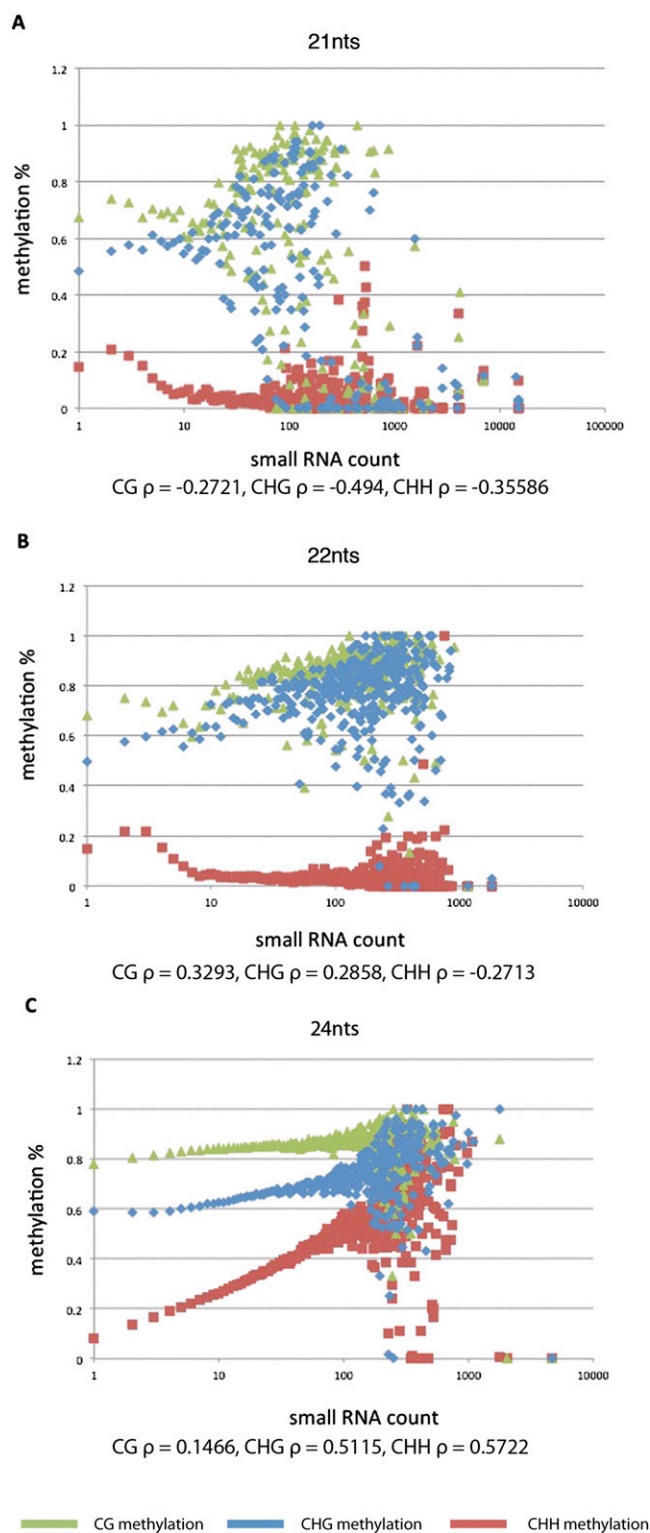
To assess the influence of DNA methylation on splicing, we first collected donor and acceptor sites genome-wide and divided them into low and high methylation categories. Methylation was measured at cytosines on both strands in each methylation context within the donor GT (or GC) and acceptor AG dinucleotides (Methods). Next, we scored the number of RNA-seq reads that corresponded to spliced and unspliced transcripts at each site (Supplemental Table S7). Strikingly, we found that acceptor sites with high levels of CHG methylation were much less efficiently spliced than sites with low levels of methylation. Next, we collected those genes in the genome that were alternatively spliced in at least one tissue (from ZmB73 4a.53 WGS; <http://ftp.maizesequence.org/release-4a.53/working-set/>), for which we had sufficient RNA-seq data to examine splicing, and for which the alternate exons displayed differing levels of methylation at their acceptor sites. Although only 28 genes met these criteria, 23 of them preferentially used the acceptor site with reduced CHG methylation (Fig. 4C). In contrast, CHH methylation at the donor site did not correlate with splicing efficiency.

Exon methylation in human cells has been proposed to influence splicing by recruiting the DNA binding protein CTCF and slowing down transcription, favoring splicing of adjacent exons (Shukla et al. 2011). However, plants do not possess CTCF, and the CHG methylation we report here does not occur in mammals. While the mechanism by which methylation influences splicing in maize is unclear, CHG methylation in plants can be guided by small RNA and by histone H3K9 methylation, and we speculate that splice site methylation might be mediated by these mechanisms.

### Inheritance and functional significance of differential methylation

In addition to determining methylation profiles for the individual inbreds, we also sought to identify differences in methylation between different maize lines. In unique regions of the genome, we identified 9635 highly differentially methylated regions (HDMRs)

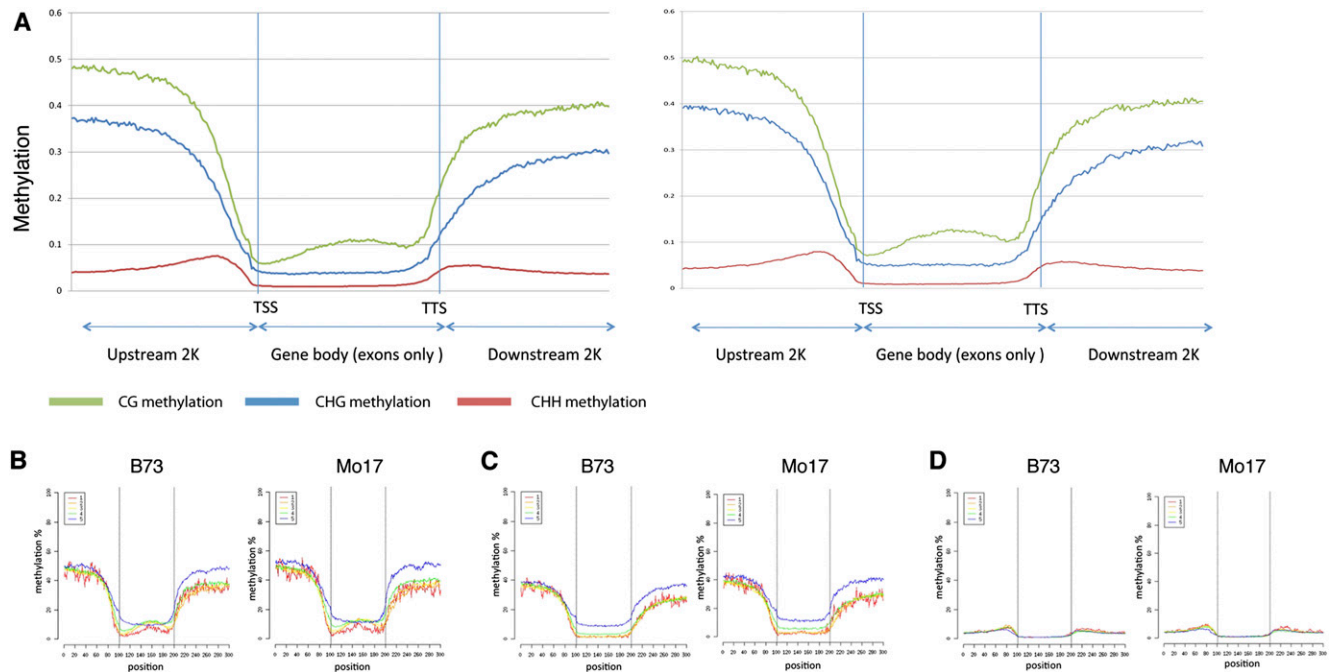
**Figure 1.** Genome-wide patterns of DNA methylation. (A) Circle diagram of all 10 maize chromosomes. The tracks represent sequence scaffolds with centromeric regions labeled in red (A), read coverage (B), methylation at CG (C), CHG (D), and CHH (E), 24-nt (F), 22-nt (G), and 21-nt (H) small RNAs. Within each pair, the outer track is B73, the inner track is Mo17. Track I represents methyl filtration reads, J genes, and K repeats from the B73 genome. “High” on the Density scale means a high density of features presented in a given track. (B) An image from the Integrative Genomics Viewer software (Robinson et al. 2011) showing information from an 8-kb region on chromosome 4 (7,224757–7,232785 bp). The first eight data tracks show methylation levels for B73 and Mo17 in different sequence contexts. Methylation levels are displayed on a scale from 0 to 1. The 24-nt small RNA coverage graphs indicate coverage depths for positions with 24-nt small RNA sequencing data for B73 and Mo17, respectively.



**Figure 2.** Genome methylation and small RNA. Each point represents a contiguous window of sequence that matches increasing numbers of smRNA reads in size classes of 21 (A), 22 (B), or 24 nt (C) on the x-axis (small RNA count). (Small RNA count) Number of smRNA reads that map to this window and nowhere else with the same number or fewer mismatches. The average methylation level (0–1.00) of each window is plotted on the y-axis (see Methods). Methylation was calculated in CG (green), CHG (blue), and CHH (red) contexts. ( $\rho$ ) Spearman's rank correlation coefficient ( $P$ -value < 0.001).

with a substantial difference in CG, and 6671 HDMRs in CHG context between B73 and Mo17 (Supplemental Tables S9, S10). HDMRs were chosen such that they were <10% methylated in one line and >70% methylated in the other, at least 200 bp long, and had at least 20  $m^5C$  sites with data (Methods). DNA methylation in the CHH context in maize is qualitatively different from in the CG and CHG contexts. Whereas in the CG and CHG contexts the genome is largely methylated at  $\sim 70\%$  with sparsely distributed regions of a few hundred bases with very low levels of methylation, typically <5%, in the CHH context the genome is mostly methylated at a very low level with narrow peaks of methylation where a few cytosines are methylated above 50%. Owing to these qualitative differences, we analyze methylation in the CHH context by different methods. HDMRs represent clusters of DNA methylation differences and may potentially serve as a foundation for epigenetic biomarkers (Schmitz et al. 2011). HDMRs with elevated methylation in all contexts have been previously shown to behave as epialleles in *Arabidopsis* (Vaughn et al. 2007; Becker et al. 2011; Schmitz et al. 2011) and arise along with the presence of 24-nt siRNA (Teixeira et al. 2009). In this respect, epialleles resemble cycling transposons in maize that arise in a spontaneous or in a developmentally programmed fashion (Martienssen and Colot 2001) and are regulated by RNA-dependent DNA methylation (Slotkin and Martienssen 2007).

To investigate the inheritance of these HDMRs, we compared their distribution in parents and in nine matched recombinant inbred lines (RILs) derived from the intermated B73  $\times$  Mo17 (IBM) population (Sharopova et al. 2002). Methylation profiles were generated for each of these lines, and the 9635 epipolymorphic regions were scored for HDMRs as before. One line was sequenced three times to assess reproducibility, and no significant differences were found ( $r^2 = 0.95$ – $0.97$ ). Chromosome-wide analysis revealed a close correlation of CG methylation patterns in B73 and Mo17 with corresponding SNP haplotypes in RILs on all 10 chromosomes (Supplemental Fig. S6). To examine this more closely, we mapped 1335 genetic markers in the RILs and were able to identify the parental origin of each segment, with distinct breakpoints reflecting recombination (Fig. 5A). Interestingly, while the majority of methylation polymorphisms were inherited along with the parental segment, there were many exceptions in which a gain (or occasionally, a loss) of methylation shifted the methylation level to the other allele within a large epihaplotype. Specifically, 1772 HDMRs switched epigenotypes: 5.1%–7.2% gained methylation in an RIL, while only 2.1%–2.3% lost methylation (Supplemental Table S10). For both Mo17 and B73, RILs switching to hypermethylated are more common than RILs becoming hypomethylated. Additionally, RILs becoming hypermethylated from a Mo17 background are more frequent than RILs becoming hypermethylated from a B73 background. In all cases, the  $P$ -value is  $< 2.2 \times 10^{-16}$  by a  $\chi^2$  test. In total,  $\sim 6$  Mb of DNA was differentially methylated between the two parents, of which 0.9 Mb switched methylation state in the RILs. Many HDMR switched in more than one RIL, with up to seven out of 10 RILs gaining methylation at the same HDMR (Supplemental Table S11). To assess predictability of frequency of these switches, a simulation was run in which methylation levels in each segment were sampled randomly with replacement (Methods). A  $t$ -test indicated that each class of switch was much more recurrent than expected by chance ( $P < 10^{-16}$ ). HDMRs that consistently switch in the RILs from unmethylated to methylated are potential candidates for epialleles. These can be investigated by looking at how many of the HDMRs become hypermethylated in all the RILs we have the opportunity to observe. In the CG

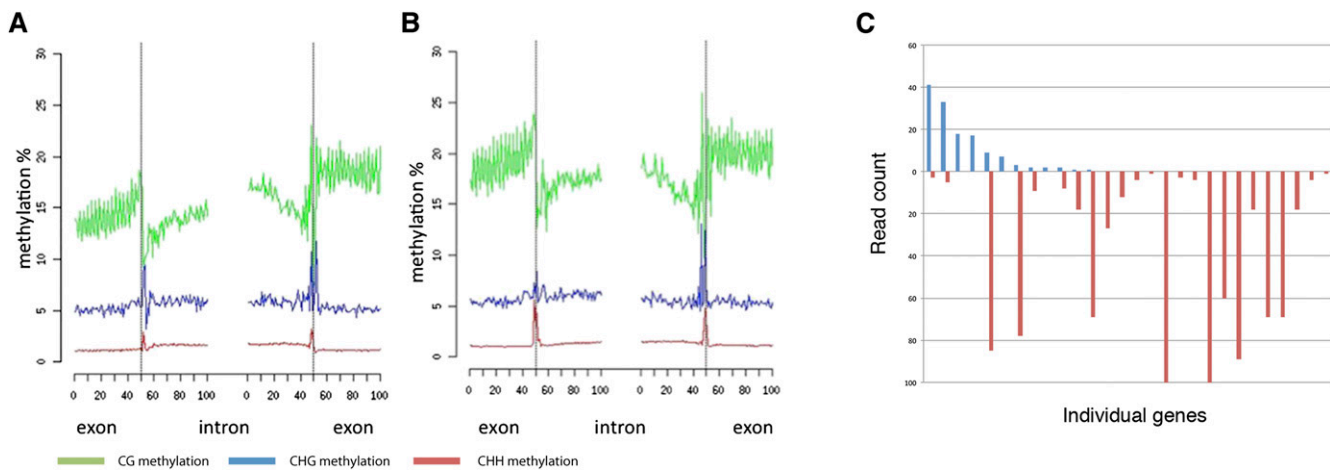


**Figure 3.** Exon methylation. (A) Methylation levels in each context, CG (green), CHG (blue), and CHH (red), are summed across all exons via the “metagene” analysis (see Methods) for each inbred. Genes were divided according to expression levels ([1] high; [5] low), and CG methylation (B), CHG methylation (C), and CHH methylation (D) was plotted for each inbred.

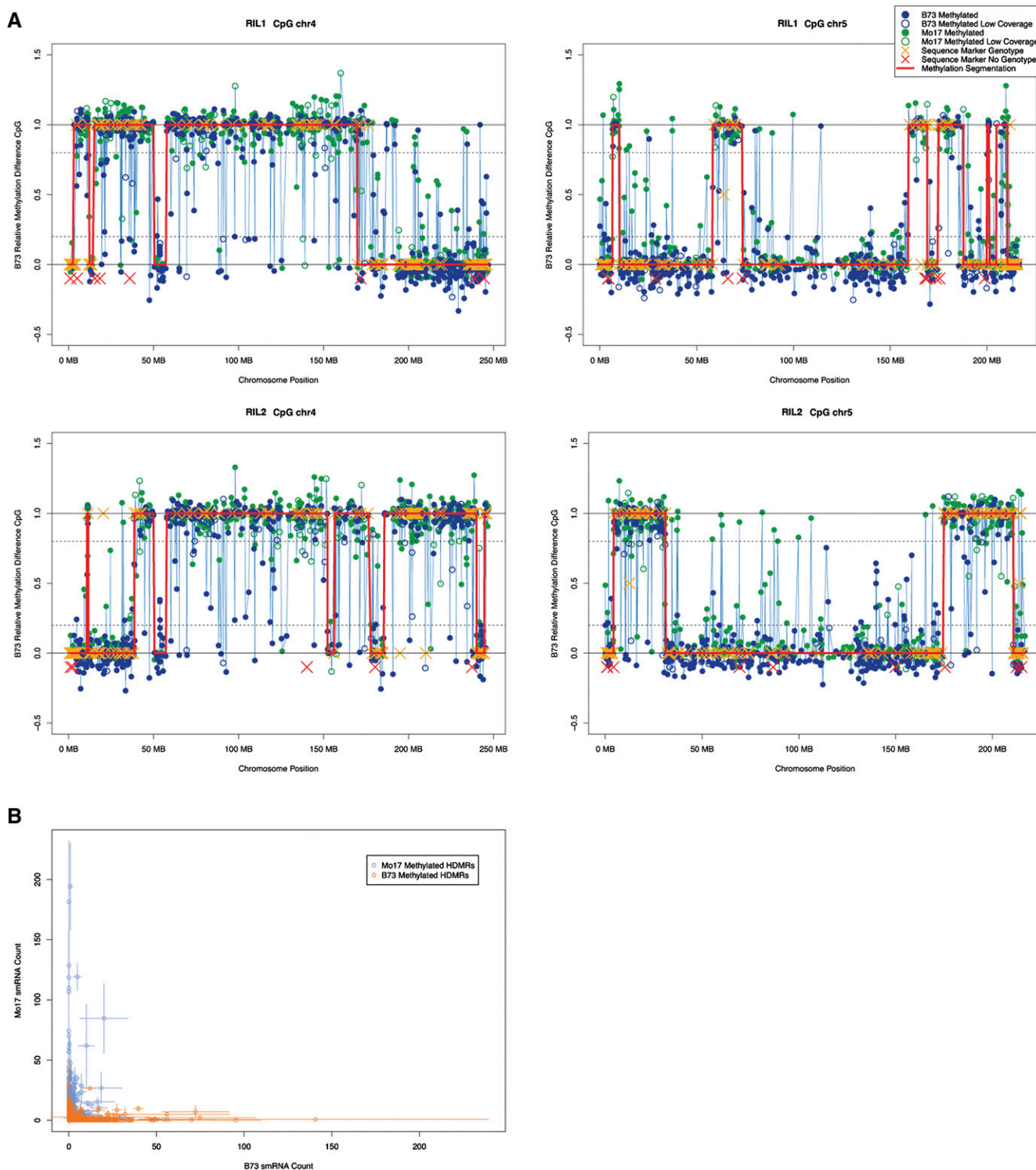
context there are 8543 HDMRs where at least three of the RILs are hypomethylated in either B73 or Mo17 genetic backgrounds. Of these, 1392 have at least one RIL hypermethylated, and 207 of these are hypermethylated in all RILs observed. In the CG context, there are 5920 HDMRs where at least three of the RILs are hypomethylated in either B73 or Mo17 genetic backgrounds. Of these, 944 have at least one RIL hypermethylated, and 96 of these are hypermethylated in all RILs observed.

Scatterplots were constructed of small RNA levels for each HDMR in each parental inbred (Fig. 5B). While some HDMRs were

restricted to CG methylation, most also affected CHG methylation (data not shown), and very strong positive correlations between small RNA presence and methylation context were observed (Figs. 1, 5). At high resolution, it was clear that the regions corresponding to abundant small RNA matches were also highly methylated at CHH sites (e.g., Fig. 1B), as expected from the genome-wide correlation (Fig. 2), but the flanking regions were methylated in a symmetric CG or CHG context. This pattern is consistent with RNA-dependent DNA methylation being triggered by small RNA in hybrids and then spread by symmetric DNA methylation. Focus-



**Figure 4.** Splice site methylation. (A) Metagene analysis of splice junctions on the antisense strand, in CG (green), CHG (blue), and CHH (red) contexts. (B) Metagene analysis of splice junctions on the sense strand, in CG (green), CHG (blue), and CHH (red) contexts. (C). Alternatively spliced genes ( $n = 28$ ) analyzed for CHG methylation levels at alternative acceptor sites (Supplemental Table S7B). The number of RNA-seq reads is plotted at alternate acceptor sites with high (blue, above the line) and low (red, below the line) methylation for each individual gene.



**Figure 5.** Inheritance and stability of DNA methylation patterns. (A) 9635 regions in the genome where B73 and Mo17 have different CG methylation levels were found using a moving window and a  $\chi^2$  test based on C and T counts at CG sites. The regions were chosen so that one parent is <10% methylated and the other >70%. The regions are at least 200 bp long and have at least 20 CG sites. There is one point on the plots for each of these 9635 regions. The x-axis is the region number in chromosome order. The y-axis value is  $(\%C \text{ in RIL} - \%C \text{ in B73}) / (\%C \text{ in Mo17} - \%C \text{ in B73})$ . Thus, data points approaching 1 are similar to Mo17, and those approaching 0 to B73 in methylation level. Mapped (orange) and unmapped (red) markers indicate recombination break points. (B) Scatterplots of small RNA levels that match HDMR in each inbred. Error bars represent variation in small RNA levels found in three biological replicates.

ing only on splice acceptor sites, we found that 1241 acceptors in Mo17 retained high CHG methylation and 249 switched to low methylation. The ratio of spliced-to-unspliced transcripts at these junctions was 4.1 and 64.9, respectively, indicating that differential methylation between inbreds has profound effects on splicing. We have also checked for potential association between the methylation state of HDMRs and expression levels of nearby genes. We have identified 4071 genes located within 1 kb of HDMRs, and 1598 were differentially expressed between B73 and Mo17. We found a significant correlation between hypomethylated regions and overexpression of genes within a 1-kb range, for both CG and CHG contexts (Fisher's exact test, CG:  $P < 2.20 \times 10^{-16}$ ; CHG:  $P < 2.20 \times 10^{-16}$ ) (Supplemental Table S8).

## Conclusions

We have determined the nucleotide-resolution high-coverage methylation map of the maize genome. Methylation of transposons and repeats is highly conserved between inbred lines, but methylation of gene bodies is much more variable and includes splice site and exon methylation in differing contexts. CHG acceptor splice site methylation is correlated with reduced splicing efficiency, while CG methylation of exons seems to deter disruption of highly expressed genes by transposon insertion. Differential methylation between inbred lines is largely heritable, and correlated with expression of nearby genes, but thousands of HDMR throughout the genome shift from one epiallele to the other and are stably inherited in recombinant inbred lines.

While the mechanism that guides these shifts remains unknown, we considered the possibility that they might be under genetic control of unlinked regulatory loci. An example of such a locus is the *PAI1-4* locus in *Arabidopsis*, which has allelic forms that regulate methylation of the unlinked *PAI2* locus in *trans* via small RNA (Bender 2004; Schmitz et al. 2013). However, if HDMRs were under genetic control, they should segregate among the RILs. With only nine RILs, it is difficult to exclude this possibility for all HDMR, but cosegregation was not generally observed, and transgressive methylation (differing from both parents) was extremely rare. Instead, we found that HDMR shifts were recurrent, biased toward one parent or the other, and correlated with matching parental small RNA. For these reasons, we favor the idea that the shifts in HDMR methylation we have observed are due to interactions between the parental alleles in heterozygous hybrid individuals.

In maize, certain heteroallelic combinations result in a heritable shift from the expressed toward the silent allele. This process is known as paramutation (Arteaga-Vazquez and Chandler 2010; Erhard and Hollick 2011) and depends on RNA interference (Alleman et al. 2006) and on RNA-dependent DNA methylation (Hale et al. 2007; Erhard et al. 2009). Paramutation at the *R*, *B*, *P*, and *Pl* loci is restricted to individual alleles associated with repeated sequences, and these alleles are not present in B73 or Mo17, so that these loci are not affected in our RILs (data not shown). Nonetheless, our results suggest that paramutation-like methylation changes affect >10% of all HDMRs in the maize genome and are heritable for up to eight generations. Increases in methylation outnumbered decreases by more than 2:1. Programmed epigenetic changes of this magnitude would have profound implications for maize breeding. For example, traits associated with epialleles would arise in doubled haploids or clonal apomicts just as readily as in sexual progeny and would not be detected in association with single nucleotide polymorphisms (SNPs) (McMullen et al. 2009).

Epiallele interactions could lead to altered gene regulation in hybrids, as well as to inbreeding depression, thereby contributing to heterosis (Hollick and Chandler 1998).

## Methods

### Biological materials

For DNA and RNA preparations, all seeds were grown in the incubator at 25°C in darkness on wet paper towels set in glass Pyrex dishes. After 5 d, shoots at the coleoptilar stage were excised and stored at -80°C.

### Genomic DNA preparation, library construction, and bisulfite conversion

DNA was prepared from nuclei as described in Bedell et al. (2005). In brief, coleoptiles were ground with mortar and pestle in LN<sub>2</sub>, and nuclei were washed and separated from organelles by centrifuging through 30% OptiPrep (Axis-Shield PoC, Oslo, Norway, catalog no. 1114542) solution. Library construction and bisulfite conversion were carried out, essentially, as described in Hodges et al. (2009). For each library, 1–5 µg of genomic DNA was sheared using Covaris S220 Adaptive Focused Acoustics ultrasonicator. Libraries were constructed following standard protocol using the NEBNext DNA Sample Prep Master Mix Set 1 (NEB E6040) and Illumina-compatible paired-end adaptors in which all cytosines were methylated. Each library (50 ng) was treated with sodium bisulfite using the EZ DNA Methylation-Gold Kit (Zymo Research D5005) according to the manufacturer's protocol. Each purified library (5–10 ng) was amplified using the Expand High Fidelity PLUS PCR system (Roche 03300242001), which is capable of efficiently amplifying uracil-containing templates. PCR reactions (50 µL final volume) containing 200 µM each dNTP, 1 µM each primer, 2.5 mM MgCl<sub>2</sub>, and 2.5 units of Expand HiFi enzyme were performed according to the manufacturer's instructions for 18 cycles. Amplified libraries were run on 2% MetaPhor agarose (Lonza 50108) gel. Fragments of 220–350 bp were excised from gel and purified using the QIAquick PCR Purification Kit (Qiagen 28104).

### DNA sequencing on Illumina platform

DNA concentrations were quantified on a Bioanalyzer (Agilent), diluted to 10 nM, and loaded on flow cells to generate clusters. Libraries were sequenced on Illumina GAII or HiSeq2000 machines using the paired-end 50 cycle protocol.

### RNA preparation and RNA-seq library construction

Total RNA was prepared by grinding tissue in TRIzol reagent (Invitrogen 15596-026) on dry ice and processed following the protocol provided by the manufacturer. To remove DNA, an aliquot of total RNA was treated with RQ1 DNase (Promega M6101), followed by phenol:chloroform:isoamyl alcohol extraction, chloroform:isoamyl alcohol extraction, and ethanol precipitation. Total RNA (20 µg) was used for poly(A)<sup>+</sup> selection using oligo(dT) magnetic beads (Invitrogen 610-02), eluted in water, and used for RNA-seq library construction with the ScriptSeq Kit (Epicentre SS10906) or mRNA-seq Sample Prep Kit (Illumina RS-100-0801) according to the manufacturer's protocol. Three libraries from B73 and Mo17 (biological replicates) were each amplified with 15 cycles of PCR generating 485002759 and 636007745 reads, respectively, that aligned to the B73 genome.



### Small RNA library construction

Small RNA fractions (up to 40 nt in length) were isolated by running 20  $\mu$ g of total RNA (prepared as described above) using the FlashPAGE Kit (Ambion AM10010 and AM9015) followed by ethanol precipitation. RNA was resuspended in 8  $\mu$ L of water and processed using the ScriptMiner Kit (Epicentre SMMP101212). The library was amplified using 15 cycles of PCR. Alternatively, samples of total RNA were sent to FASTERIS (Switzerland) for library construction and sequencing. Three libraries from B73 and Mo17 each (biological replicates) generated 14,234,461 and 10,761,747 reads, respectively, that aligned to the B73 genome uniquely. The same libraries generated 4,841,242 and 4,816,161 reads, respectively, that aligned to the TE exemplar database uniquely.

### Processing of bisulfite sequencing data and mapping of reads

Reads from the IlluminaGA pipeline were trimmed to remove the low-quality bases at the 3' end using the same algorithm as used in BWA and then, adaptor/linker sequences at the 3' ends were trimmed using the FASTX-Toolkit. Reads were then mapped to maize reference genome v1 and the transposable element sequence library from the maize transposable element (TE) database (<http://maizetadb.org/~maize/>) using RMAP (Smith et al. 2008) with a maximum of four mismatches. Sequences that mapped to more than one position were removed to retain only reads that mapped uniquely. Duplicate reads (PCR duplicates) starting and ending at the same positions were removed from each data set.

Single-base methylation levels were calculated as the number of cytosines (C) called divided by the sum of Cs and thymines (T) called at each cytosine site in the target region. For context-specific methylation levels, only the cytosines in that context were used in calculations. The bisulfite conversion rate was calculated using reads that mapped to the chloroplast genome (which were present at a very low level due to DNA preparation protocol). Since the chloroplast genome is not methylated, cytosines mapping to the chloroplast genome were measured for conversion into thymines. High conversion rates were regularly achieved (98%–99%) and considered maximal, given that a small proportion of the chloroplast genome is included in the nuclear genome and is methylated. Overall levels of methylation are highly comparable to the results of HPLC analysis (Montero et al. 1992).

B73 reads covered 89 million of Cs in CG, 81 million in CHG, and 306 million in CHH context, which represented 94.25%, 94.61%, and 93.55% of Cs in each context, respectively, that could be mapped with 50-bp paired-end reads. There were 77, 70, and 257 million Cs in each respective context that were covered by at least three reads, representing 89.58%, 89.98%, and 86.22% of the mappable Cs.

Mo17 reads covered 84 million of Cs in CG, 76 million in CHG, and 282 million of Cs in CHH context, representing 82.04%, 82.84%, and 82.50% of Cs in each context, respectively, that could be mapped with 50-bp reads. There were 61, 56, and 200 million Cs in each respective context that were covered by at least three reads, representing 71.21%, 72.08%, and 66.96% of the mappable Cs.

The overlap in sequenced Cs between two inbreds was 72, 66, and 244 million Cs in the CG, CHG, and CHH context, respectively, ~80% and 85% of the mapped Cs for B73 and Mo17, respectively, in each context.

### RNA-seq data analysis

Sequenced mRNA reads were processed by using TopHat (Trapnell et al. 2009)/Cufflinks (Trapnell et al. 2010) with default parameters. The maize reference genome v1 was used as the reference genome and the working gene set (gene build 4a.53; [maizesequence.org](http://maizesequence.org)) was used as the gene model with no novel junction.

Methylation profiles for “metagenes” were generated by combining the profiles in three regions: 2 kb upstream of transcription start sites (TSS), gene body, and 2 kb downstream from transcription termination sites (TTS). Gene body methylation consisted of concatenated exons only. Maize introns were excluded to avoid confusion with methylated transposons, which are frequently inserted in introns. For each region, sequences from all genes in ZmB73 v1 (4a.53) filtered gene set were divided into 100 bins, then the average methylation of each bin across each region was calculated and plotted.

For the analysis of gene expression levels against methylation in genes, the expression level was based on the FPKM value for each gene (ZmB73 v1 filtered gene set) derived from cufflinks. Genes were divided into five groups according to FPKM values: 1 is  $\text{FPKM} > 100$ , 2 is  $100 \geq \text{FPKM} > 10$ , 3 is  $10 \geq \text{FPKM} > 1$ , 4 is  $1 \geq \text{FPKM} > 0$ , 5 is  $\text{FPKM} = 0$ . Methylation profiles for each group of genes was done as described above.

### Splicing data analysis

All splice junctions in the 4a.53 filtered gene set were stacked (50 bp exon + 50 bp intron for donor, 50 bp intron + 50 bp exon for acceptor). Then the methylation level of each base pair was calculated as  $C/(C + T)$ .

The methylation of the donor site was calculated from the first nucleotide of both strands on the 5' end of the intron as the ratio of  $C/(C + T)$ , and the methylation of the acceptor site was calculated from the last nucleotide of both strands on the 3' end of the intron using the same formula. We categorized  $\geq 0.8$  as high methylation and  $\leq 0.2$  as low methylation.

To find splicing sites that were skipped in alternatively spliced genes, RNA reads covering at least 10 bp of exon or intron across the splice junctions were identified. To find used splice sites, RNA reads covering at least 10 bp of both exons flanking the intron were identified. Only reads in correct strand and direction regarding transcription orientation of each gene were kept. We did  $\chi^2$  tests on the contingency table of unspliced read counts and spliced read counts in high and low methylation. The  $P$ -value of the  $\chi^2$  test shows the significance of the hypothesis that the proportion of unspliced read count increase is associated with methylation state.

### Calculation of methylation in transposable elements

Reads mapped to exemplar TEs were normalized as follows:  $\text{TE} = (10^9 * C)/(N * L)$ , where  $C$  is the number of reads mapped to the TE exemplar sequence,  $N$  is the total number of mapped reads, and  $L$  is the length of TE in base pairs (adapted from Mortazavi et al. 2008). Methylation in TEs was estimated using methylation data calculated as described above.

The TE copy number was estimated from the number of reads matching each exemplar.

For testing correlation between TE expression and methylation, RNA-seq reads were mapped to TE exemplar sequences using Bowtie with a maximum of two mismatches. Sequences mapped to more than one position were removed. The calculations of normalized read count and methylation were described above.

### Small RNA data analysis

The smRNA reads were mapped to maize reference genome v1 and the transposable element exemplar database using Bowtie (Langmead et al. 2009) with a maximum of two mismatches. Sequences that mapped to more than one position were removed.

To estimate correlations between methylation in TEs and small RNA abundance, smRNA reads were plotted for each TE, and

the normalized read count was calculated as above and plotted against methylation data.

For plotting genome-wide distributions of small RNA read counts against methylation, small RNA reads were mapped to the maize B73 v1 ref genome, and the cytosines were grouped according to small RNA read depth as follows: all cytosines covered by only one smRNA read as one group, all cytosines covered by two smRNA reads as another group, and so on. The average methylation was calculated for each group as  $(C)/(C + T)$ . In Figure 2, each point represents a group of cytosines with the smRNA read depth of the group ( $x$ -axis) plotted in respect to the average methylation of the group ( $y$ -axis).

### Calculation of methylation in centromeres

To estimate methylation levels in centromeric regions, bisulfite-treated reads were mapped against maize *CRM* retrotransposons (from the exemplar database) and maize CentC1 repeats (GenBank; AY530216.1) using RMAP, and methylation was calculated as above.

### Analysis of *Mutator* insertion sites

To test the effect of exon methylation on *Mu* insertion preference, the midpoint coordinate of each target site duplication for all uniquely mapped *Mu* insertions in all.Mu.2chr.uniqIns.#49125E.txt were used (Liu et al. 2009). Insertion sites outside genes annotated in the ZmB73 4a.53 filtered gene set were removed. The target site duplication midpoint was extended 50 bp both upstream and downstream. Then, the average cytosine methylation levels in CG/CHG/CHH contexts were calculated for those regions based on B73 and Mo17 whole-genome methylation data.

### Highly differentially methylated regions (HDMRs)

Highly differentially methylated regions were identified in the CG and CHG contexts by computing a  $\chi^2$  statistic over moving windows with 15 cytosine sites from either strand in both B73 and Mo17. Cytosine (C) and thymine (T) counts at cytosine sites in the reference genome were used. Additionally, sites determined to be possible C-to-T mutations were removed from analysis. These were determined by looking at the sequence calls on the strand opposite to the cytosine site. If more than half the calls in either B73 or Mo17 were adenine (A) rather than guanine (G), the site was deemed to be a possible mutation site. Initial regions were found by grouping adjacent windows with a  $\chi^2$  value over 100. Regions with at least 20 cytosine sites, at least 200 bases long, where one sample had a methylation level below 10% and the other above 70%, are referred to as highly differentially methylated regions (HDMRs).

### Sequence genotyping RILs

RILs were genotyped at 1335 sequence markers. These were mapped to the zmb73v1 genome by using the best BLAST match. Genotypes were one of four values: B73, Mo17, heterozygous, or no call.

### Scoring RILs at HDMRs

RILs were assigned similarity scores to B73 or Mo17 at each of the HDMRs. For each HDMMR, the percent methylation level is computed by counting the number of Cs and Ts on reads in each sample at cytosine sites within each HDMMR and then computing

the percent of Cs in these bases. The similarity score is  $(\%C \text{ in RIL} - \%C \text{ in B73})/(\%C \text{ in Mo17} - \%C \text{ in B73})$ . This will give a score near 0.0 when the methylation level in the RIL is close to B73 and near 1.0 when similar to Mo17.

### RILs switching methylation state

The similarity scores were segmented into contiguous regions matching one or the other parental line using a segmentation algorithm based on the Kolmogorov Smirnov (KS) test statistic (Grubor et al. 2009). Segmentation values less than 0.4 and greater than 0.6 were rounded to zero and one, respectively. Segments with values between 0.4 and 0.6 are considered regions of either heterozygous or mixed methylation. RILs with a similarity score more than 0.80 away from the methylation segment were considered as switching to the state of the other parental line. HDMRs where more than one RIL switched methylation state will be referred to as recurrent switch HDMRs. To test if there were more recurrent switch HDMRs than would be expected if these switches were occurring randomly, a bootstrap test of significance (Efron 1979) was used to compute a  $P$ -value for the number of recurrent switch HDMRs observed. This was done by resampling with replacement the methylation levels of the HDMRs, recomputing the similarity scores, and counting the number of recurrent HDMRs repeated 1000 times. The resampling distribution was not significantly different from a normal distribution, assessed by a KS test. Then a  $t$ -test with equal variance was used to compute a  $P$ -value for the observed recurrent count compared with the bootstrap samples. All statistical tests were computed using R version 2.13.1 (R Development Core Team 2011).

### False-positive HDMRs

We estimate an upper bound on false-positive detection of large methylation state changes in the RILs by analyzing regions where both B73 and Mo17 are either highly methylated or hypomethylated and the RIL is in the opposite state. The median size of the CG HDMRs is 508 bases. We split the maize genome into non-overlapping 500-base regions and count for each RIL how many of these regions are above 60% methylated in B73 and Mo17 or below 10% methylated in both and also have more than 60 C/T base calls in the RIL. These are regions that might potentially have detectable false-positive methylation state changes similar to what is observed in the HDMRs. Regions of significant difference were found using a  $\chi^2$  test and tested for parents above 60% methylated and RIL below 10% and reverse. In the CG context, there are between 289,000 and 1.03 million windows that meet these criteria in the nine RILs, and between 67 and 453 regions where the RIL is differentially methylated compared with the parent lines, giving estimated false-positive rates between 0.000232 and 0.000485. In the CHG context, there are between 208,000 and 1.01 million windows that meet these criteria in the nine RILs and between 47 and 492 regions where the RIL is differentially methylated compared with the parent lines, giving estimated false-positive rates between 0.000226 and 0.000474.

### Hypomethylated regions in B73 and Mo17

We applied a hidden Markov model (Durbin et al. 2007) to divide the genome into contiguous segments of hypermethylation and hypomethylation. The model includes a two-parameter transition matrix, measuring the frequency of transitioning from a hypomethylated to a hypermethylated state set at 0.0001 and from a hypermethylated to a hypomethylated state set at 0.0001 for both the CG and CHG contexts, two probability parameters  $p_U$

and pM, corresponding to the expected proportion of unconverted cytosine observed in a hypomethylated or hypermethylated region, respectively. The parameter pU is initialized as the mean proportion of Cs in B73 at C sites with an observed proportion <50%, and pM is initialized as the mean proportion of Cs in B73 at C sites with an observed proportion at least 50%. We then compute the posterior probabilities and determine the maximum likelihood parameters to update the transition and emission parameters and iterate until the emission parameters converge to within a tolerance of 0.001. For regions containing at least 10 cytosine sites, there is a total of 116 Mb and 109 Mb of hypomethylated DNA, with a median region size of 823 and 806 bases, in the CG context in B73 and Mo17, respectively, and 182 Mb and 186 Mb total hypomethylated DNA, with a median region size of 1118 bases and 1029 bases, in the CHG context.

## Data access

The data from this study have been deposited in the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) and are accessible through GEO Series accession number GSE39232.

## Acknowledgments

We thank Emily Hodges and Asya Stepansky for help with bisulfite sequencing protocols, Andrew Olson for help with expression profiling analyses, and Elena Ghiban and other members of the Genome Center for help with Illumina sequencing. We also thank Shiran Pasternak, Peter Bommert, and Pavel Regulski for help with graphic design and figure development. This work was supported by the Cold Spring Harbor Laboratory/Pioneer-Dupont Research Collaboration and the Agricultural Research Service of the United States Department of Agriculture. R.A.M. is a Howard Hughes Medical Institute–Gordon and Betty Moore Foundation Investigator in Plant Biology.

## References

- Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Sikkink K, Chandler VL. 2006. An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**: 295–298.
- Arteaga-Vazquez MA, Chandler VL. 2010. Paramutation in maize: RNA mediated *trans*-generational gene silencing. *Curr Opin Genet Dev* **20**: 156–163.
- Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, Sanmiguel PJ, Bennetzen JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet* **5**: e1000732.
- Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K, Weigel D. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* **480**: 245–249.
- Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rholing T, Fries J, Bradford K, et al. 2005. Sorghum genome sequencing by methylation filtration. *PLoS Biol* **3**: e13.
- Bender J. 2004. DNA methylation of the endogenous PAI genes in *Arabidopsis*. *Cold Spring Harb Symp Quant Biol* **69**: 145–153.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- Durbin R, Eddy SR, Krogh A, Mitchinson G. 2007. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press, Cambridge, UK.
- Effron B. 1979. Bootstrap methods: Another look at the jackknife. *Ann Stat* **7**: 1–26.
- Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh CT, Jia Y, Gendler K, Freeling M, et al. 2011. Heritable epigenetic variation among maize inbreds. *PLoS Genet* **7**: e1002372.
- Erhard KF Jr, Hollick JB. 2011. Paramutation: A process for acquiring *trans*-generational regulatory states. *Curr Opin Plant Biol* **14**: 210–216.
- Erhard KF Jr, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB. 2009. RNA polymerase IV functions in paramutation in *Zea mays*. *Science* **323**: 1201–1205.
- Gen T, Dong Y, Jiang J, Dawe RK. 2012. Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res* **40**: 1550–1560.
- Gottlieb TM, Wade MJ, Rutherford SL. 2002. Potential genetic variance and the domestication of maize. *Bioessays* **24**: 685–689.
- Grubor V, Krasnitz A, Troge JE, Meth JL, Lakshmi B, Kendall JT, Yamrom B, Alex G, Pai D, Navin N, et al. 2009. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* **113**: 1294–1303.
- Hale CJ, Stonaker JL, Gross SM, Hollick JB. 2007. A novel Snf2 protein maintains *trans*-generational regulatory states established by paramutation in maize. *PLoS Biol* **5**: e275.
- Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, et al. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol* **28**: 1097–1105.
- Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, et al. 2009. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* **19**: 1593–1605.
- Hollick JB, Chandler VL. 1998. Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. *Genetics* **150**: 891–897.
- Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci* **108**: 2322–2327.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220.
- Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536.
- Liu S, Yeh CT, Ji T, Ying K, Wu H, Tang HM, Fu Y, Nettleton D, Schnable PS. 2009.  $\mu$  transposon insertion sites and meiotic recombination events colocalize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet* **5**: e1000733.
- Martienssen RA, Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**: 1070–1074.
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325**: 737–740.
- Montero LM, Filipiński J, Gil P, Capel J, Martínez-Zapater JM, Salinas J. 1992. The distribution of 5-methylcytosine in the nuclear genome of plants. *Nucleic Acids Res* **20**: 3207–3210.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* **5**: 621–628.
- Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, Dike S, de la Bastide M, Martienssen RA, McCombie WR. 2003. Maize genome sequencing by methylation filtration. *Science* **302**: 2115–2117.
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reddy AS. 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* **58**: 267–294.
- Richards EJ. 2011. Natural epigenetic variation in plant species: A view from the field. *Curr Opin Plant Biol* **14**: 204–209.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
- Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR. 2011. Transgenerational epigenetic instability is a source of novel methylation variants. *Science* **334**: 369–373.
- Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, et al. 2013. Patterns of population epigenomic diversity. *Nature* **495**: 193–198.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. *Science* **326**: 1112–1115.
- Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, et al. 2002. Development and mapping of SSR markers for maize. *Plant Mol Biol* **48**: 463–481.

- Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* **34**: 3955–3967.
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74–79.
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–285.
- Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–472.
- Smith AD, Xuan Z, Zhang MQ. 2008. Using quality scores and longer reads improves accuracy of Solexa read mapping. *BMC Bioinformatics* **9**: 128.
- Tanurdzic M, Vaughn MW, Jiang H, Lee TJ, Slotkin RK, Sosinski B, Thompson WE, Doerge RW, Martienssen RA. 2008. Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol* **6**: 2880–2895.
- Teixeira FK, Heredia F, Sarazin A, Roudier F, Boccarda M, Ciaudo C, Cruaud C, Poulain J, Berdasco M, Fraga MF, et al. 2009. A role for RNAi in the selective correction of DNA methylation defects. *Science* **323**: 1600–1604.
- Topp CN, Zhong CX, Dawe RK. 2004. Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci* **101**: 15986–15991.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* **25**: 1105–1111.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.
- Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, et al. 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol* **5**: e174.
- Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW. 2009. Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell* **21**: 1053–1069.
- Wei F, Stein JC, Liang C, Zhang J, Fulton RS, Baucom RS, De Paoli E, Zhou S, Yang L, Han Y, et al. 2009. Detailed analysis of a contiguous 22-Mb region of the maize genome. *PLoS Genet* **5**: e1000728.
- Yang L, Takuno S, Waters ER, Gaut BS. 2011. Lowly expressed genes in *Arabidopsis thaliana* bear the signature of possible pseudogenization by promoter degradation. *Mol Biol Evol* **28**: 1193–1203.

Received December 11, 2012; accepted in revised form May 30, 2013.