# Identification and Characterization of DNA Methylation Variation within Maize

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Steven Richard Eichten

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Nathan M. Springer

May 2013

**Acknowledgements**

This work would not exist if not for the help of many people who have guided me during my graduate career. I would like to thank Dr. Nathan Springer for his guidance, support, insight and friendship. Working with Dr. Springer from my undergraduate summer internship through my graduate career has been a fantastic and enjoyable experience. His guidance has helped me become the scientist I am today. Others in the Springer lab, Dr. Ruth Swanson-Wagner, Amanda Waters and Pete Hermanson have provide excellent feedback as well as support in the development of the work presented here. I also thank my committee members for their guidance and assistance during my graduate career; Dr. Peter Tiffin, Dr. Gary Muehlbauer, Dr. Robert Stupar, and Dr. Jane Glazebrook.

I would also like to thank my mother, Sharon Eichten, as well as Alison Heussler and all of my family and friends for providing me support and encouragement to continue my education.

**ABSTRACT**

DNA methylation is a genetic modification known to repress the activity of transposable elements, repetitive sequences, and in some cases genes. Although DNA methylation is often found in common locations across different individuals, evidence has shown that DNA methylation can vary between individuals at certain loci and can therefore have the opportunity to create a unique regulatory environment for the surrounding sequence. Beyond this, the relationship between DNA methylation state and the genetic content of an individual is still unclear. DNA methylation may act as a downstream effect of certain genetic signals, or it may act independently of genetic state as an epigenetic modification. The goal of this thesis is to profile the DNA methylation landscape across maize (*Zea mays*) and identify the genomic regions that display differential DNA methylation patterns. These regions of differential methylation are then further studied to understand their stability across generations, their influences on gene expression, as well as their connection to the genetic context they are found. The chapters describe the identification of thousands of differentially methylated regions (DMRs) between maize lines. These DMRs are shown to occur throughout the genome and have high stability across generations. In contrast, few DMRs are found across different tissues within the same genotype. DMRs are shown to often be associated with the local genetic varitaion. This genetic relationship is highlighted, along with the discovery of a mechanism of genetic control by the spreading of DNA methylation from certain retrotransposable elements. These results indicate that DMRs are present in maize and are created through both epigenetic and genetic means.

# Table of Contents

# List of Tables

**Chapter II – Heritable epigenetic variation among maize inbreds**



**Chapter III – Variation in DNA methylation patterns is more common among maize inbreds than among tissues**



**Chapter IV – Spreading of heterochromatin is limited to specific families of maize retrotransposons**



**Chapter V – Epigenetic and genetic influences on DNA methylation variation in maize populations**

# List of Figures

**Chapter I – Literature review of DNA methylation**



**Chapter II – Heritable epigenetic variation among maize inbreds**



**Chapter III – Variation in DNA methylation patterns is more common among maize inbreds than among tissues**

**Chapter IV – Spreading of heterochromatin is limited to specific families of maize retrotransposons**



**Chapter V – Epigenetic and genetic influences on DNA methylation variation in maize populations**

# Chapter 1
# Literature review of DNA methylation

## SOURCES OF HERITABLE INFORMATION

Heritable information is commonly assumed to be contained within the DNA sequence of an individual's genome. Variation in DNA sequence between individuals, as well as populations, can lead to heritable phenotypic variation. This genetic variation often results in allelic differences between individuals of a species in the coding sequence of genes or in regulatory regions. In addition to changes in DNA sequence, structural variation such as copy number variation (CNV) and presence / absence variation (PAV) of sequences, including genes, can contribute to heritable variation (Springer et al., 2009; Swanson-Wagner et al., 2010). The assessment of genetic variation, most commonly in the form of single nucleotide polymorphisms, has been the basis for most genome-scale methods for identifying the genomic regions that contribute to heritable variation, such as QTL or GWAS analyses (Brachi et al., 2011). Although these approaches have been used successfully to identify heritable genetic variation there is also evidence for a substantial amount of "missing heritability" that is not captured by these SNP-based approaches (Maher 2008). It has been hypothesized that epigenetic variation, which may not be captured by SNPs, may contribute to this missing heritability and may be important for more accurate predictions of phenotype (Slatkin, 2009). Epigenetic variation is heritable variation that is not directly connected to DNA sequence polymorphisms (Bird 2007) that might act in tandem with genetic and environmental factors to contribute to the resulting phenotype of an individual.

The term "epigenetic" is often used to describe a variety of chromatin modifications such as histone tail modifications and DNA methylation (Jablonka & Raz, 2009). Although these chemical modifications certainly can play a role in epigenetic information, they are not necessarily epigenetic. These chromatin modifications may be the result of genetic sequences that "program" the marks or they might be transient modifications with no heritability. Our ability to profile the genome-wide distribution of these chromatin modifications has improved and it is important to understand how these genome-wide distributions of chromatin modifications inform our understanding of genetic and epigenetic variation.

The ability to assess epigenetic information, in the form of heterochromatin modifications, has been limited and its influence on phenotype has not been explored in detail. In addition, the potential instability of epigenetic marks could result in unstable behaviors for traits that are less likely to be observed for traditional genetic variation. Although the theoretical impacts of DNA methylation have been explored (Richards, 2006; Johannes et al., 2008; Weigel & Colot, 2012), there is relatively little known about the prevalence of epigenetic natural variation in plant species.

Any 'useful' epigenetic information would be expected to impact the phenotype of the individual(s) displaying the variation. The most likely method of action would be the repression of gene expression by methylation of the gene promoter or possibly increased transcription through the methylation of gene bodies (Zhang et al., 2006; Vaughn et al., 2007; Zilberman et al., 2007; Cokus et al., 2008; Lister et al., 2008). Studies in *Arabidopsis thaliana* have shown negative correlations between DNA methylation presence and mRNA abundance of genes (Zhang et al., 2008; Schmitz et al., 2012). Theoretical methods to integrate epigenetic information, such as DNA methylation, with traditional genetic assessments of traits have been outlined (Johannes et al, 2008) in which combining these factors may allow for more complete heritability estimates.

**ESTABLISHMENT AND MAINTNANCE OF DNA METHYLATION**

The methylation of cytosines (5-methylcytosine) in genomic DNA of some eukaryotes is one of the best characterized examples of chromatin modification with studies on the mechanisms of targeting and heritability as well as the functional impacts of this modification (Chan et al., 2005; Stroud et al., 2013). 5-methylcytosine is generated by the activity of DNA methyltransferases which add a methyl group to cytosine residues following DNA replication. The enzymatic addition of a methyl group to cytosines can be separated into two categories: *de novo* methylation of previously unmethylated DNA and maintenance methylation which would allow a methylated state to be maintained through DNA replication and subsequent cellular division (Chan et al., 2005). The establishment of DNA methylation is facilitated by the RNA directed DNA methylation pathway (RdDM; Law & Jacobsen, 2010) in which small RNAs guide the targeting of the DRM2 *de novo* methyltransferase to specific genomic locations. As long as the *de novo* methyltransferase is being actively targeted to a region by the RdDM machinery, the majority of cytosines in that region will be methylated and DNA methylation will be observed in

CG, CHG and CHH (H is any base that is not guanine) sites. In the absence of an active targeting of *de novo* DNA methylation it is still possible for DNA methylation to be maintained in certain sequence contexts. The maintenance of DNA methylation at symmetrical (across strands) CG and CHG sites is facilitated by MET1 and CMT3 in *Arabidopsis thaliana* with homologous genes found in maize (Steward et al., 2000; Papa et al., 2001; Makarevitch et al., 2007). CG methylation is found as a symmetric mark on both strands of DNA. It is maintained following DNA replication by MET1 as this enzyme is associated with the replication machinery and add a methyl-group to hemi-methylated sites on the symmetric site on the unmethylated strand (Law & Jacobsen, 2010). CHG methylation is also symmetric, and is maintained by the methyltransferase CMT3. There is evidence for positive feedback between H3K9me2 histone modifications and CHG methylation as the chromodomain of CMT3 recognizes this chromatin modification to reinforce silencing at these regions (Du et al., 2012). In contrast to CG and CHG sites, CHH methylation is an asymmetric context of DNA methylation. This assymetric CHH methylation cannot utilize maintenance methods that utilize hemi-methylated templates. CHH methylation therefore is not strictly maintained, but requires continually targeted *de novo* establishment through DMR2 and the RdDM pathway (Law & Jacobsen, 2010).

The loss of DNA methylation can occur through passive or active means. The maintenance of DNA methylation requires specific enzymatic activity. If this maintenance does not occur, DNA replication will quickly lead to hemi-methylated and subsequently non-methylated daughter strands of DNA. In this way, DNA methylation can be lost rapidly through entirely passive means. DNA methylation can also be actively removed by the activity of DNA glycosylases such as DEMETER (DME, Choi et al., 2002) and ROS1 (Gong et al., 2002) which excise the methylated cytosine and replace it with an unmethylated cytosine (Zhu et al., 2009). If active demethylation enzymes are repressed, DNA methylation is increased at specific genomic loci (Gong et al., 2002; Penterman et al., 2007; Lister et al., 2008). These studies provide evidence that specific genomic loci are targeted for active demethylation while much of the genome would lose DNA methylation primarily through passive mechanisms.

The mechanisms for methylating DNA are generally similar in flowering plants and mammals, however the specific targeting mechanisms and functional roles of DNA methylation may differ (Law & Jacobsen, 2010; Feng et al., 2010; Feng & Jacobsen, 2011). Two unique aspects of DNA methylation in plants indicate slight differences in DNA methylation processes in plants and

mammals. The first is that plants contain the CMT3 methyltransferase which catalyzes CHG methylation while mammals do not contain CMT3 homologs and generally lack CHG methylation(Chan et al., 2005; Lee et al., 2010). The second major change is that the RNA-directed DNA methylation pathway utilizes two plant-specific RNA polymerases, PolIV and PolV in the processing of small RNAs to allow active targeting of CHH methylation (Law & Jacobsen, 2010). This specific function may be partially fulfilled by PolII in mammals but there is evidence for specialized RNA polymerases for RdDM in plants. Outside of mammals, many DNA methylation pathways are conserved across other eukaryotes, however the level of conservation and divergence varies across the eucaryota domain (Zemach et al., 2010). Some model eukaryotes such as *C. elegans, Drosophila*, and certain yeasts appear to have no, or extremely low levels of DNA methylation (Gowher et al., 2000; Bird 2002) indicating DNA methylation is not a requirement for proper eukaryotic function.

**FUNCTIONAL CONSEQUENCES OF DNA METHYLATION**

DNA methylation can lead to stable phenotypic variation within plant and animal species (Rasmusson et al., 1997; Richards 2006; Feinberg 2007; Jirtle & Skinner 2007; Vaughn et al., 2007; Richards 2008; Johannes et al., 2008). The presence of DNA methylation is often considered to be a mark of heterochromatin that would result in lower levels of transcription. However, genome-wide profiles of DNA methylation and gene expression have suggested that the relationship of DNA methylation and functional state are not entirely repressive (Zilberman et al., 2007; Vaughn et al., 2007). Initial studies using tiling microarrays in *Arabidopsis* thaliana allowed for patterns of DNA methylation to be assessed (Lippman et al., 2004; Zilberman et al., 2007; Vaughn et al., 2007; Zhang et al., 2008). Transposable elements and repetitive DNA were universally targeted by DNA methylation, with a higher proportion of methylation observed in the centromeric regions of chromosomes. Some genes displayed internal gene body methylation which was often linked to moderately expressed genes (Zilberman et al., 2007). In contrast, DNA methylation found within the promoter regions of genes was often inversely correlated with gene expression (Zhang et al., 2008). In maize, genome wide profiling of DNA methylation by sequencing DNA digested with McrBC, a methylation-sensitive restriction enzyme, showed evidence for mutually exclusive patterns of DNA methylation and histone modification H3K27me3 near genes with low, or no expression (Wang et al., 2009). More recent studies using bisulfite sequencing technology, a method of developing single-base resolution of

methylated cytosines in all three methylation contexts (CG, CHG, and CHH; Lister & Ecker, 2009), have expanded this work highlighting unique regions in which CG, CHG, and CHH methylation are present (Cokus et al., 2008; Lister et al., 2008; Schmitz & Ecker, 2012). CG and CHG methylation are often found in similar locations whereas CHH methylation, as a non-symmetric site requiring active *de novo* methylation for its presence, is often rare and transient. Gene body methylation is predominantly in the CG context paired with a depletion of CHH methylation and is often associated with genes that show moderate to high levels of constitutive expression (Zilberman et al., 2007).

There is evidence that DNA methylation plays an important role in suppressing the activity of transposable elements, pseudogenes, and repetitive sequences (Chan et al., 2005; Slotkin et al., 2007). As active transposons can insert and disrupt genes, an organism's ability to limit their transposition can prevent unwanted genomic disruption (Yoder et al., 1997). Transposons often contain cryptic promoters that allow small RNAs to be produced if the transposon is not already suppressed. These small RNAs are used as targeting factors in the RdDM pathway leading to the silencing of transposable elements (Law & Jacobsen, 2010). Indeed when genes that are required for the RdDM pathway are inhibited, active transcription of transposable elements and transposition is greatly increased (Miura et al., 2001).

**GENOME-WIDE DISCOVERY OF DNA METHYLATION VARIATION WITHIN PLANTS**

The availability of tools for genome-wide profiling of DNA methylation levels have allowed researchers to identify DNA methylation variation among individuals or tissues in several plant species. Several different approaches have been used to study natural variation in DNA methylation and the role of DNA methylation in cryptic variation. The straight-forward approach involves comparing DNA methylation levels in different individuals of the same species. Initial studies in *Arabidopsis thaliana* used a microarray platform to study DNA methylation levels for a single chromosome (Vaughn et al., 2007). The DNA methylation patterns over transposable elements are highly conserved, however DNA methylation within gene bodies was found to be highly polymorphic across different ecotypes (Vaughn et al., 2007). A recent detailed study by Schmitz et al. (2013) examined the DNA methylation patterns across 152 wild accessions of *Arabidopsis thaliana* finding thousands of differentially methylated regions (DMRs) across the lines. These DMRs were found as both CG-specific DMRs that were enriched within gene bodies, and C-DMRs (all sequence contexts) which were enriched for

repetitive elements (Schmitz et al., 2013). A study in two cultivars of rice found differential methylation at over 7 percent of cytosines between the lines studied (Chodavarapu et al., 2012). Regions of variable methylation were observed and associated with variable production of small interfering RNAs (siRNAs) between the two parental lines as well as their hybrid offspring.

A second approach involves the analysis of spontaneous DNA methylation variation. Several Arabidopsis lines were generated for studying spontaneous mutation accumulation (Shaw et al., 2000). Genomic re-sequencing documented the rate of spontaneous DNA mutations in these lines after 30 generations (Ossowski et al., 2010). Whole-genome bisulfite sequencing revealed the rate of spontaneous epimutations in DNA methylation (Becker et al., 2011; Schmitz et al., 2011). DMRs identified between early and late generations were more often found away from transposable elements and often displayed instability in subsequent generations (Becker et al., 2011). While the frequency of DNA methylation changes at single bases was much higher than the rate of DNA mutations the rate of changes in larger DMRs was quite similar to DNA mutations (Becker et al., 2011; Schmitz et al., 2011).

The third approach has invesTrends in Geneticsated the phenotypic variation that appears following the loss of DNA methylation. Several epiRIL populations were developed by using mutations that affect epigenetic regulation (Reinders et al., 2009; Johannes et al., 2009). These populations are expected to segregate for altered levels of DNA methylation but not for sequence polymorphisms. The epiRILs populations developed show substantial phenotypic variation suggesting the potential for cryptic variation that is normally suppressed by DNA methylation. Many DMRs created in the epiRIL population displayed stable inheritance over multiple generations (Johannes et al., 2009). Non-parental methylation states were also observed leading to a unique epigenetic pattern at loci that were fixed between the parental lines (Reinders et al., 2009). The evidence for stable, heritable DNA methylation patterns across segregating individuals displaying phenotypic variation indicates a possible role of DNA methylation as heritable information.

**GENETIC AND EPIGENETIC CONTRIBUTION TO DNA METHYLATION VARIATION**

Understanding the sources of DNA methylation variation is important as DNA methylation can impact gene expression and phenotype. While DNA methylation is generally thought of as an epigenetic (not programmed by DNA sequence) mark it is also possible that the targeting of DNA

methylation to particular regions is influenced by DNA sequence variation. Many of the known loci showing DNA methylation variation (Martin et al., 2009; Durand et al., 2012; Luff et al., 1999; Melquist et al., 1999) are located near examples of copy number variation or transposable element insertions that could be causal factors determining the methylation state. Indeed the hypermethylation of retrotransposon elements has been shown to spread DNA methylation beyond the borders of the element itself (Hollister et al., 2011) which suggests that polymorphism for transposon insertions could lead to variation in DNA methylation levels in flanking sequences. Other methylation variants have been linked to copy number variation in gene paralogs located in other areas of the genome in which siRNAs may play a role in directing the methylation state at the observed loci (Durand et al., 2012; Luff et al., 1999; Melquist et al., 1999).

The potential for genetic sequence variation to cause variation in DNA methylation arguably limits its contribution as a truly epigenetic mark (Weigel & Colot, 2012). However, it is important to note that even if the DNA methylation is targeted due to genetic sequence changes it may still have influences on gene expression that are beyond the impacts of the genetic change. In other words, the presence of a transposon near a gene may result in silencing due to the presence of DNA methylation. If the DNA methylation was removed but the transposon remained, the gene would be reactivated. This complicates our uses of the term epigenetic as this example has both genetic and epigenetic influences on gene expression. Richards (2006) described three classes of potential "epialleles": obligatory, facilitated, and pure (Figure 1). Obligatory epialleles are completely dependent on genetic variation such that the genetic variation will define the resulting pattern of DNA methylation at a locus. Facilitated epiallelic variation occurs when a genetic change results in a poised state in which the chromatin state can exist as either active or silent in one genetic allele but the other genetic allele always exists in the active state. Pure epiallelic variation represents true epigenetic variation that is not connected to the underlying genetic state of an organism. Obligatory epialleles are likely to be highly stable and may often be predictable based on the analysis of linked genetic variation. In contrast, facilitated and pure epialleles may be unstable and are not likely to be predictable based on DNA sequence analysis (Figure 1).

**EVIDENCE FOR GENETICALLY CONTROLLED DNA METHYLATION VARIATION**

There is evidence that at least some of the DNA methylation variation observed in plants is the result of genetic changes or obligatory epialleles. Obligatory variation does not require epigenetic "memory" as the sequence-level variation is inherited through traditional genetic means. A major consequence is that the association between chromatin state and genotype could be observed through genotyping assays of SNPs and the chromatin state can be correctly imputed. An elegant example of this can be found in the sex determination through DNA methylation regulation of *CmWIP1* in melon (Martin et al., 2009). What was originally discovered as a DNA methylation variation in the promoter of CmWIP1 resulting in differential expression of CmWIP1 was shown to be caused by a transposon insertion near the promoter. When the transposon is removed, the CmWIP1 promoter is not methylated and the gene is transcribed (Martin et al., 2009). Perhaps the best characterized example of methylation variation due to a genetic polymorphism is the PAI loci in *Arabidopsis* (Bender & Fink, 1995; Melquist et al., 1999). In certain accessions the PAI1 locus has a duplicated inverted repeat which leads to the RdDM targeted methylation of the PAI1 locus, as well as targeting the PAI2 and PAI3 loci that are located elsewhere in the genome. When this inverted repeat at PAI1 is not present, or when PAI2 / PAI3 are moved into a background without the PAI1 inverted repeat, PAI2/3 methylation is lost (Bender & Fink, 1995).

A better understanding of the prevalence of obligatory epigenetic variation and the specific genetic changes that lead to this variation is important in understanding whether the majority of DNA methylation changes can be "captured" by analysis of genotype alone. If obligatory epigenetic variation is quite common and explains the majority of DNA methylation variation then it would be expected that DNA methylation variation would not contain previously unassessed heritable variation, i.e. missing heritability, unless the sequence variations assessed were not in linkage disequilibrium with the causal variant (e.g. a PAV transposon not resolved with SNP data across lines). It is important to understand the frequency of obligatory epialleles relative to pure epialleles and to understand how commonly these are caused by transposon insertion polymorphisms, CNV or other genetic changes. Studies in mice have identified *cis*-acting methylation-determining regions (MDRs) that are genetic elements that determine nearby methylation states across mouse stem cells (Lienert et al., 2011). In plants, genes displaying allele-specific expression patterns in rice hybrids display a high level of variable

methylation states that also correlate to an increase in SNP rate between lines (Chodavarapu et al., 2012) indicating a possible connection between genetic variation and methylation state. Recent work mapping DMRs to controlling genetic regions identified thousands of 'methylQTL', regions of the genome that impact DMR state, that associate genetic information to methylation variation across a population of wild *Arabidopsis thaliana* accessions (Schmitz et al., 2013). The genetic regions appeared to be impacting the methylation state at one third of the ~3000 DMRs containing methylation differences in all three methylation contexts.

The facilitated epialleles are dependent upon a genetic change but are not strictly associated with this genetic difference (Richards, 2006; Figure 1). This class may be difficult to identify as they must be identified by population level analyses with multiple alleles including those without the genetic change (which will always exhibit one chromatin state) and those with the genetic changes (which will include exhibit both chromatin states). The reversible nature of the chromatin state at the facilitated epialleles would prevent the ability to predict the chromatin state from genotype information. If controlling sites for facilitated chromatin variation were identified, it would be possible to predict that some alleles would have an active chromatin state but other alleles (with the genetic change) could not be accurately predicted as active or inactive. Due to the difficulty in identifying facilitative chromatin variants, their impact on heritable phenotype is largely unknown.

**EVIDENCE FOR PURE EPIGENETIC VARIATION FOR DNA METHYLATION**

While there is evidence that some DNA methylation variation is driven by DNA sequence changes there is also evidence that some examples of DNA methylation variation are the result of pure epigenetic variation. These pure epialleles completely sever the link between underlying genetic content and DNA methylation state. If these pure epialleles are stable through meiotic inheritance, they may act as a source of missing heritable information that cannot be assessed through any traditional genetic association studies. This true epigenetic variation may be important for the understanding of all aspects of heritable variation of key agronomic traits. The *Lcyc* locus in *Linaria vulgaris*, as well as the *CNR* locus in tomato do not display any local sequence variation within 10kb of the gene when comparing lines displaying altered methylation states (Cubas et al., 1999; Manning et al., 2006). The SUPERMAN locus in *Arabidopsis thaliana* impacts flower development and heritable epialleles deemed 'clark kent alleles' display aberrant methylation and suppression compared to the wild type allele even

though there are no sequence polymorphisms between the two alleles (Jacobsen & Meyerowitz, 1997). In their 'methylQTL' mapping, Schmitz et al (2013) were able to identify *trans*-acting genetic factors that impacted methylation state for some loci. These remote genetic states could influence a DMR state without local genetic variation present at the DMR locus.

**STABILITY OF DNA METHYLATION AS A HERITABLE MARK**

A major difference between genetic variation and DNA methylation variation is that while genetic variation is fixed as a highly stable variant, DNA methylation may exhibit much lower levels of stability and heritability with common reversion for states of methylated / unmethylated cytosines through both active and passive mechanisms (Law and Jacobsen, 2010). The heritability of DNA methylation across generations has been assessed in a variety of ways. In rice an allele-specific methylation assay was performed on F1 hybrids from the NPB and 93-11 varieties. The parental methylation rates found in F1 were shown to be faithfully inherited almost 100% of the time when both parents displayed similar methylation states, whereas regions displaying differential methylation between the parental lines were stable ~93% of the time (Chodavarapu et al, 2012).

DMRs that are stable across multiple generations may provide an epigenetic method of gene regulation that could be selected for breeding purposes. However, DNA methylation that is controlled by genetic factors will most likely display higher stability due to the causal genetic elements being more stable themselves. Vaughn et al (2007) showed that transposable element methylation was much more stable across various genotypes and generations whereas genic methylation often displayed a higher rate of instability. DNA methylation that appears to be purely epigenetic has been estimated to have an 'epimutation' rate, the rate of DNA methylation state change per generation, greater than five-times higher than genetic variation over the same timespan (Schmitz et al., 2011; Becker et al., 2011).

**CONCLUSIONS**

There is increased interest in epigenetic phenomenon within plant systems as a possible source for previously unassessed heritable variation (Slotkin, 2009). A detailed view of chromatin marks such as DNA methylation that are inherited epigenetically as true 'epialleles' detached from sequence variation are vital in properly understanding the phenotypic impact they produce within a species.

Although experiments detailing DNA methylation variation, stability, and interaction with genetic elements have been conducted in the dicot *Arabidopsis thaliana*, there is limited work on understanding the epigenome of crop plants. Maize is a highly diverse inbreeding species (Buckler et al., 2006; Messing & Dooner 2006) that provides a useful model to study the role of epigenetic variation. The maize genome is highly complex with a large proportion of genomic sequence that is repetitive or derived from transposable elements (Rabinowicz & Bennetzen 2006; Schnable et al., 2009). As transposable elements are often the targets of DNA methylation, the overall patterns of DNA methylation across maize chromosomes leads to many low-copy genes being directly flanked by transposable elements, leading to an environment prone for epigenetic variation to occur. In addition, a number of epigenetic phenomena including imprinting (Zhang et al., 2011; Waters et al., 2011), paramutation (Chandler et al., 2000) and transposable element inactivation (Chomet et al., 1987) were initially described in maize.

This work outlines four major experiments performed to determine the level of DNA methylation variation across maize, its stability as a heritable mark, as well as its connection to the genotype across a diverse panel of maize inbreds. Profiling DNA methylation on a genome-wide scale and integrating this information with known genetic and transcriptional information is used to expand our view of the impacts DNA methylation has within this important crop.

**REFERENCES**


Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245–249.

Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83: 725–734.

Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398.

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes & Development 16: 6–21.

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology 12: 232.

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Current Opinion in Plant Biology 9: 172–176.

Chan SW-L, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Reviews Genetics 6: 351–360.

Chandler VL, Eggleston WB, Dorweiler JE (2000) Paramutation in maize. Plant Molecular Biology 43: 121–145.

Chodavarapu RK, Feng S, Ding B, Simon S a, Lopez D, et al. (2012) Transcriptome and methylome interactions in rice hybrids. Proceedings of the National Academy of Sciences of the United States of America 109: 12040–12045.

Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell 110: 33–42.

Chomet PS, Wessler S, Dellaporta SL (1987) Inactivation of the maize transposable element Activator (Ac) is associated with its DNA modification. The EMBO Journal 6: 295–302.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.

Cubas P, Vincent C, Coen E, Centre JI, Lane C, et al. (1999) An epigenetic mutation responsible for natural variation in ¯ oral symmetry: 157–161.

Du J, Zhong X, Bernatavichute Y V, Stroud H, Feng S, et al. (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151: 167–180.

Durand S, Bouché N, Perez Strand E, Loudet O, Camilleri C (2012) Rapid establishment of genetic incompatibility through natural epigenetic variation. Current biology?: CB 22: 326–331.

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447: 433–440.

Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. Current Opinion in Plant Biology 14: 179–186.

Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. Science 330: 622–627.

Gong Z, Morales-Ruiz T, Ariza RR, Roldán-Arjona T, David L, et al. (2002) ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111: 803–814.

Gowher H, Leismann O, Jeltsch A (2000) DNA of Drosophila melanogaster contains 5-methylcytosine. The EMBO Journal 19: 6918–6923.

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, et al. (2011) Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proceedings of the National Academy of Sciences of the United States of America 108: 2322–2327.

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. The Quarterly Review of Biology 84: 131–176.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. Science 277: 1100–1103.

Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. Nature Reviews Genetics 8: 253–262.

Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics 9: 883–890.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics 5: e1000530.

Law J a, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11: 204–220.

Lee T-F, Zhai J, Meyers BC (2010) Conservation and divergence in eukaryotic DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 107: 9027–9028.

Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, et al. (2011) Identification of genetic elements that autonomously determine DNA methylation states. Nature Genetics 43: 1091–1097.

Lippman Z, Gendrel A V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.

Lister R, Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Research 19: 959–966.

Luff B, Pawlowski L, Bender J (1999) An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. Molecular Cell 3: 505–511.

Maher B (2008) The case of the missing heritability. Nature 456: 18–21.

Makarevitch I, Stupar RM, Iniguez AL, Haun WJ, Barbazuk WB, et al. (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase zmet2. Genetics 177: 749–760.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948–952.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135–1138.

Melquist S, Luff B, Bender J (1999) Arabidopsis PAI gene arrangements, cytosine methylation and expression. Genetics 153: 401–413.

Messing J, Dooner HK (2006) Organization and variability of the maize genome. Current Opinion in Plant Biology 9: 157–163.

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, et al. (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature 411: 212–214.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, et al. (2010) The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science 327: 92–94.

Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM (2001) Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. The Plant Cell 13: 1919–1928.

Penterman J, Zilberman D (2007) DNA demethylation in the Arabidopsis genome. Proceedings of the National Academy of Sciences of the United States of America 104: 6752–6757.

Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology 9: 149–156.

Rasmusson DC, Phillips RL (1997) Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Science 37: 303–310.

Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, et al. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939–950.

Richards EJ (2008) Population epigenetics. Current Opinion in Genetics & Development 18: 221–226.

Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews Genetics 7: 395–401.

Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends in Plant Science 17: 149–154.

Schmitz RJ, Schultz MD, Urich M a., Nery JR, Pelizzola M, et al. (2013) Patterns of population epigenomic diversity. Nature 495: 193-198.

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich M a, et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science 334: 369–373.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Shaw RG, Byers DL, Darmo E (2000) Spontaneous mutational effects on reproductive traits of arabidopsis thaliana. Genetics 155: 369–378.

Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. Genetics 182: 845–850.

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272–285.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, et al. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics 5: e1000734.

Steward N, Kusano T, Sano H (2000) Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic Acids Research 28: 3250–3259.

Stroud H, Greenberg MVC, Feng S, Bernatavichute Y V, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152: 352–364.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research 20: 1689–1699.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology 5: e174.

Wang X, Elling AA, Li X, Li N, Peng Z, et al. (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. The Plant Cell 21: 1053–1069.

Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh CT, et al. (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. The Plant Cell 23: 4221–4233.

Weigel D, Colot V (2012) Epialleles in plant evolution. Genome Biology 13: 249.

Yoder J a, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics13: 335–340.

Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919.

Zhang M, Zhao H, Xie S, Chen J, Xu Y, et al. (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proceedings of the National Academy of Sciences of the United States of America 108: 20042–20047.

Zhang X, Shiu SH, Cal A, Borevitz JO (2008) Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. PLoS Genetics 4: e1000032.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126: 1189–1201.

Zhu J-K (2009) Active DNA demethylation mediated by DNA glycosylases. Annual Review of Genetics 43: 143–166.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39: 61–69.

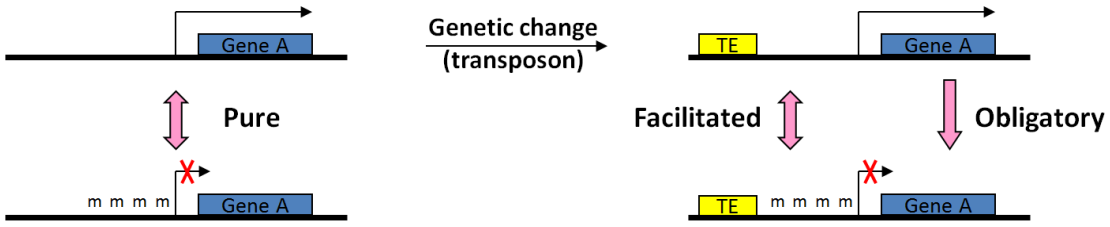



Figure 1: Diagram of DNA methylation's dependence and independence related to genetic state. A pure epiallele (shown left) indicates a change in methylation state (m) that is independent of genetic variation. DNA methylation variation occurring after a genetic change (transposon insertion, yellow box) allows for facilitated or obligatory methylation variation.

Chapter II - **CONTEXT STATEMENT**

Although DNA methylation variation had been studied at specific loci in maize, no genome-scale analysis of variable DNA methylation had been performed prior to this study. The work described in chapter II documented hundreds of differentially methylated regions between the genomes of maize inbred lines B73 and Mo17. Additionally, this chapter highlights the ability to define regions of variable methylation that do not show a clear connection to genetic variation. The work in this chapter developed the set of methods to analyze microarray data that are utilized in subsequent chapters.

Chapter II entitled 'Heritable epigenetic variation among maize inbreds' has been adapted from my work in the publication:

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, S Liu, C T Yeh,Y Jia, Gendler, K, Freeling, M, Schnable, P S, Vaughn, M W, Springer, N M . (2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics 7: e1002372. doi:10.1371/journal.pgen.1002372.

In the development of this manuscript, other researchers have played roles in the creation of data and in the development of analytical methods. RSW and PJH performed the microarray experiments to obtain the raw experimental data. MVand KG provided support in developing methods for normalizing the microarray data. I performed the analysis of the data to develop biological conclusions. Additional computational support and validations based upon existing methyl-filtration genome sequences was provided by LS,YC, and JY and PS. qPCR validations of DMR regions were conducted by SRE and AJW. Figure development and text of this document was developed by NS and SRE.

Several portions of the text as well as supplemental data tables that were present in the published manuscript have been omitted from this chapter in order to focus only on the portions of the work that I played an active role in generating or analyzing. In addition, I have removed the author list and contacts, publication abstract, author summary, and acknowledgments. Reference formatting has been modified to a standard format throughout the thesis. All omitted data is available through the initial scientific publication.

# Chapter II
# Heritable epigenetic variation among maize inbreds

Epigenetic variation describes heritable differences that are not attributable to changes in DNA sequence. There is the potential for pure epigenetic variation that occurs in the absence of any genetic change or for more complex situations that involve both genetic and epigenetic differences. Methylation of cytosine residues provides one mechanism for the inheritance of epigenetic information. A genome-wide profiling of DNA methylation in two different genotypes of *Zea mays (ssp. mays)*, an organism with a complex genome of interspersed genes and repetitive elements, allowed the identification and characterization of examples of natural epigenetic variation. The distribution of DNA methylation was profiled using immunoprecipitation of methylated DNA followed by hybridization to a high-density tiling microarray. The comparison of the DNA methylation levels in the two genotypes, B73 and Mo17, allowed for the identification of approximately 700 differentially methylated regions (DMRs). Several of these DMRs occur in genomic regions that are apparently identical by descent in B73 and Mo17 suggesting that they may be examples of pure epigenetic variation. The methylation levels of the DMRs were further studied in a panel of near-isogenic lines to evaluate the stable inheritance of the methylation levels and to assess the contribution of *cis*- and *trans*- acting information to natural epigenetic variation. The majority of DMRs that occur in genomic regions without genetic variation are controlled by *cis*-acting differences and exhibit relatively stable inheritance. This study provides evidence for naturally occurring epigenetic variation in maize, including examples of pure epigenetic variation that is not conditioned by genetic differences. The epigenetic differences are variable within maize populations and exhibit relatively stable transgenerational inheritance. The detected examples of epigenetic variation, including some without Trends in Geneticshtly linked genetic variation, may contribute to complex trait variation.

**INTRODUCTION**

Much of the heritable variation within a species is a consequence of differences in the primary DNA sequence of different individuals. However, there is growing evidence for heritable variation in the absence of DNA sequence polymorphisms, termed epigenetic variation (Bird, 2007). Cytosine methylation is one of the molecular mechanisms that can contribute to epigenetic variation and often acts to suppress the activity of transposable elements, repetitive sequences, pseudogenes, and in some cases otherwise active genes (Chan et al., 2005; Slotkin & Martienssen, 2007). There is evidence that epigenetic changes can lead to stable phenotypic variation in plant and animal species (Rasmusson & Phillips, 1997; Richards, 2006; Feinberg, 2007; Jirtle & Skinner, 2007; Vaughn et al., 2007; Richards, 2008; Johannes et al., 2008). However, the abundance and role of epigenetic, relative to genetic, variation has not been well characterized. Maize (*Zea mays*) provides a useful model to study the role of epigenetic variation. Genetically, maize is a highly diverse species (Buckler et al., 2006; Messing & Dooner, 2006) with a large, complex genome with many interspersed genic and repetitive regions (Rabinowicz & Bennetzen, 2006; Schnable et al., 2009). While in the past this complex genomic structure has complicated the ability to perform genome-wide analyses it also is likely to contribute to higher levels of epigenetic variation relative to less complex genomes such as Arabidopsis (Slotkin & Martienssen, 2007; Weil & Martienssen, 2008). In addition, there are outstanding resources for the analysis of quantitative trait variation in maize (Yu et al., 2008; McMullen et al., 2009) that may allow for a better understanding of the relative roles of genetic and epigenetic variation in controlling quantitative trait variation.

In plants, the majority of genome-wide methylation studies have been conducted in Arabidopsis (Lippman et al., 2004; Zhang et al., 2006; Zilberman et al., 2007; Vaughn et al., 2007). In these studies DNA methylation was frequently associated with heterochromatic regions, transposable elements, and repetitive DNA (Lippman et al., 2004). In general, lower levels of methylation occur within gene promoter sequences; however when present, promoter methylation shows a negative correlation with gene expression (Zhang et al., 2006). Within gene bodies, regions of DNA methylation have been observed uniquely in the CG context, but no major impact on gene expression is associated with this modification (Zilberman et al., 2007). The exact role of gene body methylation is unclear, but it may preferentially affect moderately-transcribed genes (Zhang et al., 2006), and be under different regulatory control than that of transposable element

methylation (Miura et al., 2009). Similar genome-wide patterns of DNA methylation have also been observed in rice and poplar (Feng et al., 2010). A recent analysis of DNA methylation in maize used a 0.3X coverage sequencing of McrBC digested DNA to show evidence for mutually exclusive patterns of DNA methylation and H3K27me3 near genes with low, or no expression (Wang et al., 2009).

DNA methylation has been proposed to play a role in generating variation that could provide adaptation to environmental stresses (Richards, 2006; Finnegan, 2002; Kalisz & Purugganan, 2004; Lukens & Zhan, 2007; Richards, 2011). Two groups have recently developed epiRIL populations in which epigenetic states were altered by passage through DNA methylation mutants (Reinders et al., 2009; Johannes et al., 2009; Roux et al., 2011; Johannes & Colome-Tatche , 2011). The existence of quantitative trait variation in these populations suggests that alteration of DNA methylation patterns can induce phenotypic change although it is difficult to rule out the potential for primary sequence changes due to activated transposition. These studies have been very useful for documenting an important role for DNA methylation in regulating complex traits but do not provide information on natural variation for epigenetics states. There is evidence that DNA methylation patterns at specific loci can vary within Arabidopsis ecotypes (Vaughn et al., 2007; Riddle & Richards, 2002; Riddle & Richards, 2005; Woo et al., 2007; Zhang et al., 2008) and there are several specific examples of epigenetic variation that result in phenotypic variation in a variety of species (Bender & Fink, 1995; Jacobsen & Meyerowitz, 1997; Cubas et al., 1999; Morgan et al., 1999; Chandler et al., 2000; Stokes et al., 2002; Rakyan et al., 2003; Suter et al., 2004; Manning et al., 2006). However, there are limited analyses of genome-wide methylation variation conducted in plant species. A detailed contrast of chromosome 4 methylation patterns in *Arabidopsis thaliana* ecotypes shows very similar targeting of transposable elements and repetitive sequences, yet genic partial methylation states were highly polymorphic across ecotypes (Vaughn et al., 2007). A study in maize found evidence for variable effects of CHG methylation on transcription patterns in different inbreds of maize (Makarevitch et al., 2007).

Despite the evidence for variation in DNA methylation patterns among individuals of the same species relatively little is known about the nature of the inheritance of these methylation differences. One study in Arabidopsis found that gene body methylation was only partially heritable and was lost at a relatively high frequency (Vaughn et al., 2007). Richards (2006)

provided a description of how methylation variation may be dependent upon, conditioned by, or independent of DNA sequence change and termed these as examples of obligatory, facilitated or pure epialleles, respectively. Obligatory epialleles exhibit different levels of DNA methylation but are entirely dependent upon DNA sequence changes at linked or unlinked sites. In contrast, facilitated and pure epialleles exhibit stochastic variation that can be conditioned with or without genetic differences, respectively. In addition, there is evidence that genetic variation at unlinked genomic regions can direct epigenetic changes elsewhere in the genome, presumably through an RNA-direct DNA methylation (RdDM) pathway (Bender & Fink, 1995; Slotkin et al., 2005; Arteaga-Wazquez et al., 2010).

We sought to characterize the variation in DNA methylation patterns in two maize inbred genotypes, B73 and Mo17. Genome-wide profiling of DNA methylation patterns was used to assess the relationship of methylation to chromosomal and gene structures. Although the majority of the genome shows highly similar methylation patterns in both inbreds there are several hundred differentially methylation regions (DMRs) found throughout the maize genome. The analyses of several identical-by-descent regions of the B73 and Mo17 genomes provides evidence that epigenetic variation can occur in the absence of nearby genetic polymorphisms. A population of near-isogenic lines was used to further characterize the heritable behavior of the DMRs and to assess the genomic regions that controlled the differential methylation.

**RESULTS**

An array platform containing 2.1 million long oligonucleotide probes was designed to profile genomic DNA methylation patterns in low-copy sequences throughout the maize genome (see Methods). Methylated DNA immunoprecipitation (meDIP) was performed on fragmented genomic DNA using a 5-methylcytosine antibody. This approach is very useful for providing cost-effective quantitative estimates of DNA methylation density (Zhang et al., 2006). This method can detect substantial differences in the proportion of methylated cytosines in genomic regions but cannot accurately assess individual bases or differentiate the different types of DNA methylation such as CG, CHG, CHH. The enrichment of methylated DNA was confirmed (Figure 1A) by assessing the enrichment for a region known to be methylated and lack of enrichment for a region known to lack significant DNA methylation (Haun et al., 2007). meDIP was performed

on three biological replicates of leaf blade tissue isolated from the third expanded leaf of greenhouse-grown B73 and Mo17 seedlings; the resulting enriched fractions were labeled and hybridized to the array along with un-enriched control input DNA. Linear model ANOVA was used to estimate values for input DNA, B73 methylation, Mo17 methylation and relative methylation in B73 and Mo17 (Figure 2).

The probe sequences were designed based on the sequence of the B73 reference genome, but previous studies have documented abundant DNA sequence polymorphisms (Gore et al., 2009) and structural variants between B73 and Mo17 (Springer et al., 2009; Belo et al., 2010). We invesTrends in Geneticsated the methylation levels at sequences that exhibit structural genomic variation such as copy number variation (CNV) and presence-absence variation (PAV). Comparative genomic hybridization (CGH) data were obtained from the hybridization of input B73 and Mo17 genomic DNA. DNAcopy (Venkatraman & Olshen, 2007) was performed, followed by expectation maximization (Dempster et al., 1977) model analysis to identify segments that exhibit significantly more copies in Mo17 than in B73 (M>B CNV) and to identify segments that exhibit significant fewer or no copies in Mo17 relative to B73 (M<B CNV and PAV). While there are examples of CNV or PAV that show high levels of methylation, there is little evidence for substantial differences in the overall methylation levels of sequences that exhibit structural genome variation relative to sequences that do not show structural variation (Figure 3, 4A-C). Even so, for subsequent analyses, we focused on regions that do not have any evidence for CNV/PAV based on the CGH data. The values from the remaining probes were adjusted using the B73-input vs. Mo17-input hybridization coefficient to control for differential hybridization efficiency while still estimating methylation differences.

Microarray probes were selected to target low-copy regions of the maize genome by using repeat-masked sequences (provided by J. Stein and D. Ware). This repeat masking does not, however, remove all multi-copy sequences. The number of exact (100% identity and coverage) or close matches (>90% identity and coverage) was determined for each probe. Slightly over half (58%) of the probes present on the array have only a single perfect match in the B73 reference genome and no other close matches. As the numbers of perfect or close matches for probes increase there is a significant increase in the levels of methylation they detect (Figure 4D-E). This copy-number dependent increase in methylation levels is observed in both B73 and Mo17 (data not shown). The subsequent genome-wide analyses of DNA methylation are

confined to the subset of probes that are present as a single copy within the B73 genome (Table 1). The genome-wide analysis of methylation levels in Mo17 is further restricted to those probes that do not exhibit evidence for substantial differences in CGH values in the two inbreds (Table 1). By focusing on these subsets of probes the effects of probe copy number and genomic polymorphism on the detected methylation levels are minimized.

The distribution of per-probe methylation estimates provides evidence for a bi-modal distribution (Figure 1B) with the two distributions accounting for methylated and un-methylated genomic regions. Application of expectation maximization allows classification of the methylation status of each probe (Table 1). The genomic distribution of DNA methylation patterns was visualized across each of the maize chromosomes (Figure 2C and Figure 5). Similar to other species, methylation levels are higher in the pericentromic regions of maize chromosomes than at the ends of chromosomal arms. There are several regions of higher methylation throughout the chromosome that do not correlate with the centromeric position and do not correlate well with cytologically visible features such as knobs or rDNA sites. In general, the relative levels of DNA methylation are inversely correlated with gene density. The relative levels of DNA methylation in parental lines also show a negative correlation with recombination rates measured in a set of intermated B73 x Mo17 RILs (Liu et al., 2009). However, the exact parents for this population may have slight differences relative to the B73 and Mo17 profiled in this study and we have not measured actual DNA methylation profiles in any specific RIL genotype.

***Comparative genomic analysis of DNA Methylation dynamics for maize genes***

The location of each probe was determined relative to the gene models of annotation 5a.59 (www.maizesequence.org). Version 5a.59 of the maize working gene set contains 104,369 annotated genes which include 39,384 genes models that are part of the high-confidence filtered gene set (FGS) and another 64,985 genes that were rejected from the FGS. In both B73 and Mo17 the FGS genes show substantially lower methylation levels within and surrounding the genes relative to the rejected genes (Figure 6A). The reasons for rejecting genes from the FGS include low confidence FGENESH models, probable transposons and probable pseudogenes. Rejected genes that fall into these classes exhibit significantly higher methylation than the genes in the FGS (Figure 7). The methylation pattern for FGS genes has reduced methylation in the 300bp upstream of the transcription start site (TSS) then has a short “peak” of methylation in

the very beginning of the gene which drops off quickly in the 3' direction. There is a region of low methylation at the 3' ends of FGS genes, but methylation returns to genome-wide average levels within 500bp of the transcription termination site. This distribution of methylation levels, particularly the increased methylation at the 5' ends of genes, is distinct from patterns observed in other species (Lister et al., 2008; Feng et al., 2010) but is consistent with a previous report from maize (Wang et al., 2009). The methylation pattern observed across the gene body is related to the distribution of CpG dinucleotides (Figure 8A-B). However, the analysis of the region immediately upstream of the transcription site reveals that this region with increased CpG content does not show increased DNA methylation levels which confirms hypomethylation of these promoter regions and provides evidence that observed methylation levels are not strictly driven by CpG content (Figure 8A). In addition to the dynamics of methylation along the length of the genes, there are also significant differences in the methylation levels of exons, introns and UTRs relative to intergenic probes (Figure 8B). Introns show relatively low methylation levels throughout the gene body while exon sequences exhibit relatively high methylation in the 5' end and low methylation in the 3' end of the gene (Figure 8C). Many of these differences reflect the relatively high CpG content of the first exons of maize genes.

To determine the relationship between DNA methylation and gene expression, the relative expression levels of FGS genes in B73 leaf tissue were used to divide genes into five categories: non-expressed; and four quartiles based on RNA-seq data from Li et al. (2010). As expected, highly expressed genes show the lowest levels of methylation. There are significant differences in DNA methylation values among all quartiles of genes except between the two quartiles containing highly expressed genes (Figure 6B-C). Genes that are not expressed have higher levels of methylation in nearby regions as well as within the gene body. We proceeded to assess methylation levels of FGS genes in a comparative genomics context. Schnable et al. (Schanble et al., 2011) used comparative genomic approaches to identify homoeologous regions of the maize genome derived from a whole genome duplication event and to then assign them to sub-genome 1 and sub-genome 2 based on the level of fractionation observed. Sub-genome 1 has retained a larger proportion of the ancestral genes and generally exhibits higher mRNA expression levels as compared to sub-genome 2. Despite the trend for lower expression levels for genes in sub-genome 2 (Schnable et al., 2011), there was no evidence for differences in methylation levels in genes present in sub-genome 1 relative to sub-genome 2 (Figure 8D). However, there was evidence for substantial differences in the methylation levels of genes in

the FGS that are in syntenic positions relative to sorghum and/or rice relative to FGS genes that are located in non-syntenic positions (Figure 6D). The non-syntenic genes are enriched (chi-square p value <0.001) for genes that are not expressed or are in the lowest quartile of expressed genes based on the data of Li and coworkers (Li et al., 2010).

***Variability for B73-Mo17 methylation***

A visual analysis of the B73 and Mo17 methylation patterns revealed that while the majority of loci exhibit very similar patterns, there are examples of altered methylation levels between the two genotypes (Figure 2). Two different approaches were used to discover differentially methylated regions (DMRs) in B73 and Mo17. One approach identified individual probes that exhibit significant (q<0.001) differences in the contrast of B73 and Mo17 methylation (Table 2). A second approach implemented the DNAcopy segmentation algorithm on the relative methylation values followed by expectation maximization to identify segments of at least three probes that exhibit altered methylation (Figure 10A; Table 3). The per-probe analysis is capable of identifying small regions with altered methylation whereas the analysis of segments defined by adjacent probes will identify larger, high-confidence DMRs.

The DMRs identified by segmentation were further characterized as they have evidence for altered methylation from multiple adjacent probes encompassing a region of at least several hundred base pairs. There are nearly 700 DMR segments each that exhibit either B73 or Mo17 hypermethylation (Table 3). The Mo17 hypermethylation segments include a total of 500kb of DNA while the B73 hypermethylation segments include a total of 350kb of DNA. The majority of the DMRs (674/690) are less than 5kb in length and only one segment is over 10kb (Figure 9A). The majority of the DMRs occurred in intergenic regions and relatively few even overlap with a FGS gene or a member of the WGS (Table 3). Those genes that were contained within DMRs were enriched for non-syntenic genes, inclusion in sub-genome 1, and for those that are not expressed in leaf tissue (Figure 9B -D). A genomic visualization of probes and/or segments of differential methylation (Figure 2; Figure 5) revealed a non-uniform genomic distribution.

The DMRs may be conditioned by local sequence differences in B73 and Mo17 or may be the result in stochastic epigenetic differences that are not directly attributable to genetic differences. We focused on genomic regions of low diversity to identify potential examples of epigenetic differences that are not directly attributable to local sequence changes. As

previously reported (Springer et al., 2009) there are several large seemingly non-polymorphic regions in B73 relative to Mo17. These are likely identical-by-descent (IBD) regions that represent shared inheritance of the same haplotype block in these two different inbred genotypes from a common parent (Springer et al., 2009). We analyzed 10 putative IBD regions in the B73-Mo17 genome that are at least 2 Mb in length, have no evidence for structural variation, and have extremely low SNP densities (Table 4). The SNP rates in these regions (1 every 44.2kb) are below the levels of sequence error rates reported for the B73 reference genome (Schnable et al., 2009). Despite the near-absence of genetic variation within these regions there are 52 differentially methylated probes and 9 DMRs within these regions. The large low diversity region on chromosome 8 provides several examples of altered methylation levels within a large region that lacks sequence differences (Figure 10B and C). We used the Mo17 whole-genome shotgun sequences and targeted PCR to confirm that absence of any InDels within 2kb of each of the nine DMRs located within the IBD regions. The majority (8/9) DMRs in IBD regions did not have any InDels. Only one of these DMRs in an IBD region exhibit sequence polymorphism. At this DMR there is evidence for a recent insertion of a repetitive element in the B73 allele and B73 is more highly methylated than Mo17.

***Characterization of inheritance for differential methylation***

A set of 33 DMR regions consisting of 14 regions of B73 hypermethylation and 19 regions of Mo17 hypermethylation was selected for further characterization (Table 5). These included eight DMRs that were present in the IBD regions. Quantitative PCR-based assays were developed to assess relative methylation levels following digestion with the methylation dependent enzymes MspJI and FspEI. The methylation differences observed in the full-genome profiling were confirmed for 28/33 of these regions in independent biological samples of B73 and Mo17 DNA (Table 5). The differential methylation was also assessed using the methylation-sensitive enzymes HpaII and/or PstI for ten of these same DMRs including the three that had not been supported by methylation-dependent digests. All DMRs (4/4) that include a PstI (CHG sensitive) site were validated and 8 of the 10 DMRs that had a HpaII site were validated. Two of the three regions that were not conclusively validated by the methylation-dependent enzyme digests did exhibit differential methylation when tested with HpaII. The other DMR was not supported by assays with either methylation-dependent or –sensitive enzymes. The

classification of differential methylation was also supported by an analysis of read counts from methyl-sensitive and insensitive sequencing libraries from Gore et al (2009).

We assessed relative methylation levels for 13 of the DMRs in selected genotypes from a population of near-isogenic lines (NILs) derived from B73 and Mo17 (Eichten et al., 2011). The levels of methylation in NILs can be used to evaluate the relative contribution of linked and unlinked genomic regions and to test for paramutation-like transfer of information between alleles. The expected results for each of these potential scenarios are shown in Figure 11A. The genotype for the chromosomal region containing the DMR is expected to predict the methylation state in the NIL if the methylation change is purely epigenetic or if linked sequence polymorphisms regulate methylation levels. Alternatively, if unlinked genomic regions are directing the methylation levels at DMRs introgressed into a NIL then the DMR is expected to exhibit methylation levels similar to the recurrent parent. For each of the DMRs we selected several genotypes that provided an introgression of the locus into either a B73 or Mo17 genomic background. In addition, as a control we monitored DNA methylation levels in several NIL genotypes that did not have an introgression at the locus of the DMR. In general, the control assays show a high stability of DNA methylation levels at these DMRs. The analysis of the NILs with introgressions at the DMR loci reveal that 10/13 have methylation levels that can be predicted by the haplotype of the region surrounding the DMR. This could reflect stable inheritance of epigenetic variation or *cis*-linked genetic changes that are directing the methylation difference. The other three DMRs that were mapped have DNA methylation patterns that are influenced by genomic regions that are unlinked to the DMR locus, suggesting that the methylation levels of these loci are controlled by *trans*-acting loci. Four (of the 13) DMRs that were mapped are located within the IBD regions and each of these exhibited methylation patterns that were controlled by *cis*-linked regions despite the absence of closely linked genetic variation within these regions.

The relative DNA methylation patterns for these 13 DMRs were also assessed in a panel of 10 other inbred lines of maize and two teosinte inbred lines (TILs) (Figure 11B). Each of these DMRs exhibits at least one other genotype with high or low levels of methylation indicating that the B73-Mo17 states are not unique within maize.

## DISCUSSION

Maize has a rich history of serving as a model for epigenetic studies. The first examples of imprinting and paramutation were discovered in maize (Chandler & Stam, 2004; Springer & Gutierrez-Marcos, 2009) and there have been a number of pioneering studies on the epigenetic regulation of transposable element in maize (Fedoroff et al., 1995; Lisch et al., 2002). While these discoveries have been enabled by the ease of genetic studies in maize it is also likely that the complex organization of the maize genome with many interspersed transposons and genes has led to numerous examples of epigenetic regulation. In this study we have performed a genome-wide characterization of DNA methylation levels in two inbred lines of maize and found hundreds of loci with differences in DNA methylation levels. This study of natural epigenetic variation also demonstrates the utility of near-isogenic lines for characterizing epialleles.

### *Methylation dynamics along maize chromosomes and genes*

Our data provide evidence for higher levels of DNA methylation in the pericentromeric regions of maize chromosomes. This is quite similar to observations in Arabidopsis (Lippman et al., 2004; Zhang et al., 2006; Zilberman et al., 2007). Although there is a general negative correlation between methylation density and recombination, methylation density does not exactly mirror recombination rates. Recombination happens much more frequently near the ends of the chromosomes and quickly drops to a lower level internally (Liu et al., 2009). The DNA methylation patterns exhibit a much more gradual change along the length of the chromosomes, potentially suggesting that the differences in recombination rate along the length of a chromosome are not directly related to DNA methylation levels. In general, the methylation patterns within maize gene bodies are similar to the density of CG sites. However, it is clear that the methylation levels in regions immediately upstream and downstream of gene bodies are not reflective of CG density. Interestingly, while short maize genes have elevated CG content throughout the gene body the longer maize genes have elevated CG content only in the first 500-1000bp.

We also noted very distinct patterns of DNA methylation in maize genes. In general, methylation is lower in high-confidence genes (i.e., members of the FGS), highly expressed genes and genes in syntenic positions relative to other grass species. Putative genes that were rejected during stringent genome annotation (often partial length sequences or putative

transposons) are more highly methylated. The lack of detectable methylation differences between genes located in the maize1 and maize2 sub-genomes was surprising because there is evidence that maize1 genes are generally more highly expressed than maize2 genes (Schnable et al., 2011). Therefore, DNA methylation is unlikely to provide the mechanism for controlling the expression differences between the two sub-genomes of maize. The significant difference in average methylation levels for syntenic and non-syntenic genes is correlated with lower expression levels for non-syntenic genes. The non-syntenic genes that are conserved among the grasses but exhibit different genomic positions are the result of gene movement, perhaps mediated by transposons. Approximately 1/3 of the maize FGS genes are not in syntenic positions relative to other grasses, yet nearly all genes invesTrends in Geneticsated by classical genetics to date belong to the fraction of the genome located in syntenic positions (Schnable & Freeling, 2011). Their higher levels of methylation may reflect the presence of transposon sequences near these genes or may result from the insertion of a gene into new chromosomal environments that are lacking some of their ancestral regulatory sequences.

***Many stable DMRs are found in B73 and Mo17***

Several groups have demonstrated that perturbation of epigenetic information can affect quantitative traits (Reinders et al., 2009; Johannes et al., 2009; Roux et al., 2011; Johannes & Colome-Tatche, 2011). In addition, the existence of natural epigenetic variation has been demonstrated for individual loci (Riddle & Richards, 2002; Riddle & Richards, 2005; Woo et al., 2007; Zhang et al., 2008; Bender & Fink, 1995; Jacobsen & Meyerowitz, 1997; Cubas et al., 1999; Morgan et al., 1999; Chandler et al., 2000; Stokes et al., 2002; Rakyan et al., 2003; Suter et al., 2004; Manning et al., 2006) or chromosomes (Vaughn et al., 2007). A primary objective of this study was to document the prevalence and distribution of variable methylation levels in different maize genotypes. There are hundreds of examples of differential methylation in B73 and Mo17. In general, many of these DMRs are located in intergenic regions and may reflect differences in transposon silencing among the genotypes. However, at least 71 of the 690 variable methylation regions are found within 500bp of a high-confidence gene (FGS). Following the discovery of these regions we were able to pursue further characterization using a population of NILs. The NILs provide a useful tool for assessing the stability of DNA methylation patterns and for testing whether the epigenetic variation is caused by genetic differences elsewhere in the genome.

The analysis of DNA methylation levels at several DMRs within the NILs addressed the stability of the DNA methylation patterns. The near-isogenic lines were developed by three back-crosses followed by at least four rounds of self-pollination (Eichten et al., 2011). In an analysis of the control lines (lines without an introgression at the DMR locus for eleven *cis*-controlled DMRs) there is evidence for stable inheritance as 85% of the assays reveal the expected methylation level. There are a small number of assays (5/150) that exhibit a completely changed methylation state and another 17/150 exhibit a partial gain or loss of DNA methylation. These examples may reflect inaccuracies in our measurements of DNA methylation or actual instability of DNA methylation patterns. In general, we observe relatively stable inheritance with rare examples of both gains and losses of DNA methylation. We did not observe evidence for paramutation-like effects where methylation levels were affected by heterozygosity for the DMRs.

Stable differences in DNA methylation levels between two genotypes can be the result of differences in epigenetic state that are faithfully propagated to offspring. Alternatively, they may be the result of genetic changes elsewhere in the genome that direct epigenetic modifications at an unlinked site. For example, structural rearrangements of the PAI1-PAI4 locus on chromosome 1 of Arabidopsis control the methylation state of the two other PAI loci on chromosomes 1 and 5 (Bender & Fink, 1995). The conditioning of DNA methylation state by either linked or unlinked genomic regions would be examples of obligatory epialleles where a genetic change at one locus programs variable methylation at another locus. We studied the contribution of linked and unlinked genomic regions to the methylation differences between B73 and Mo17 for 13 of the DMRs using a series of near-isogenic lines (Eichten et al., 2011). Three of the regions exhibit evidence for *trans*-acting control of DNA methylation patterns. The remaining ten loci have methylation patterns that are either stably inherited or are continuously directed by local sequence changes.

***Evidence for pure epialleles in maize***

A major unresolved question about epigenetic variation is whether the majority of epigenetic variation exhibits strong linkage disequilibrium with nearby genetic differences (Richards, 2008). If genetic markers, such as SNPs, are in strong linkage disequilibrium with epigenetic changes then the functional consequences of epigenetic differences would likely be revealed by assays of linked genetic differences. In particular, “obligatory” epialleles are entirely conditioned by

nearby genetic changes (Richards, 2006). Alternatively, "facilitated" epialleles exhibit stochastic variation in epigenetic state with a conditioning genetic change, and "pure" epialleles exhibit stochastic variation in epigenetic state independent of any genetic changes (Richards, 2006). Both facilitated and pure epialleles will show differences in epigenetic state that are not completely linked to, or predicted, by nearby genetic polymorphisms.

We were interested in whether some of the epigenetic changes between B73 and Mo17 might be due to epigenetic changes that are not directly caused by nearby genetic differences such as transposon insertions or by unlinked rearrangements that might direct methylation in *trans* via RNA-directed DNA methylation. This led us to focus on the 10 extended B73-Mo17 identical-by-descent regions. These regions are most likely the result of shared inheritance of a chromosomal region from a genotype that was used in the pedigree of both B73 and Mo17. At least one genotype, CI187-2, is present in the pedigree of both B73 and Mo17 (Hansey et al., 2011). It is therefore possible that these regions could exhibit identity by descent (IBD). It is worth noting that other maize genotypes have alternative haplotypes in these regions so they are not the result of large selective sweeps among all maize genotypes (Swanson-Wagner et al., 2010). The absence of detected structural variation and few SNPs within these regions suggest that any observed epigenetic differences between B73 and Mo17 are not the result of nearby genetic polymorphisms. We found 9 DMRs (of 690 genome-wide) within these IBD regions and only one of these has evidence for a nearby genetic change. The number of DMRs within the IBD regions (9) is very close to the number we would have expected (13) based on the frequency per Mb within the whole genome. The finding that all four of these regions that were assessed in NILs show stable inheritance provides evidence for heritable epigenetic information in the absence of genetic differences. Our initial focus on IBD regions allowed the discovery of epigenetic variation without nearby genetic changes given the extended regions of identity. However, it is likely that many of the DMRs that are located in non-identical by descent genomic regions may also be the result of purely epigenetic changes. We noted that one-third of the 690 DMRs do not contain any SNPs in B73 relative to Mo17 within 1kb of the DMR and may represent epigenetic differences that are not conditioned by genetic differences.

This study provides a detailed view of the distribution of cytosine methylation in two maize inbreds. The evidence for faithfully inherited methylation differences, even in the absence of nearby genetic polymorphisms, provide evidence for at least partially stable epigenetic variation

in maize that would not be revealed by high-resolution analyses of genetic differences. There are likely functional consequences of the altered methylation levels in B73 and Mo17. There are several examples in which a DMR within an identical by descent region is near the promoter for a FGS gene and several of these genes exhibit differential expression in other tissues of B73 and Mo17 (data not shown). Further characterization of the relationship between expression variation and methylation variation may identify examples of epigenetic variation that affect phenotypic differences among inbred lines. This study, in combination with recent analyses of epiRILs in Arabidopsis (Reinders et al., 2009; Johannes et al., 2009; Roux et al., 2011; Johannes & Colome-Tatche, 2011), provides evidence for heritable epigenetic information that may contribute to quantitative trait differences within species. Future research is required to uncover evidence for the contribution of the variable methylation we have described in this study to phenotypic differences among maize genotypes.

**EXPERIMENTAL METHODS & PROCEDURES**

**Plant materials and DNA isolation**: Three replications of B73 and Mo17 seedlings were grown in a randomized block design. The seeds for each replication came from a unique, single source (ear). For each replication, 10 seedlings were grown in pots (5 seedlings per pot) that were assigned random positions. Seedlings were grown under controlled conditions in a greenhouse at the University of Minnesota (St. Paul, MN) with a light cycle of 15 hours lights on and 9 hours lights off each day. Seedlings were watered daily as needed. After 18 days of growth, the 3rd leaf (L3) of each plant was harvested and pooled with other plants from the same pot/replication and immediately frozen in liquid nitrogen. DNAs were isolated using the CTAB method. Phenol:chloroform extraction and subsequent precipitation in 0.1x volume Na-Acetate (3M) and 2x volume 100% EtOH was conducted to purify the DNA samples. 15-30ug of gDNA in 650-700uL nuclease-free water was sonicated for five, ten- second pulses as per the methods of (Haun & Springer, 2008). Samples were quantified and run on 1.5% agarose gels to verify that DNAs were fragmented to 200-400bp.

**Array Design and annotation**: A NimbleGen 2.1M feature long oligonucleotide array was designed using B73 RefGen2 assembly (provided by the Arizona Genomics Institute). The maize genome exhibits a complex architecture with many repetitive sequences interspersed with low-

copy genic sequences (Schnable et al., 2009). A repeat masked version of the pseudomolecule sequences from RefGenv2 of the B73 genome (provided by J Stein and D Ware) were used to design probes to low-copy regions. Thermally balanced probes were designed every ~200bp across the low-copy portion of the maize genome. The actual spacing varies in some cases to allow for ideal probe selection. In addition, a higher density of probes (one probe every ~56bp) was used for chromosome 9 to determine whether higher probe density provided increased resolution for methylation detection. The probes were each annotated with respect to their copy number in the B73 genome, the number of close matches and their location relative to gene models. Syntenic orthologs of maize genes in sorghum, rice, and brachypodium were identified using the combined synonymous substitution rate of syntenic blocks method described in (Schnable et al., 2011). A maize gene was considered to be recently inserted if orthologous locations could be identified in rice, sorghum, and brachypodium by the syntenic conservation of up and downstream genes, but no homologous gene nor unannotated homologous sequence was identified in any species at the predicted orthologous location.

**Immunoprecipitation of Methylated DNA, labeling and hybridization**: Methylated DNA was immunoprecipitated with an anti-5-methylcytosine monoclonal antibody from 400ng sonicated DNA using the Methylated DNA IP Kit (Zymo Research, Orange, CA; Cat # D5101). For each replication and genotype, whole genome amplification was conducted on 50-100ng IP DNA and also 50-100ng of sonicated DNA (input control) using the Whole Genome Amplification kit (Sigma Aldrich, St. Louis, MO, Cat # WGA2-50RXN). For each amplified IP input sample, 3ug amplified DNA were labeled using the Dual-Color Labeling Kit (Roche NimbleGen, Cat # 05223547001) according to the array manufacturer's protocol (Roche NimbleGen Methylation UserGuide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA was labeled with Cy3. The no anti-body control (negative control) was also labeled, using the average volumes required for the experimental samples. Samples were hybridized to the custom 2.1M probe array (GEO Platform GPL13499) for 16-20hrs at 42°C. Slides were washed and scanned according to NimbleGen's protocol for the GenePix4000B scanner. Images were aligned and quantified using NimbleScan software (Roche NimbleGen) producing raw data reports for each probe on the array.

**Normalization and linear modeling**: Pair files exported from NimbleScan were imported into the Bioconductor statistical environment (http://bioconductor.org/). Microarray data channels were

assigned the following factors: B73, Mo17, B73 input, or Mo17 input depending on sample derivation. Non-maize probes and vendor-supplied process control probes were configured to have analytical weights of zero. Variance-stabilizing normalization was used to account for array-specific effects. Factor-specific hybridization coefficients were estimated by fitting fixed linear model accounting for dye and sample effects to the data using the limma package (Smyth, 2004). To compute biologically relevant information about B73 and Mo17 DNA methylation, the following contrasts were then computed: B73 IP vs B73 input (B73 methylation); Mo17 input vs B73 input (CGH and differential hybridization efficiency); Mo17 IP vs [Mo17 input vs B73 input] (Mo17 methylation corrected for differential hybridization efficiency); B73 IP vs [Mo17 IP vs [Mo17 input vs B73 input]] (differential DNA methylation corrected for differential hybridization efficiency). Moderated t-statistics and the log-odds score for differential MeDIP enrichment were computed by empirical Bayes shrinkage of the standard errors with the False Discovery Rate controlled to 0.05 (Smyth, 2004). Full results are available for download from the following URL: http://genomics.tacc.utexas.edu/data/eichten-plos-genetics-2011-a. Microarray results were deposited with NCBI GEO under accession GSE29099.

**Defining CGH copy-number variations:** CGH data were obtained from the hybridization of un-enriched B73 and Mo17 genomic DNA (B73 and Mo17 input channels). DNAcopy (Venkatraman & Olshen, 2007) was performed to identify segments showing similar hybridization patterns based on chromosomal order of probes. Defined segments were analyzed by expectation maximization model analysis (Dempster et al., 1977; Fraley & Raftery, 2006) to identify segments that fall into three predicted sub-distributions with non-uniform variances. For each segment, the posterior probability that it occurred in each of the three distributions was determined. Segments that had > 0.95 probability of falling into either the first or third sub-distributions were defined as containing more copies in Mo17 than in B73 (M>B CNV) or significant fewer to no copies in Mo17 relative to B73 (M<B CNV and PAV) respectively.

**Analysis of variable methylation**: To define probes with differential methylation between the B73 and Mo17 inbreds, the significance values developed from the B73 vs Mo17 relative methylation linear model probes were used. Probes with a significance value of < 0.001 were considered differentially methylated between the two inbreds. The direction of the variation was determined based on the positive or negative value of the B73 methylation minus the Mo17

methylation state. A total of 4172 B73 hypermethylated and 5367 Mo17 hypermethylated probes were classified using this method.

To identify segments showing differential methylation between B73 and Mo17, the DNAcopy algorithm (Venkatraman & Olshen, 2007) was used on 1,088,820 Mo17 unique probes in the B73 vs. Mo17 relative methylation linear model results. The EM algorithm (Dempster et al., 1977; Fraley & Raftery, 2006) was used to estimate the mixing proportion, mean, and variance associated with three predicted sub-distributions with non-uniform variances found within the B73 vs Mo17 segments. For each segment, the posterior probability that it occurred in each of the three distributions was determined. Segments that had > 0.95 probability of falling into either the first or third sub-distributions were called as Mo17 hypermethylated or B73 hypermethylated segments respectively.

**qPCR**: Quantitative PCR (qPCR) was performed to evaluate the efficiency of 5-methylcytosine immunoprecipitation using regions of the *Mez1* gene known to have methylation (5'region) and non-methylation (Exon 9) (Haun & Springer, 2008). Primers were designed within the methylated region (forward primer 5'- TTGTGTCGAGGTCTCGAATG-3', reverse primer 5'-TGTTGAAGCGCATTAGCACT -3') and within the non-methylated region (forward primer 5'-CAACAAAGTGAAAGCTCTTCAACTGCAA-3', reverse primer 5'-CACAACACTCCCCTAGTCCCTCAAAAGTT-3'). Primer amplification and efficiency were tested in B73 and Mo17 genomic DNA. Three technical replications were included for each of two biological replications of B73 and Mo17 IP and input DNA samples. The relative amount of immunoprecipitated DNA (percentage of the input control DNA for each sample) was calculated (Figure 1). As expected, the IP negative control and qPCR no template controls either did not amplify or amplified approximately 10 cycles after the experimental samples (>1000 fold difference, data not shown). *Mez1* qPCR reactions were conducted using 100ng DNA and Light Cycler480 SYBR Green I Master (Roche, Cat # 04707516001) on the LightCycler480 instrument (Roche) in accordance with Roche's protocol for SYBR Green on the LightCycler480.

Primers were designed for 33 regions within DMRs. 1 microgram of genomic DNA was digested for 16hr with MspJI or FspEI (New England Biolabs). Mock digestions were performed substituting glycerol for restriction enzyme. qPCR reactions were performed using 37ng DNA and SsoFast EvaGreen Supermix (BioRad) on the Chromo4 instrument (BioRad) in accordance with SsoFast protocol. The difference between digest C(t) and mock C(t) was calculated for each

genotype tested. As our selected enzymes target methylated cytosines, higher methylation leads to increased digestion and subsequently longer C(t) times. DMRs in B73 and Mo17 were validated as higher methylation levels for larger differential C(t) value between the inbred lines. NIL and diverse inbred samples were compared across individual primer pairs and methylation state was determined by comparing C(t) difference values.

**REFERENCES**


Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, et al. (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proceedings of the National Academy of Sciences of the United States of America 107: 12986–12991.

Belo A, Beatty MK, Hondred D, Fengler KA, Li B, et al. (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theoretical and Applied Genetics 120: 355–367.

Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83: 725–734.

Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398.

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Current Opinion in Plant Biology 9: 172–176.

Chan SW-L, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Reviews Genetics 6: 351–360.

Chandler VL, Eggleston WB, Dorweiler JE (2000) Paramutation in maize. Plant Molecular Biology 43: 121–145.

Chandler VL, Stam M (2004) Chromatin conversations: mechanisms and implications of paramutation. Nature Reviews Genetics 5: 532–544.

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157–161.

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical SocietySeries B (Methodological) 39: 1–38.

Eichten SR, Foerster JM, de Leon N, Kai Y, Yeh CT, et al. (2011) B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiology 156: 1679–1690.

Fedoroff N, Schlappi M, Raina R (1995) Epigenetic regulation of the maize Spm transposon. BioEssays 17: 291–297.

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447: 433–440.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America 107: 8689–8694.

Finnegan EJ (2002) Epialleles - a source of random variation in times of stress. Current Opinion in Plant Biology 5: 101–106.

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, et al. (2009) A first-generation haplotype map of maize. Science 326: 1115–1117.

Hansey CN, Johnson JM, Sekhon RS, Kaeppler SM, Leon N de (2011) Genetic Diversity of a Maize Association Population with Restricted Phenology. Crop Science 51: 704–715.

Haun WJ, Laoueille-Duprat S, O'connell MJ, Spillane C, Grossniklaus U, et al. (2007) Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. The Plant Journal 49: 325–337.

Haun WJ, Springer NM (2008) Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. The Plant Journal 56: 903–912.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. Science 277: 1100–1103.

Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. Nature Reviews Genetics 8: 253–262.

Johannes F, Colome-Tatche M (2011) Quantitative epigenetics through epigenomic perturbation of isogenic lines. Genetics 188: 215–227.

Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics 9: 883–890.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics 5: e1000530.

Kalisz S, Purugganan MD (2004) Epialleles via DNA methylation: consequences for plant evolution. Trends in Ecology & Evolution 19: 309–314.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, et al. (2010) The developmental dynamics of the maize leaf transcriptome. Nature Genetics 42: 1060–1067.

Lippman Z, Gendrel A V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.

Lisch D (2002) Mutator transposons. Trends in Plant Science 7: 498–504.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (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, et al. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genetics 5: e1000733.

Lukens LN, Zhan S (2007) The plant genome's methylation status and response to stress: implications for plant improvement. Current Opinion in Plant Biology 10: 317–322.

Makarevitch I, Stupar RM, Iniguez AL, Haun WJ, Barbazuk WB, et al. (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase zmet2. Genetics 177: 749–760.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948–952.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Messing J, Dooner HK (2006) Organization and variability of the maize genome. Current Opinion in Plant Biology 9: 157–163.

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, et al. (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. The EMBO Journal 28: 1078–1086.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23: 314–318.

Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology 9: 149–156.

Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, et al. (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proceedings of the National Academy of Sciences of the United States of America 100: 2538–2543.

Rasmusson DC, Phillips RL (1997) Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Science 37: 303–310.

Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, et al. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939–950.

Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews Genetics 7: 395–401.

Richards EJ (2008) Population epigenetics. Current Opinion in Genetics & Development 18: 221–226.

Richards EJ (2011) Natural epigenetic variation in plant species: a view from the field. Current Opinion in Plant Biology 14: 204–209.

Riddle NC, Richards EJ (2002) The control of natural variation in cytosine methylation in Arabidopsis. Genetics 162: 355–363.

Riddle NC, Richards EJ (2005) Genetic variation in epigenetic inheritance of ribosomal RNA gene methylation in Arabidopsis. The Plant Journal 41: 524–532.

Roux F, Colome-Tatche M, Edelist C, Wardenaar R, Guerche P, et al. (2011) Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. Genetics 188: 1015–1017.

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PloS ONE 6: e17855.

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America 108: 4069–4074.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Slotkin RK, Freeling M, Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nature Genetics 37: 641–644.

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272–285.

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3: 3.

Springer NM, Gutierrez-Marcos F. J (2009) Imprinting in Maize. Handbook of Maize. University of Warwick Warwick HRI: Springer New York. 429–440.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, et al. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics 5: e1000734.

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in Arabidopsis disease resistance. Genes & Development 16: 171–182.

Suter CM, Martin DI, Ward RL (2004) Germline epimutation of MLH1 in individuals with multiple cancers. Nature Genetics 36: 497–501.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research 20: 1689–1699.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology 5: e174.

Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23: 657–663.

Wang X, Elling AA, Li X, Li N, Peng Z, et al. (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. The Plant Cell 21: 1053–1069.

Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Current Opinion in Genetics & Development 18: 188–192.

Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes & Development 21: 267–277.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539–551.

Zhang X, Shiu SH, Cal A, Borevitz JO (2008) Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. PLoS Genetics 4: e1000032.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126: 1189–1201.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39: 61–69.

Table 1. Methylation levels in subsets of probes

| Data Set | Probes | % Probes | % methylated (50%pp) |
|---|---|---|---|
| B73_methylation | 2,120,701.00 | 100.00 | 50.11 |
| B73_methylation_unique^* | 1,202,553.00 | 56.71 | 52.49 |
| Mo17_methylation~ | 1,940,644.00 | 91.51 | 41.37 |
| Mo17_methylation_unique^*~ | 1,088,820.00 | 51.34 | 49.11 |

**Filters (completed in order indicated above)**

^ = every third chromosome 9 probe for similar spacing relative to other chromosomes

* = unique probes (only one perfect match in genome)

~ = CGH filter (only probes with CGH values >-1)

Table 2. Probes with variable DNA methylation in B73 and Mo17

| **Probe Class** | # Probes | Mean B73 Methylation ($log_2$(meDIP/input)) | Mean Mo17 Methylation ($log_2$(meDIP/input)) | % present in segments | % Syntenic | % Intergenic |
|---|---|---|---|---|---|---|
| Mo17 hypermethylation | 5367 | -1.15 | 0.81 | 20.4 | 51.6 | 67.4 |
| B73 hypermethylation | 4172 | 1.05 | -1.03 | 23.2 | 37.5 | 72.3 |
| All probes | 1088820 | -0.21 | -0.21 | NA | 56.2 | 54.2 |

Table 3. Segments with variable DNA methylation levels in B73 and Mo17

| **Segment Class** | # Segments | Segment Mean | Avg # Probes | Avg Seg. Length | Avg # of FGS Genes | Avg # of WGS Genes | #Segs with at least 1 FGS | #Segs with at least 1 WGS | % of probes q<0.0001 |
|---|---|---|---|---|---|---|---|---|---|
| Mo17 hypermethylation | 402 | -1.6125 | 5.6 | 1241 | 0.119 | 0.303 | 46 | 111 | 39.9% |
| No change | 919 | 0.0010 | 1149.9 | 2148870 | 39.410 | 96.210 | 877 | 910 | NA |
| B73 hypermethylation | 288 | 1.7846 | 4.9 | 1219 | 0.087 | 0.285 | 25 | 74 | 44.4% |
| Unclassified | 269 | -0.1459 | 92.9 | 246008 | 3.000 | 8.530 | 197 | 238 | NA |

Table 4. IBD regions in the B73-Mo17 genome

| Chromosome | Start (Mb) | Stop (Mb) | Length (Mb) | Fold reduction in SNP diversity | Variable methylation probes | Variable methylation segments |
|---|---|---|---|---|---|---|
| 1 | 116 | 119 | 3 | 49.1 | 0 | 0 |
| 2 | 86 | 88 | 2 | 31.4 | 1 | 1 |
| 2 | 136.5 | 140.5 | 4 | 30.2 | 2 | 0 |
| 2 | 178.5 | 185 | 6.5 | 42.2 | 5 | 0 |
| 3 | 162.5 | 165 | 2.5 | 40.9 | 4 | 2 |
| 4 | 126 | 130 | 4 | 39.3 | 1 | 0 |
| 4 | 163 | 166.5 | 3.5 | 25.0 | 1 | 1 |
| 5 | 54 | 56 | 2 | 25.4 | 2 | 1 |
| 5 | 206 | 210 | 4 | 27.3 | 11 | 0 |
| 8 | 142.5 | 160 | 17.5 | 36.3 | 25 | 4 |

Table 5. Characterization of variable methylation segments

| Seg ID | Chr | Start | Stop | Assay | Mspj1 relative met[a] | FspEI relative met[a] | Confirmed | IBD region | *Cis*/ *trans* control |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 1 | 86899 | 87699 | MDMR_40 | -9.23 | -9.14 | Yes | | *cis* |
| 36 | 1 | 19,520,608 | 19,522,008 | MDMR_43 | -2.84 | -7.66 | Yes | | |
| 42 | 1 | 19982857 | 19984474 | MDMR_82 | -3.79 | -5.22 | Yes | | |
| 94 | 1 | 39036362 | 39039562 | MDMR_24 | -5.14 | -5.87 | Yes | | *trans* |
| 213 | 1 | 160563568 | 160563968 | BDMR_22 | 2.73 | 2.44 | Yes | | *cis* |
| 215 | 1 | 162,538,673 | 162,540,073 | MDMR_44 | -7.81 | -11.09 | Yes | | |
| 247 | 1 | 190761322 | 190762322 | MDMR_27 | -5.65 | -10.1 | Yes | | *trans* |
| 710 | 2 | 144,048,581 | 144,048,981 | BDMR_45 | -1.38 | -0.72 | No | | |
| 788 | 3 | 7,683,400 | 7,686,600 | MDMR_73 | -3.7 | -8.04 | Yes | Yes | *cis* |
| 792 | 3 | 8,346,040 | 8,348,062 | MDMR_74 | -6.36 | -9.79 | Yes | Yes | *cis* |
| 794 | 3 | 8,357,591 | 8,358,191 | BDMR_78 | 3.76 | 9.4 | Yes | Yes | |
| 999 | 3 | 183,380,573 | 183,381,173 | BDMR_3 | -0.82 | -0.17 | No | | |
| 1060 | 3 | 205,814,185 | 205,816,207 | MDMR_41 | -6.3 | -13.2 | Yes | | |
| 1064 | 3 | 206,677,794 | 206,678,782 | BDMR_48 | 1.9 | 3.34 | Yes | | |
| 1066 | 3 | 209,258,739 | 209,259,431 | BDMR_51 | 1.31 | 3 | Yes | | |
| 1193 | 4 | 108,425,289 | 108,425,889 | BDMR_79 | -6.21 | 1.17 | No | Yes | |
| 1252 | 4 | 140,005,016 | 140,005,441 | MDMR_1 | -4.73 | -3.93 | Yes | | *cis* |
| 1280 | 4 | 160,954,336 | 160,955,336 | MDMR_37 | -5.36 | -9.87 | Yes | | *cis* |
| 1479 | 5 | 69,250,995 | 69,252,274 | BDMR_49 | 5.58 | 4.69 | Yes | | *cis* |
| 1493 | 5 | 96,853,806 | 96,854,606 | BDMR_53 | 3 | 1.64 | Yes | Yes | *cis* |
| 1603 | 6 | 60,183,875 | 60,184,275 | MDMR_8 | -4.9 | -3.95 | Yes | | |
| 1697 | 6 | 161,113,703 | 161,114,709 | MDMR_4 | -4.34 | -9.87 | Yes | | |
| 1840 | 7 | 150,216,544 | 150,218,384 | BDMR_31 | 2.94 | 2.02 | Yes | | |
| 1895 | 8 | 97,274,412 | 97,276,412 | MDMR_36 | -3.58 | -1.43 | No | | |
| 1938 | 8 | 143,704,696 | 143,705,542 | MDMR_75 | -5.34 | -10.08 | Yes | Yes | |
| 1940 | 8 | 145,875,662 | 145,877,862 | MDMR_76 | -3.41 | -2.52 | Yes | Yes | *cis* |
| 1946 | 8 | 151,080,956 | 151,083,158 | MDMR_77 | -1.34 | ND | Yes | Yes | |
| 1968 | 9 | 3,855,741 | 3,856,077 | BDMR_62 | 2.84 | 2.9 | Yes | | |
| 2005 | 9 | 20,864,861 | 20,865,197 | MDMR_13 | -2.25 | -1.99 | Yes | | |
| 2033 | 9 | 37,257,975 | 37,258,311 | BDMR_59 | 2.52 | -0.83 | No | | |
| 2178 | 9 | 116,238,414 | 116,238,750 | BDMR_32 | 2.9 | 2.93 | Yes | | |
| 2209 | 9 | 145,760,392 | 145,760,896 | BDMR_47 | 3.06 | 3.89 | Yes | | *cis* |
| 526 | 10 | 126586236 | 126586636 | MDMR_38 | -5.66 | -6.09 | Yes | | *trans* |

[a]The relative methylation is calculated as the (B73 mock Ct - B73 digest Ct) - (Mo17 mock Ct - Mo17 digest Ct). Values above zero reflect higher methylation in Mo17 while values below zero reflect higher methylation levels in B73.

[b]The number of methylated and unmethylated inbreds (from Figure 11B) is reported.

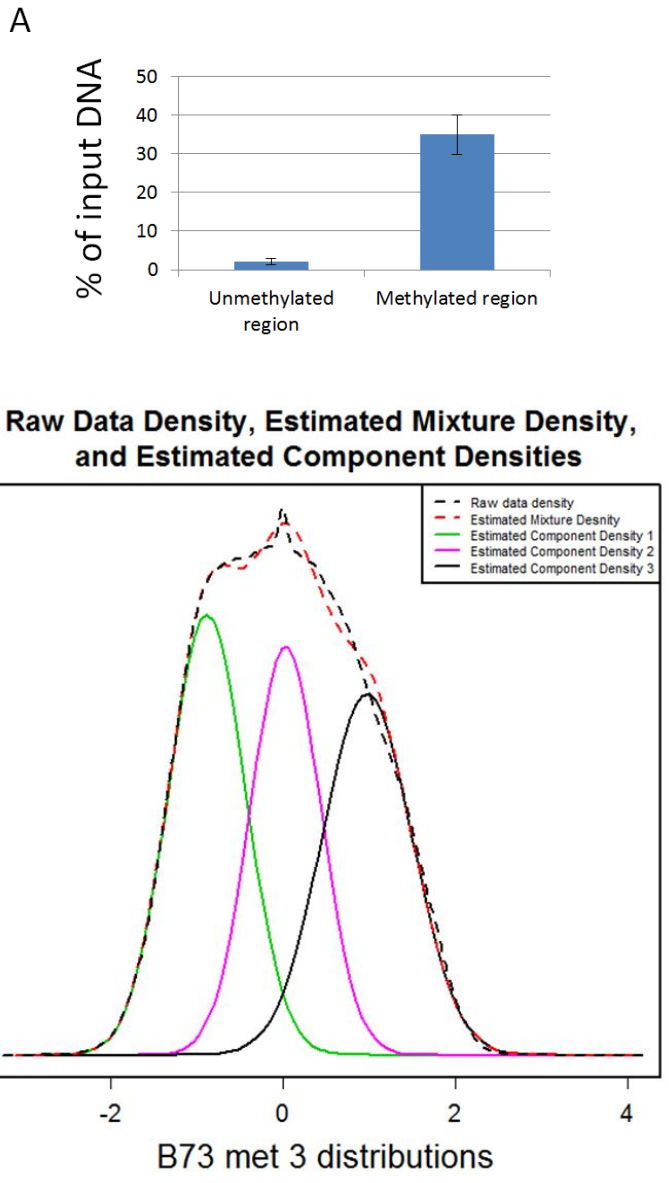



Figure 1: Enrichment of methylated DNA by immunoprecipitation. (A) The percent of input DNA recovered following 5-methylcytosine immunoprecipitation of three biological replicates of B73 was determined for two different regions by qPCR. The unmethylated region is 5,270 to 5,380 of Mez1 (exon 9) and the methylated region is from -1,238 to -1,038 of Mez1. (B) A density plot is used to visualize the distribution of all B73 log2(IP/input) values (black dotted line). This observed distribution can be approximated by an expectation maximization model that assumes three normal distributions (solid lines that add up to the red dashed line). Values with a high posterior probability of being sampled from the black distribution are assigned as methylated.

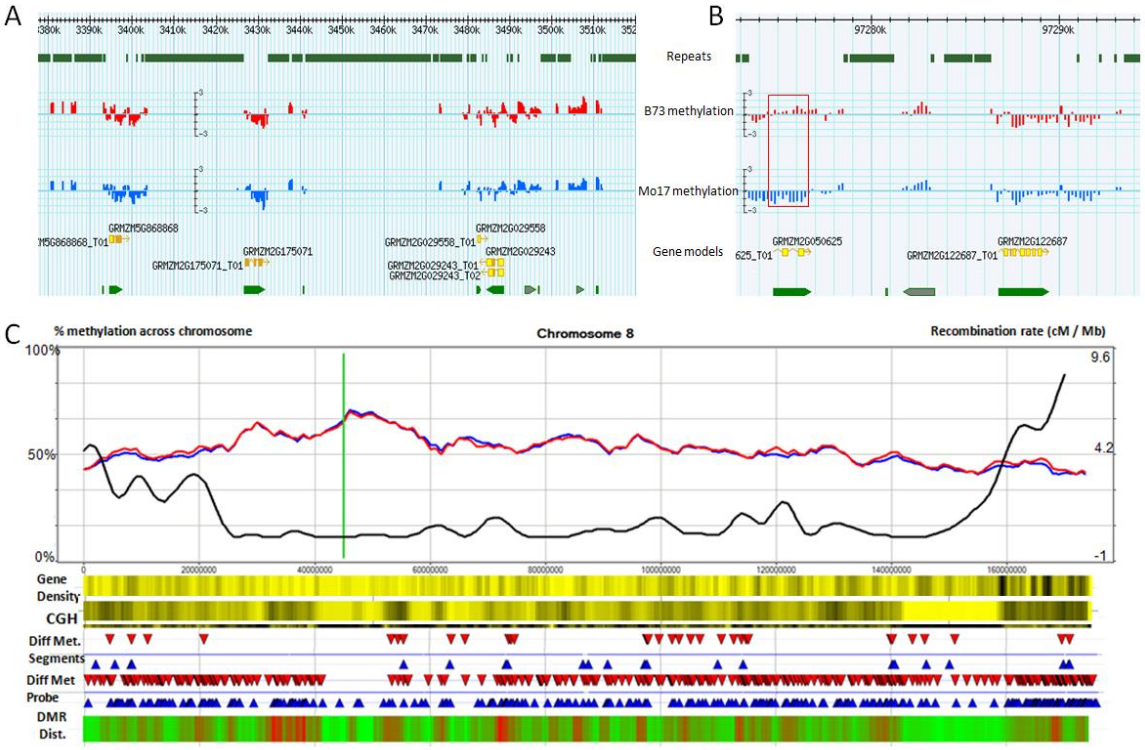



Figure 2: Synopsis of chromosome-level methylation and variation. (A - B) A Gbrowse view is presented for 140kb of chromosome 8 (3,380kb-3,520kb) and a closer view of 22kb (97,273-97,294K) that includes a differentially methylated region (red box). The top track shows the regions that are annotated as repetitive sequences using the MIPS/Recat repeat catalog. The next two tracks show the B73 and Mo17 methylation levels for each of the probes within these regions. The bottom tracks illustrate the gene models. The gray arrows indicate gene models that were rejected from the FGS. (C) Shows a chromosomal view of methylation levels in B73 and Mo17. The percentage of methylation is plotted as a 5Mb window sliding 1Mb downstream across the chromosome Blue and red lines indicate percent methylation of all probes for B73 and Mo17 respectively. The green line indicates centromere position. The black line shows the cm/Mb across the chromosome. The first heatmap provides a visualization of gene density. Yellow and black values indicate lower and higher relative gene density values respectively. The second heatmap provides visualization for the genomic structural variation between B73 and Mo17 using Comparative Genomic Hybridization (CGH) values . A 5Mb sliding window across the chromosome indicates regions of high diversity (black) to low diversity (yellow). The differential methylation regions (DMRs) are shown using red (Mo17 hypermethylation) and blue (B73 hypermethylation) arrows. The next track shows the location of individual probes that have significant ($q<0.001$) methylation variation between B73 and Mo17. The final heat map indicates the relative enrichment for differentially methylation probes across the chromosome with enriched regions indicated by red and regions with depleted levels of methylation variation in green.

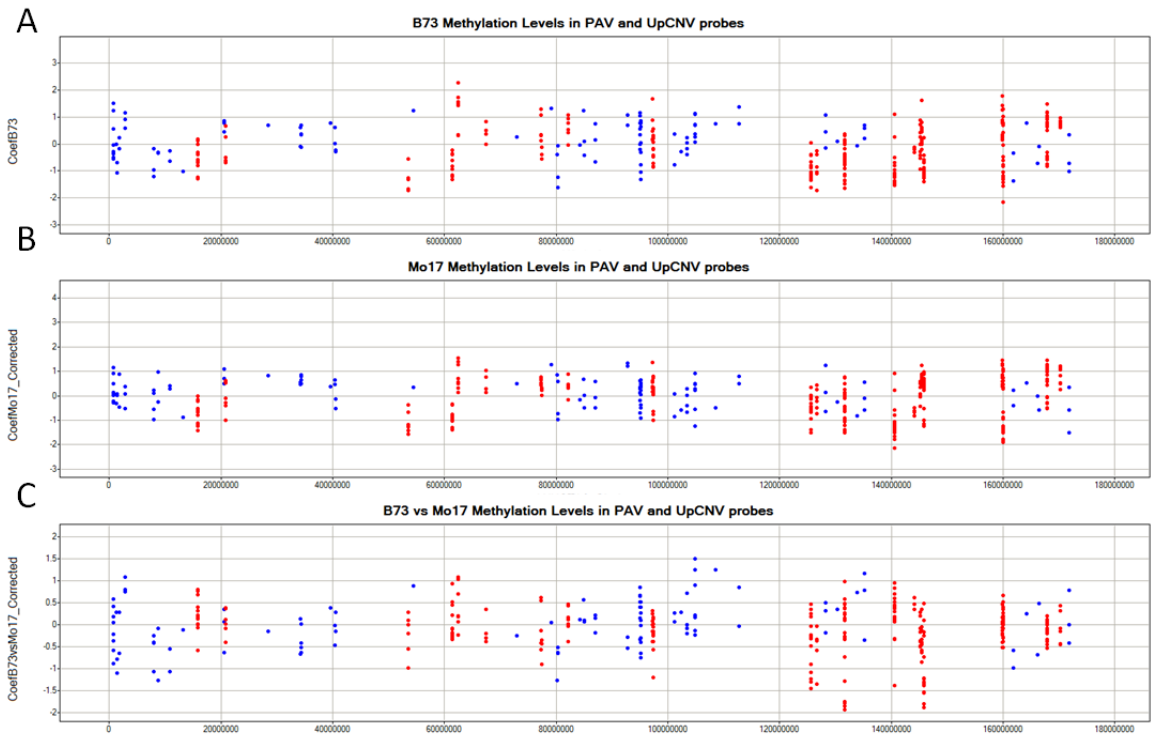



Figure 3: Examples of UpCNV and PAV probes showing both high and low levels of DNA methylation in B73 (A) and Mo17 (B). Regions of decreased and increased methylation levels for PAV (B>M, Blue) and UpCNV (M>B, Red) loci are present throughout chromosome 8. Variable methylation of PAV and UpCNV also occur throughout the chromosome (C).

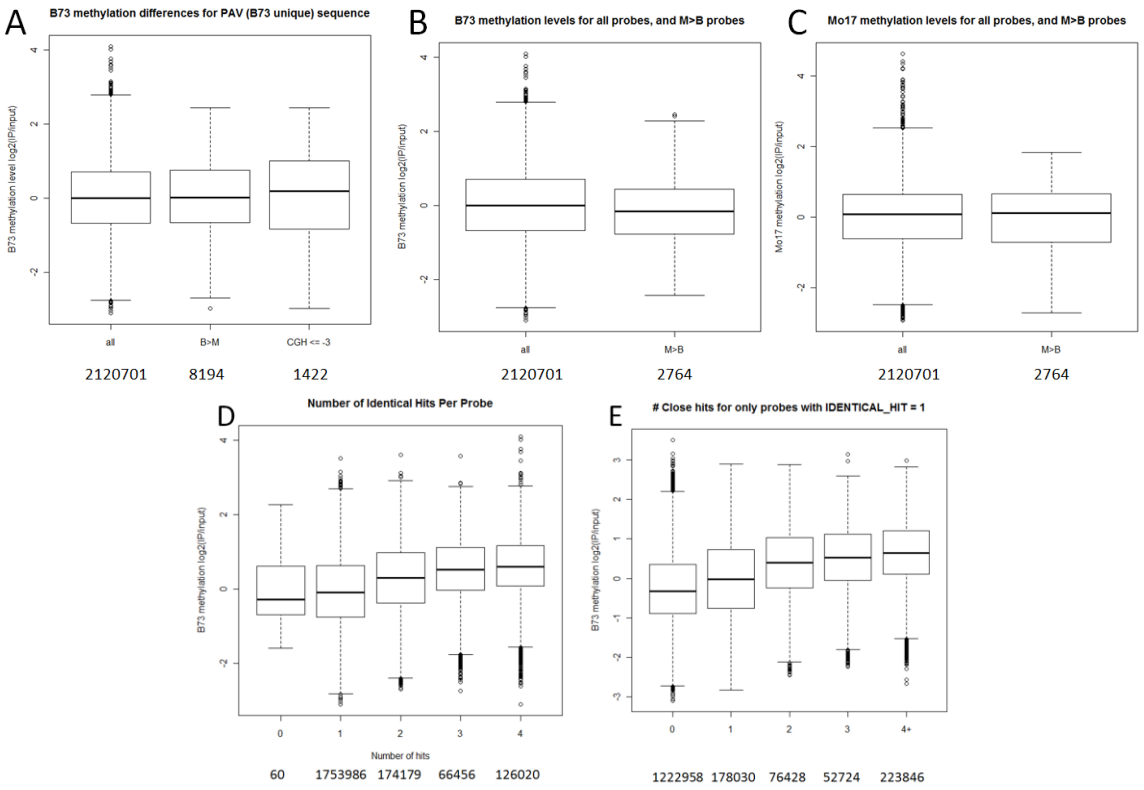



Figure 4: Copy number and genomic structural variation effects on methylation levels. (A) The distribution of B73 methylation values is shown for all probes as well as for probes in B>M segments and a subset of B>M segments that likely represent PAV sequences as the Mo17 signal is substantially lower than the B73 signal. There are no significant differences in the average methylation levels of these probes. In (B) and (C) the methylation of M>B probes is shown for B73 and Mo17, respectively. These likely represent sequences with copy number gains in Mo17 relative to Mo17 but there is not a substantial difference in the methylation of these sequences relative to other genomic sequences. (D) A boxplot is used to show the distribution of B73 methylation values for all probes with 1, 2, 3, or 4+ copies in the B73 genome. The methylation level significantly increases as the number of perfect matches increases. (E) A similar plot is used to show how the number of close (>90% identity and coverage) matches is similarly related to increased methylation values.

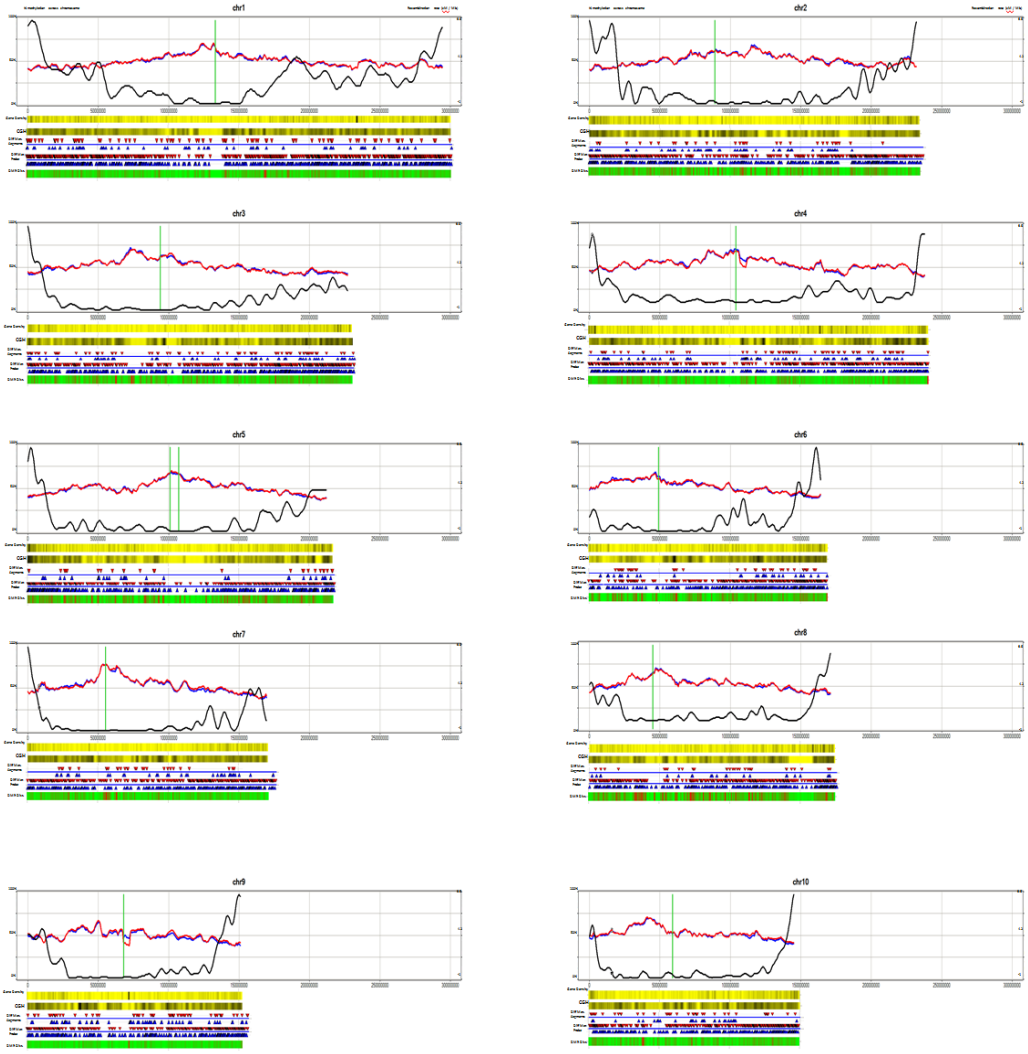


Figure 5: Percent methylation across maize chromosomesThe percentage of methylation is plotted as a 5Mb window sliding 1Mb downstream across the chromosome Blue and red lines indicate percent methylation of all probes for B73 and Mo17 respectively. The green line indicates centromere position. The black line shows the cm/Mb across the chromosome. The first heatmap provides a visualization of gene density. Yellow and black values indicate lower and higher relative gene density values respectively. The second heatmap provides visualization for the genomic structural variation between B73 and Mo17 using Comparative Genomic Hybridization (CGH) values. A 5Mb sliding window across the chromosome indicates regions of high diversity (black) to low diversity (yellow). The differential methylation regions (DMRs) are shown using red (Mo17 hypermethylation) and blue (B73 hypermethylation) arrows. The next track shows the location of individual probes that have significant ($q<0.001$) methylation variation between B73 and Mo17. The final heat map indicates the relative enrichment for differentially methylation probes across the chromosome with enriched regions indicated by red and regions with depleted levels of methylation variation in green.

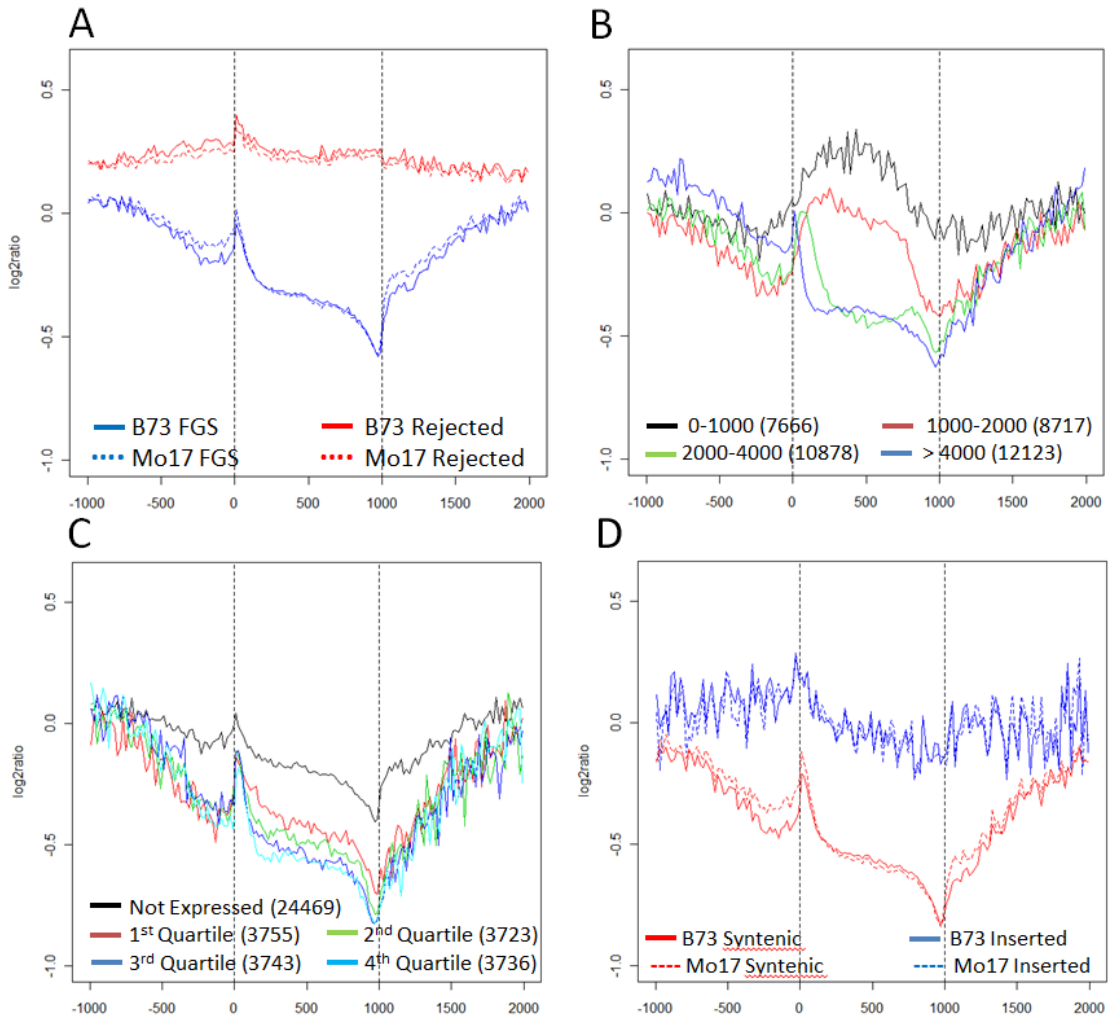
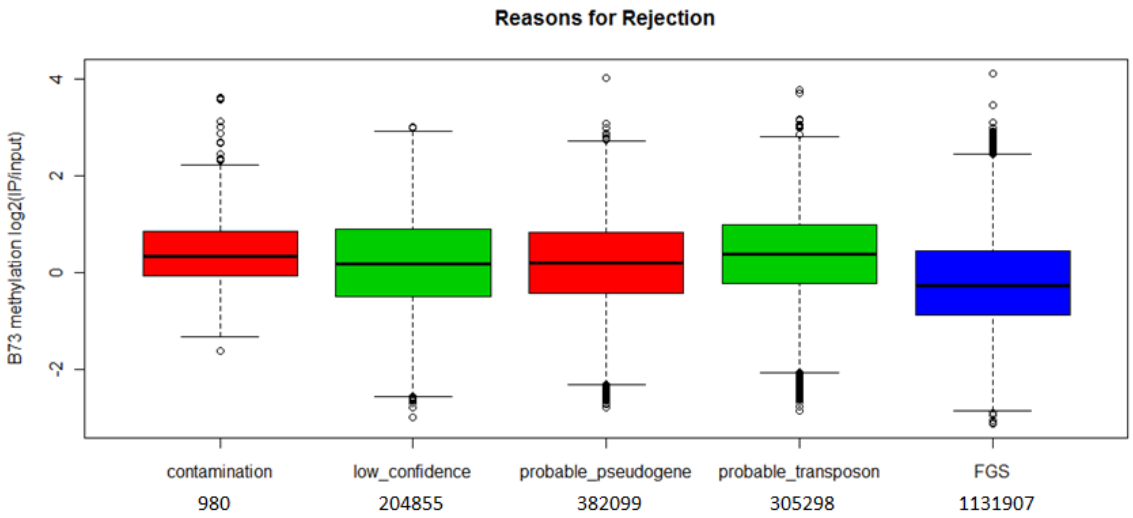



Figure 6: Gene body methylation and expression levels in maize genes. (A) The relative methylation levels (log2(IP/input)) was assessed for probes within 1000bp of the transcription start and termination site. The relative position for probes within the gene was normalized to a scale of 1000. The genes in the filtered gene set (FGS) exhibit a lower methylation and a more dynamic pattern across the length of the gene than the rejected genes. The vertical dashed lines indicated the beginning and end of transcription for each gene. (B) The FGS genes were divided into different length categories to assess the level distribution of gene body methylation. (C) The relative expression level for all FGS genes was assessed using published RNA-seq data from B73 leaf tissue (Li et al., 2010) and genes were assigned as not expressed or quartile 1-4 based on their expression level. In general, the genes show similar patterns of methylation but the higher expressed genes exhibit lower levels of methylation within and around the gene. (D) Each of the FGS genes was also classified according to whether it was located in a syntenic position relative to sorghum and/or rice or in a non-syntenic position. The syntenic genes exhibit much lower levels of methylation than the non-syntenic genes. This difference between syntenic and non-syntenic genes can also be seen in the regions immediately surrounding the gene.




```
Groups, Treatments and means
a         contamination            0.3995816
a         probable_transposon      0.3667027
 b        probable_pseudogene      0.2013132
 b        low_confidence           0.1919736
  c       FGS                      -0.1810878
```

Figure 7: Increased methylation at rejected genes. The genes in the working set that were rejected from the Filtered Gene Set (FGS) include possible contamination (bacterial sequences), low confidence FGENESH models, probable transposons and probable pseudogenes. Significance of classification was determined by one way ANOVA and group mean separation by Tukey's HSD. Groups of factors are provided. Genes in each of these rejection categories exhibit significantly higher methylation levels than genes in the uniquely-grouped FGS.

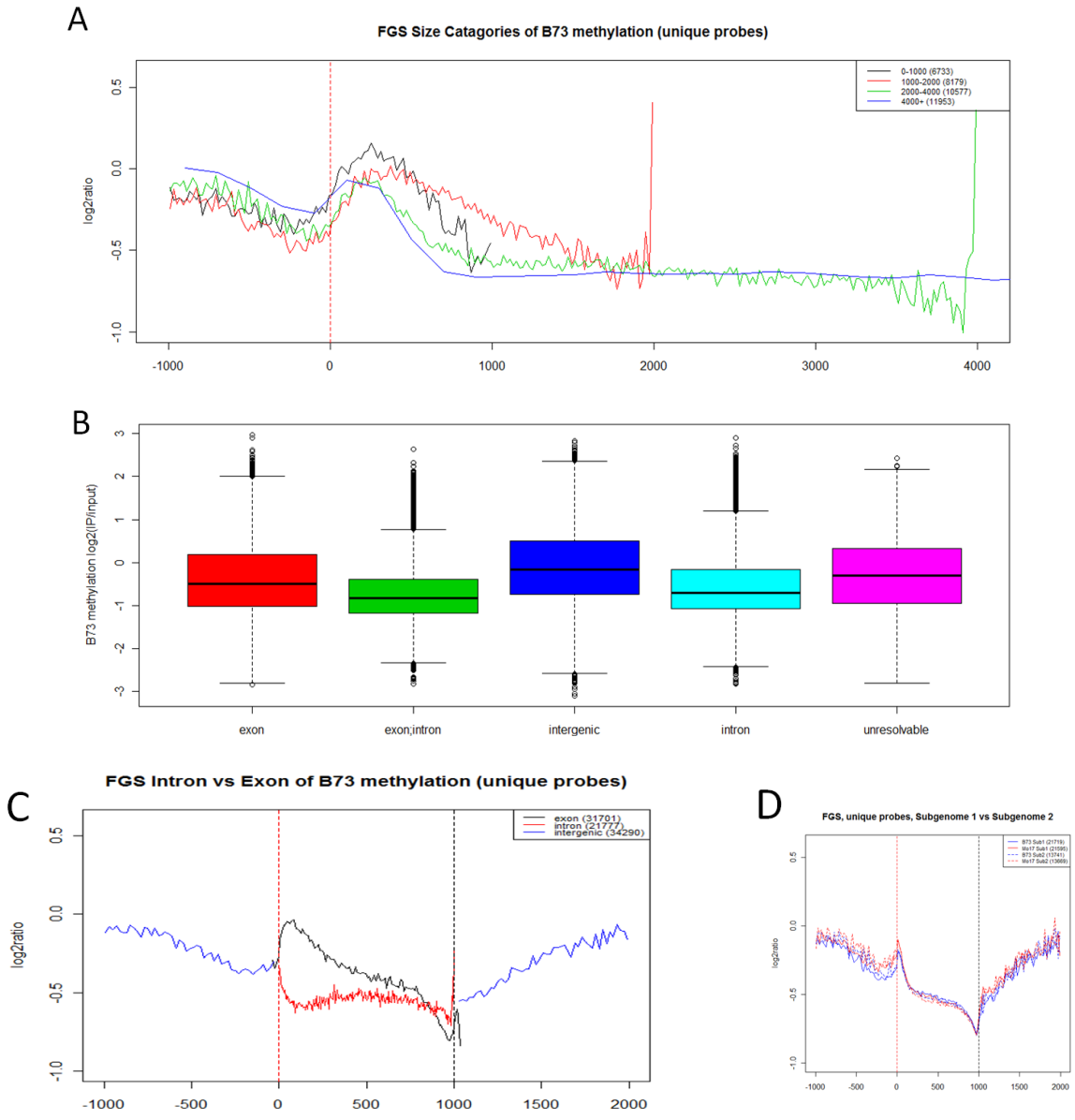



Figure 8: High levels of methylation in first ~700bp of gene body. (A) The average level of methylation was assessed for genes in various length classes in the first 4000bp of the gene using the actual distance from the TSS rather than the relative distance. Genes exhibit substantial gene body methylation in the first 500-700 bp. (B) Methylation levels are higher in intergenic sequences than in exons and introns. The lowest levels of methylation are observed in introns and at exon/intron boundaries. (C) A profile of the methylation patterns along genes for only exon (black) or intron (red) shows that gene body methylation at the 5' end of genes is confined to exons. Similarly, the reduced methylation at the 3' end of genes is more pronounced in exons than in introns. (D) Methylation levels are not affected by sub-genome 1 and 2. The FGS genes were all classified based on whether they were located in regions of the maize genome classified as sub-genome 1 or sub-genome 2 (Based on Schnable et al., 2011). There is no evidence for altered methylation levels for genes in sub-genome 1 relative to sub-genome 2.

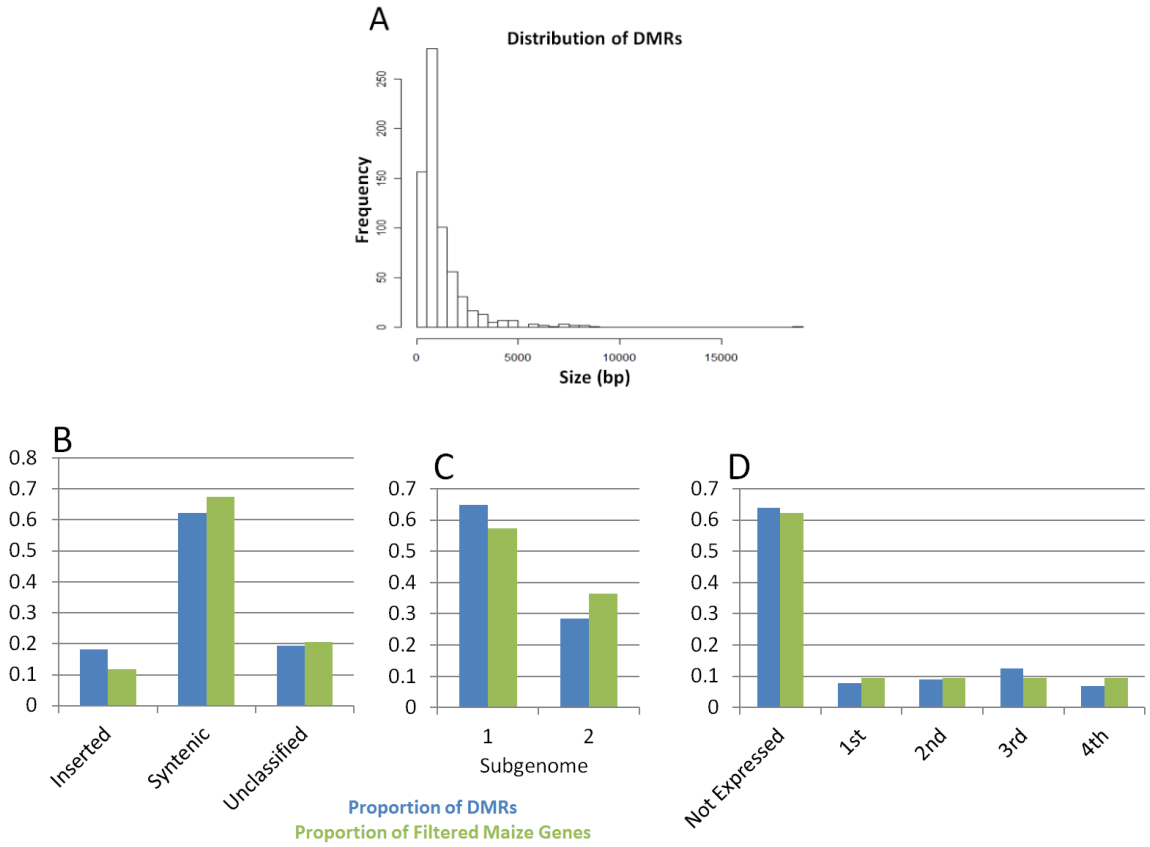



Figure 9: Characterization of maize DMRs. (A) A histogram is used to show the distribution of the length of the DMRs identified between B73 and Mo17. The majority of DMRs are <1000bp. (B) The DMRs were analyzed to assess enrichments for syntenic positioning. The proportion of DMRs (blue) that are within annotated syntenic or non-syntenic (inserted) positions in the genome was developed. This proportion is compared to the proportion of maize filtered genes (green) that are annotated as syntenic or non-syntenic (inserted). A slight enrichment for DMRs with inserted positions was observed. (C) The maize sub-genome classification (Schnable et al., 2011) was used to compare the proportion of DMRs within each class (blue) compared to the proportion of filtered genes (green). A slight enrichment of DMRs present in sub-genome 1 was observed. (D) For each DMR, the nearest gene within 5kb was identified as a possibly regulated by DMR state. The proportion of genes nearby DMRs (blue) and all maize filtered genes (green) across expression quartiles from maize leaf RNA-seq was calculated. The majority of genes are not expressed for both genes nearby DMRs and the filtered gene set as a whole. No enrichment for genes near DMRs was observed for any of the classes.

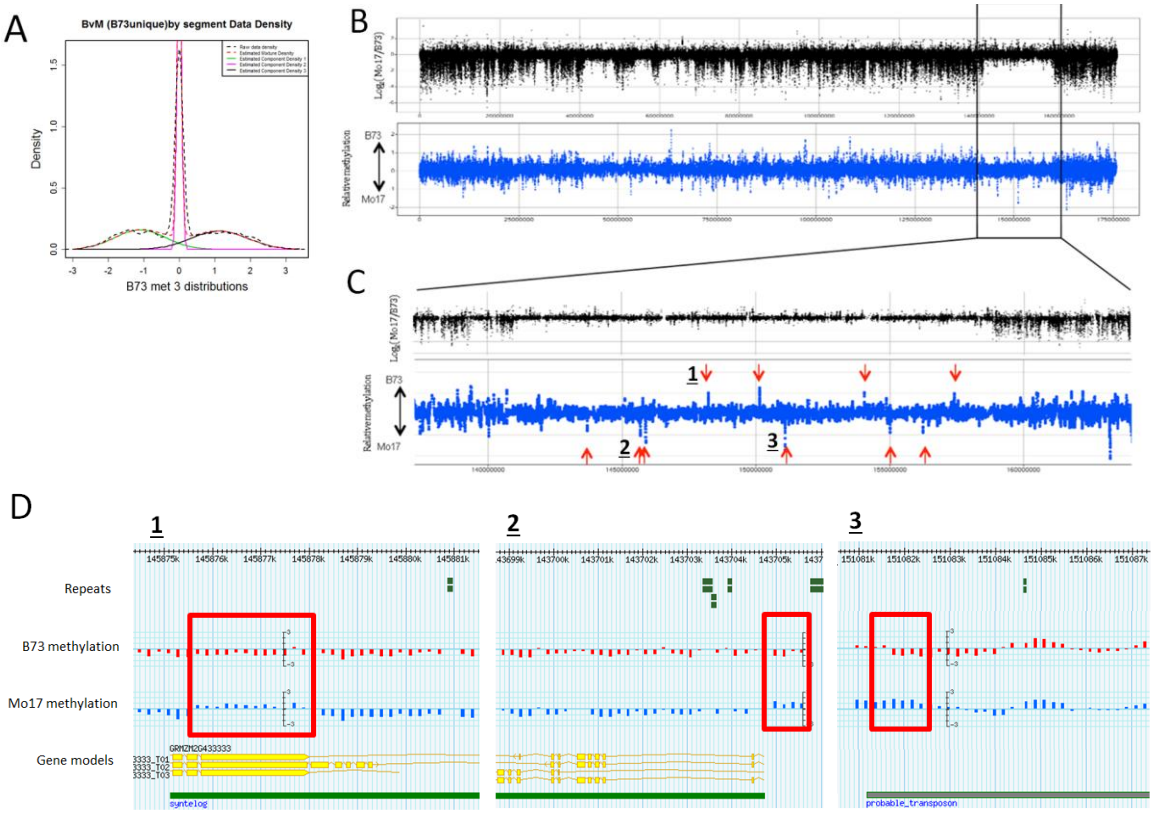



Figure 10: DMRs in B73 and Mo17. (A) The relative methylation levels for all probes were used to perform DNAcopy segmentation followed by expectation maximization. The black dashed lines show the observed distribution of the segment means. This distribution can be approximated (red dashed line) by a model that is derived from three normal distributions including B73 hypermethylation (right peak), Mo17 hypermethylation (left peak) and unchanged regions (middle peak). (B) The structural variation across chromosome 8 is shown in the plot with black spots. The blue spots show relative methylation in B73 and Mo17. The region of low structural diversity (boxed region) is magnified in (C). (D) Gbrowse views for three DMRs located within this region are shown. The data tracks show the position of repeats, genes, B73 methylation, Mo17 methylation and relative methylation. Each bar showing methylation values represents an individual probe. The actual DMRs are shown by the red boxes.

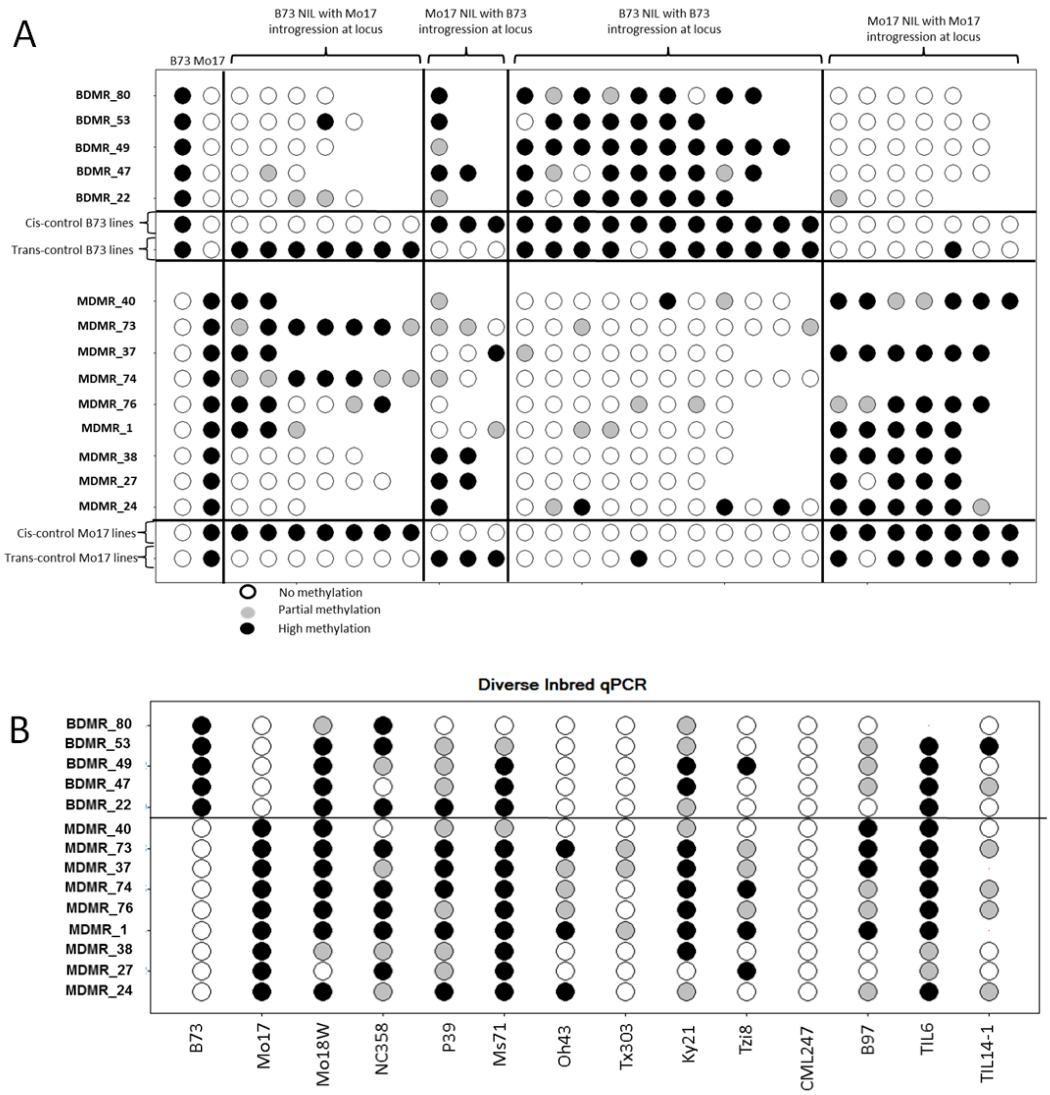



Figure 11: Variable DNA methylation patterns in near-isogenic lines and diverse inbreds. (A) The relative DNA methylation levels in selected near-isogenic lines was tested by digestion with the methylation dependent restriction enzyme MspjI followed by qPCR. Different subsets of NILs were selected and analyzed for each of 14 DMRs. The first two columns show the data from B73 and Mo17. Open circles reflect low methylation levels and black circles indicate high methylation levels. Intermediate methylation levels are indicated by gray color. The next group of 2-7 genotypes show the data from NILs that have B73 as the recurrent parent (>95% of the genome) and have introgression of the Mo17 haplotype in the region containing the DMR. The variable number of genotypes tested reflects the fact that some DMR loci are have more NILs with an introgression than others. The next group of 1-3 genotypes are NILs that are primarily Mo17 but have B73 introgressed at the DMR. The next two groups provide "control" genotypes of B73-like or Mo17-like NILs that do not have an introgression at the DMR. The expected patterns for *cis* (local) inheritance of DNA methylation or *trans* (unlinked) control of DNA methylation or shown. (B) The same type of assays were performed on a panel of 12 diverse inbred genotypes, including two inbred teosinte lines, to monitor the frequency for the hyper- and hypo-methylated states.

Chapter III - **CONTEXT STATEMENT**

Epigenetic variation has been proposed to be heritable through both meiotic and mitotic cellular divisions. DNA methylation has been hypothesized to play roles both in constitutive heterochromatic silencing of genomic regions that is constant throughout development and generations as well as in facultative heterochromatin that would vary during development. There is limited evidence for a specific role of DNA methylation in the regulation of proper development by acting as a reversible regulator of gene expression. In this chapter I study the relative level of variation for DNA methylation among different tissues and different genotypes in order to document the potential role of DNA methylation in development and natural variation. There is little evidence for DNA methylation variation across diverse developmental tissues found within maize and the few examples that were found do not appear to be associated with tissue-specific gene expression. In contrast, DNA methylation variation that occurs between genotypes is stable across all tissues assessed within a single genotype. This work highlights that DNA methylation states appear to be stably inherited through mitotic cellular divisions and there is little evidence that DNA methylation playing a prominent role in proper maize development.

Chapter III entitled 'Variation in DNA methylation patterns is more common among maize inbreds than among tissues' has been adapted from my work in the publication:

Eichten SR, Vaughn M, Hermanson P, Springer NM (2012) Variation In DNA Methylation Patterns is More Common among Maize Inbreds than among Tissues. The Plant Genome. doi: 10.3835/plantgenome2012.06.0009

In the development of this manuscript, other researchers have played roles in the creation of data and assistance to my analysis. The experimental design and microarray experiments were performed by SRE and PJH. MVassisted SRE in the developing methods for proper normalization of microarray data to facilitate analysis. Text and figures of this chapter were developed by SRE with assistance from NS.

Many of the journal specific sections as well as supplemental data tables have been omitted from this chapter. This includes the author list and abbreviations list. Figures have been renumbered as necessary to meet thesis guidelines. All omitted data is available through the initial scientific publication.

# Chapter III

# Variation in DNA methylation patterns is more common among maize inbreds than among tissues

Chromatin modifications, such as DNA methylation, can provide heritable, epigenetic regulation of gene expression in the absence of genetic changes. A role for DNA methylation in meiotically stable marking of repetitive elements and other sequences has been demonstrated in plants. Methylation of DNA is also is also proposed to play a role in development through providing a mitotic memory of gene expression states established during cellular differentiation. We sought to clarify the relative levels of DNA methylation variation among different genotypes and tissues in maize (*Zea mays L*.). We have assessed genome-wide DNA methylation patterns in leaf, immature tassel, embryo, and endosperm tissues of two inbred maize lines: B73 and Mo17. There are hundreds of regions of differential methylation present between the two genotypes. In general, the same regions exhibit differential methylation between B73 and Mo17 in each of the tissues that were surveyed. In contrast, there are few examples of tissue-specific DNA methylation variation. Only a subset of regions with tissue-specific variation in DNA methylation show similar patterns in both genotypes of maize and even fewer are associated with altered gene expression levels among the tissues. Our data indicates a limited impact of DNA methylation on developmental gene regulation within maize.

**INTRODUCTION**

Epigenetic variation can result in altered gene expression or phenotype without requiring changes in DNA sequence. Epigenetic information can provide gene regulation during development and differentiation of cells to reinforce the "memory" of transcriptional states. In some cases, epigenetic changes can also be heritable and result in transgenerational "memory." Epigenetic marks within the genome are often encoded through chromatin modifications including DNA methylation and histone modifications. Different types of chromatin modifications can have varying functional consequences and there is also evidence that different types of modifications may have varying roles in meiotic and mitotic epigenetic

memory. Herein we are focused on the role of DNA methylation in providing epigenetic memory during development (mitotic) and among individuals of different genetic backgrounds (meiotic).

There is fairly strong evidence that DNA methylation patterns can vary among individuals of the same species. There are also many examples of epigenetic variation that generates epialleles resulting in phenotypic variation without sequence changes (Bender and Fink 1995; Jacobsen & Meyerowitz 1997; Cubas et al., 1999; Morgan et al., 1999; Chandler et al., 2000; Stokes et al., 2002; Rakyan et al., 2003; Suter et al., 2004; Manning et al., 2006). Genome-wide scans in Arabidopsis have found DNA methylation differences among ecotypes (Vaughn et al., 2007) and in spontaneous mutant accumulation lines (Becker et al., 2012; Schmitz et al., 2012). There are also examples of variable DNA methylation levels among maize inbreds (Eichten et al., 2011). The variation in DNA methylation patterns among individuals of the same species includes examples that are correlated with genetic changes and those that are not linked to DNA sequence change (Richards, 2006). The genetic changes that are correlated with DNA methylation variation often include transposon insertions or structural rearrangements (Richards 2006; Girard & Freeling, 1999; Feschotte 2008; Weil & Martienssen, 2008).

Several groups have also proposed that DNA methylation also plays an important role in direct regulation of genes during development (Bird 1997; Richards 1997). There is fairly strong evidence to demonstrate developmental variation in DNA methylation patterns that are correlated with developmental regulation of gene expression in humans (Lister et al., 2009; Hawkins et al., 2010; Franklin et al., 1996; Reik 2007; Skinner 2011; Ohtani & Dimmeler 2011). In addition, several specific cell types or tissues related to reproduction exhibit altered DNA methylation levels in plants (Lauria et al., 2005; Gehring et al., 2009; Hsieh et al., 2009; Zemach et al., 2010). It has been suggested that demethylation in the central cell and vegetative nuclei of the male gametes (Slotkin et al., 2009) is a mechanism to reinforce transposon silencing each generation (Kohler & Weinhofer-Molisch 2010). Although examples exist of the importance of DNA methylation for proper development in animals and in gametes or endosperm of plants, there is limited evidence that DNA methylation variation contributes to regulatory variation among plant vegetative tissues. The phenotypes observed in Arabidopsis MET1 loss-of-function lines include a variety of developmental abnormalities that may suggest a role of DNA methylation in development (Finnegan et al., 1996; Ronemus et al., 1996), though it wasn't clear

whether this was due to direct effects of altered DNA methylation patterns. Global DNA methylation levels may change in different vegetative tissues of tomato (Messeguer et al., 1991). One study suggested a difference in the methylation levels of photosynthesis-related genes during differentiating maize leaf cells (Ngernprasirtsiri et al., 1989). A recent study in *Arabidopsis* provides evidence that DNA methylation regulates the expression of WUSCHEL (Li et al., 2011).

However, there are also several pieces of evidence to suggest somewhat limited changes in DNA methylation during vegetative development. Genome-wide profiling of DNA methylation levels in different tissues of rice (Zemach et al., 2010) identified few examples of differences in DNA methylation levels among vegetative tissues compared to the differences observed between endosperm and other tissues. The analysis DNA methylation patterns in several different tissues of sorghum have shown that the majority of methylation patterns are similar across tissues; however a small number of tissue-specific DNA methylation regions (tDMRs) were identified that correlate with variable gene expression within their respective tissues (Zhang et al., 2011).

To gain insights into the role of DNA methylation in providing epigenetic memory during development or among individuals, we profiled DNA methylation patterns in two inbred genotypes of maize in four distinct tissues: 14-day leaf, endosperm, tassel, and embryo tissues. Although these complex tissues contain a variety of cell types, we are interested in broad DNA methylation variation across these diverse tissues. Maize (*Zea mays*) provides a unique and robust system to study the role of epigenetic modifications due to the complex organization of interspersed transposons and genes in the maize genome and due to substantial genetic resources (Yu et al., 2008; McMullen et al., 2009; Schnable et al., 2009). Genome-wide profiling of methylation patterns between these four tissues show that there are numerous methylation differences between the two genotypes that are observed in all tissues studied. However, there are relatively few differences in DNA methylation levels among tissues. This study shows that few tissue-specific methylation events occur within maize and suggests that the impact of DNA methylation on proper developmental regulation is limited.

**RESULTS**

The methylation state of 14-day seedling leaf, tassel, 14 DAP embryo, and 14 DAP endosperm DNA from B73 and Mo17 inbred lines was assessed by methylated DNA immunoprecipitation (meDIP) followed by hybridization to a microarray platform (see Methods). The array platform includes long oligonucleotide probes placed ~200bp apart on low-copy sequences. The analysis of the signal ratio between imunoprecipitated DNA and a non-enriched control allows for the assessment of methylation enrichment across the maize genome (B73 RefGen_v2). To prevent aberrant signal from multi-copy sequences as well as sequences absent from the Mo17 genome, all data was filtered to probes with only one unique copy in the B73 RefGen_v2 assembly and probes known to hybridize with similar efficiency in both B73 and Mo17 inbred lines (Eichten et al., 2011). The resulting 1,088,820 single-copy probes were used for a linear model based analysis that allowed for the isolation of genotype and DNA methylation effects. A linear model was developed to estimate average probe ratios for 12 tissue comparisons and four genotype comparisons (Table 1).

One approach to assessing the variation in DNA methylation among tissues and genotypes is to perform hierarchical clustering. Clustering of hypothetical SNP data (Figure 1A) or actual transcriptome (Figure 1B) data across the four complex tissues illustrates the expectation for a genetic variation (such as SNPs) or primarily among tissues (as expected for transcriptomes). The clustering of DNA methylation data (Figure 1C) reveals more substantial effects from genotype rather than tissue. Although most variation occurs between genotypes, there are some examples where tissue-specific methylation signals appear. The hierarchical clustering reveals that DNA methylation changes during development are limited relative to the differences between genotypes. We proceeded to assess common differences among genotypes throughout development and subsequently to search for the rare examples of tissue-specific methylation variation within the sampled tissues.

***Numerous examples of conserved methylation differences found between inbred lines***

There is wide-spread interest in understanding how DNA methylation patterns vary among individuals of the same species. We assessed whether different developmental stages would exhibit similar differences between two maize inbreds, B73 and Mo17. Differentially methylated regions between the two genotypes (gDMRs) were identified in each tissue by using DNAcopy to

identify multiple adjacent probes with significant variation between the genotypes (Figure 2A). Over 500 gDMRs with at least a two-fold methylation difference were found between B73 and Mo17 in each of the four tissues that were assessed. There are similar proportions of gDMRs with higher methylation each of the two genotypes; 49% have higher methylation in B73 compared to Mo17 (Figure 2A). While the number of gDMRs between B73 and Mo17 in the each tissue was different we found that many of the regions had very similar differential methylation in all tissues profiled (Figure 2B). The variation in actual number of gDMRs identified in each tissue is due to gDMRs that are near the statistical cut-offs in some tissues but pass all filters in other tissues. These results indicate that there are hundreds of gDMRs showing similar methylation variation between B73 and Mo17 in leaf, embryo, tassel, and endosperm tissue (Figure 2B).

***Few tissue-specific DNA methylation variants***

It has been postulated that DNA methylation may play a role in proper regulation of development (Bird 1997; Richards 1997). To examine the impact of DNA methylation on development, an analysis of tissue-specific DNA methylation was performed in four developmentally unique tissues. Six tissue contrasts were developed between the four tissues (seedling, embryo, tassel, and endosperm) for both B73 and Mo17 tissues. Using the same criteria, substantially fewer tissue-specific differentially methylated regions (tDMRs) were discovered in tissue contrasts compared to genotype contrasts (Figure 2C-E). There were four-times fewer tDMRs than gDMRs. Genotype DMRs often exhibit larger differences in methylation levels (median segment difference = 1.575) than tissue DMRs (median = 1.310) (Figure 3). A large portion of the DMRs (70% in B73 and 81% in Mo17) were identified in contrasts of endosperm with the other tissues.

Although there are significantly fewer tDMRs present between maize tissues, it is possible that some tDMRs play important roles in developmental regulation. Several filtering criteria were utilized to identify tissue-specific DMRs with potential functional roles in plant development. First, any tDMR that impacts development would be expected to show similar patterns in both genotypes if it plays a major role in proper development. Second, tDMRs involved in developmental gene regulation should be located near a gene that shows developmental variation for expression levels that is correlated with DNA methylation state.

We assessed whether tDMRs that were identified in a contrast of two tissues in one genotype had a conserved pattern in the same tissues of the other genotype. Only 115 (19%) of all 600 tDMRs showed similar DNA methylation patterns in both B73 and Mo17 (Figure 4A) suggesting that the majority of tissue-specific DMRs appear to be inbred specific and are most likely not involved in important developmental regulation. The 115 examples of tissue-specific DMRs that are conserved in B73 and Mo17 include 12 examples of regions that were identified in multiple tissue contrasts resulting in a set of 103 unique tDMRs. The DNA methylation state in all four tissues was assessed for each of the 103 conserved tDMRs using hierarchical clustering (Figure 2F). For both genotypes, the endosperm was most distinct relative to the other tissues. Nearly 70% of the 103 conserved tDMRs were identified in a contrast of endosperm with one of the vegetative tissues. There are smaller distinct clusters of genes for which leaf and tassel tissues show similar methylation levels compared to endosperm and embryo. Overall, there are many examples of tissue DMRs that appear to show a unique DNA methylation level in a single tissue, while showing a uniform pattern in the remaining tissues (Figure 2E, Figure 5).

***Expression variation of genes near tissue-specific DMRs***

The 103 DMRs among maize tissues that were consistent in both B73 and Mo17 were further characterized to assess whether they were associated with tissue-specific expression of nearby genes. RNA-seq was performed to assess gene transcription levels in the same tissues that were used for DNA methylation profiling (6-17 million reads per sample). There are 57 genes that were located within 5kb of the tDMRs and are expressed in at least one of the tissues (RPKM > 0.1). The expression levels for 34 of these 57 genes near the tissue-specific DMRs exhibit a 2-fold or greater change in expression when comparing the two tissues used to discover the nearby tDMR. These 34 genes include 14 examples in which DNA methylation was negatively correlated with transcript abundance, as expected, and another 20 examples in which the transcript abundance and DNA methylation were positively correlated. Methyl-dependent restriction enzyme digests followed by qPCR were used to provide a validation for tissue-specific variation in DNA methylation levels for 18 tDMRs (Figure 4D). 11 of 18 tDMRs tested selected from the 34 candidates were validated between the tissues used in the discovery of the tDMR.

A comparison of DNA methylation and transcript abundance in all tissues and genotypes identified six genes with a significant ($p<0.05$) negative correlation between DNA methylation and expression state and four of the genes had an unexpected positive correlation between

DNA methylation and expression (Figure 4). For example, tDMR55 exhibits lower DNA methylation levels in tassel and is more highly expressed in this tissue (Figure 4C). tDMR51 provides an example in which higher DNA methylation levels are correlated with higher expression. It is worth noting that many genes show variable tissue-specific expression patterns and we might expect some number of false-positives when assessing expression patterns for any 103 random genes simply due to the frequency of tissue-specific variation. Several of the 10 correlated genes are similar ($e<0.01$) to an Arabidopsis gene (data not shown). The genes with a negative correlation between DNA methylation and expression include a MYB-domain gene similar to ARR18. There was a putative histone acetylase with a positive correlation between DNA methylation and gene expression.

## DISCUSSION

Epigenetic memory likely occurs during both mitosis and meiosis. Our experiment allows us to examine the level of variation for DNA methylation, a chromatin modification often associated with epigenetic memory. We find more evidence for variation in DNA methylation patterns among genotypes than among developmental stages (Figure 1). This suggests that DNA methylation may play a more substantial role in differentiating individuals of a species than in "memory" of developmental expression differences. Although a large number of gene expression changes occur as tissues develop within a plant, there is limited evidence for localized DNA methylation variation to play a role in this developmental process.

### *Genotype-specific DMRs are highly conserved across tissues*

There is a growing effort to characterize DNA methylation differences among different individuals of the same species. Researchers are often confronted with decisions about the experimental design and are concerned that developmental variation among two individuals may complicate the comparisons. Our data suggest that DNA methylation patterns are strikingly similar regardless of tissue assessed (Figure 6). In general, if we observe differences in DNA methylation among B73 and Mo17 in one tissue we saw similar variation in all other tissues for 84% of the genotype DMRs. In a previous report we had identified 690 gDMRs among leaf tissue of B73 and Mo17. The combined analysis of all four tissues revealed a total of ~850 non-redundant gDMRs that meet all statistical criteria in at least one tissue, and show conserved

patterns of genotypic variation in all other tissues. Our data suggest that very similar DMRs would be found in any particular tissue but that the analysis of multiple tissues may allow for the most robust and complete set of genotype-specific DMRs. However, if necessary, it is possible to compare DNA methylation profiles of two individuals that vary in developmental morphology (such as wild and domesticated individuals) and still identify differences in DNA methylation due to genotype differences. Although the exact function of gDMRs among maize genotypes is still to be determined, the hundreds of regional differences in DNA methylation between B73 and Mo17 may provide a source of regulation explaining a portion of the differences between these two lines.

***Few tissue-specific DMRs indicate limited role of DNA methylation and tissue development***

There has been evidence both for (Finnegan et al., 1996; Ronemus et al., 1996), and against (Zemach et al., 2010; Zhang et al., 2011), a broad role of DNA methylation in regulating gene expression during development in plants. In order for a methylation variant to play an important, predictable role in development, it would be expected that multiple different genotypes of maize would show similar tissue-specific DMRs. We found a limited number of tDMRs among the vegetative tissues and only a small number of these exhibit tDMRs in both genotypes. It would be expected that functionally relevant tDMRs would contribute to regulation of nearby genes. We found only 10 genes among the 57 located near tDMRs for which expression was correlated to the tDMRs methylation state (Figure 4). These include six examples of negative correlation between DNA methylation and expression and four examples of positive correlation. We would expect that DNA methylation would show a negative correlation to gene expression as often methylation is associated with nearby gene silencing. The observed positive correlation of four tDMRs was surprising. It is possible that DNA methylation can be positively associated with gene expression. Studies of imprinted genes in Arabidopsis have identified examples where demethylation of certain imprinted genes is required for the gene to be silenced (Hsieh et al., 2011). From this, there may be certain genes within the genome that are not transcriptionally silenced unless demethylated. Alternatively, the existence of similar numbers of genes with positive and negative correlations with DNA methylation levels could suggest a lack of causation. Many genes exhibit tissue specific variation and we might expect some examples of correlated patterns in any set of 103 genes chosen by chance.

This study has provided evidence that DNA methylation does not show significant variation between tissues. Similar analyses of methylation across tissues within plant systems show minimal tissue-specific methylation variation indicating a minimal influence of DNA methylation perturbation on development (Zhang et al., 2011). However, it is possible that tissue-specific DNA methylation variations are present that are beyond the scope of this study. By utilizing an array-based technology to assess DNA methylation, we are unable to identify methylation variants smaller than 200bp. Because of this, any tDMR smaller than our fragment size limits may go undetected in this study. It is also possible that tDMRs may go undetected given the strict filtering criteria used in this study. For example, requiring a two-fold difference in DNA methylation levels would exclude the discovery of many of the allele-specific DNA methylation differences observed in endosperm. Beyond this, tissues selected in this study contain a variety of cell types that may display altered DNA methylation patterns, but could be masked due to sampling multiple cell types at once within each tissue. Many development studies have shown that an increased resolution of tissue development allowed for a much greater capacity to identify tissue-specific differences is gene expression (Li et al., 2010). The same may indeed hold true for epigenetic studies.

**CONCLUSIONS**

Much debate has centered on the role of epigenetic regulation in tissue specification (Richards 1997). Given the complexities of cellular and gene regulation required to correctly develop into a terminal tissue, it is possible that reversible epigenetic regulation of genes could provide a mechanism for gene expression variation during cellular differentiation. To invesTrends in Geneticsate this possibility, we performed genome-wide DNA methylation assessment across four distinct tissues of maize in two separate inbred lines. Although many conserved methylation variants could be identified across genotypes (gDMRs), very few tissue-specific DMRs (tDMRs) were identified. Tissue-specific DMRs initially identified exhibit poor agreement across genotypes as well as limited association with nearby gene expression variation. From these results, we conclude that there are very few, if any, tissue-specific methylation variants that provide developmental regulation of gene expression in maize.

**EXPERIMENTAL METHODS & PROCEDURES**

**Plant materials and DNA isolation:** Seedling leaf tissue was harvested and prepared as described in Eichten et al., 2011. For embryo, tassel, and endosperm collection, plants were grown to maturity in the University of Minnesota Agricultural Research station, Falcon Heights, Minnesota. Endosperm and Embryo were harvested from multiple B73 and Mo17 ears 14 days after self-pollination. Multiple endosperms and embryos were pooled from each ear on each date and frozen in liquid nitrogen. Immature 15-20cm tassel tissue was harvested from multiple B73 and Mo17 plants and frozen in liquid nitrogen. DNA was isolated using the CTAB method (Doyle 1987). 5-10ug of gDNA in 650-700uL nuclease-free water was sonicated for five, ten-second pulses as per the methods of Haun et al., 2008. Samples were quantified and run on 1.5% agarose gels to verify that DNAs were fragmented to 100-500bp.

**Array Design and annotation:** Two array platforms were used in this experiment. Endosperm, leaf, and embryo tissues were assayed using a NimbleGen 2.1M feature long oligonucleotide array as described in Eichten et al., 2011. The analyses in this manuscript focus on a set of ~1.4million probes from this array that are single copy in the B73 genome and that do not exhibit strong comparative genomic hybridization variation among B73 and Mo17. Tassel tissue was assayed using a NimbleGen 3x1.4M long oligonucleotide array containing the same subset of 1.4 million probes also found on the 2.1M platform. Partial array replication was performed as outlined in Supplementary Table 1.

**Immunoprecipitation of Methylated DNA, labeling and hybridization:** Methods were adapted from Eichten et al., 2011. Briefly, methylated DNA was immunoprecipitated with an anti-5-methylcytosine monoclonal antibody from 400ng sonicated DNA using the Methylated DNA IP Kit (Zymo Research, Orange, CA; Cat # D5101). For each replication and genotype, whole genome amplification was conducted on 50-100ng IP DNA and also 50-100ng of sonicated DNA (input control) using the Whole Genome Amplification kit (Sigma Aldrich, St. Louis, MO, Cat # WGA2-50RXN). For each amplified IP input sample, 3ug amplified DNA were labeled using the NimbleGen Dual-Color Labeling Kit (Cat # 05223547001) according to the manufacturer's protocol for methylation arrays (Roche NimbleGen Methylation UserGuide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA sample was labeled with Cy3. Samples were hybridized to the array for 16-20hrs at 42°. Slides were washed and scanned according to NimbleGen's protocol for the GenePix4000B or Nimblegen MS200 array scanner.

Images were aligned and quantified using NimbleScan software (Roche NimbleGen) producing raw data reports for each probe on the array.

**Normalization and linear modeling:** Pair files exported from NimbleScan were imported into the Bioconductor statistical environment (http://bioconductor.org/). Sample-dependent MeDIP enrichments were estimated for each probe by fitting a fixed linear model accounting for array, dye, and sample effects to the data using the limma package (Smyth, 2004). The following statistical contrasts were then fit: B73 seedling immunoprecipitation sample (IP) vs. B73 seedling genomic DNA control (input); B73 embryo IP vs. input; B73 endosperm IP vs. input; B73 tassel IP vs. input; Mo17 seedling IP vs. input; Mo17 embryo IP input; Mo17 endosperm IP vs. input; Mo17 tassel IP vs. input. Four between-genotype statistical contrasts were also fit: Mo17 seedling vs. B73 seedling; Mo17 embryo vs. B73 embryo; Mo17 endosperm vs. B73 endosperm; Mo17 tassel vs. B73 tassel. Finally, between-tissue statistical contrasts were developed for each genotype individually: Embryo vs. leaf; Embryo vs. Tassel; Endosperm vs. embryo; Endosperm vs. leaf; Endosperm vs. tassel; Leaf vs. tassel. Moderated t-statistics and the log-odds score for differential MeDIP enrichment was computed by empirical Bayes shrinkage of the standard errors with the False Discovery Rate controlled to 0.05. Results were formatted for the Integrative Genomics Viewer (IGV, Robinson et al., 2011) for downstream analysis.

**Analysis of variable methylation:** To identify segments showing differential methylation for all contrasts, the DNAcopy algorithm (Venkatraman and Olshen, 2007) was used on 1,088,517 Mo17 unique probes in the B73 vs. Mo17 relative methylation linear model results for all three tissues. Resulting segments were defined as differentially methylated regions (DMRs) if the segment mean of methylation values showed at least a two-fold change (<-1 or > 1 on the log2 scale). These defined segments were used to assess methylation states in other contrasts. Values were scaled for all 8 samples as (x – avg(y,z) / max(y,z,) – avg(y,z)) where x is the unscaled sample value and y,z are the unscaled values that initially used in the discovery of the DMR. To validate similar methylation states within other contrasts, the segment mean of methylation values must show a 25% change in relative methylation state.

**RNA-seq analysis:** RNA seq analysis of all tissues described above was performed. Three biological replicates of each sample DNA were prepared at the University of Minnesota BioMedical Genomics Center in accordance with the TruSeq library creation protocol (Illumina). Samples were sequenced on the HiSeq 2000 developing 6-17 million reads per replicate. Raw

reads were filtered to eliminate poor quality reads using CASAVA (Illumina). Transcript abundance was calculated by mapping reads to the maize reference genome (AGPv2) using TopHat (Trapnell et al., 2009). A high degree of correlation between replicates was observed ($r>0.98$). RPKM values were developed using 'BAM to Counts' across the exon space of the maize genome reference working gene set (ZmB73_5a) within the iPlant Discovery Environment (www.iplantcollaborative.org).

**qPCR**: qPCR validation of DMRs was adapted from Eichten et al., 2011. Briefly, primers were designed to amplify regions of 18 tDMR candidates. Sample DNA was digested with the methyl-sensitive restriction enzyme MspJI (New England Biolabs) along with glycerol mock digests. The difference between digest C(t) and mock C(t) was calculated for each sample tested. As our selected enzymes target methylated cytosines, higher methylation leads to increased digestion and subsequently longer C(t) times.

**REFERENCES**


Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245–249.

Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83: 725–734.

Bird A (1997) Does DNA methylation control transposition of selfish elements in the germline? Trends in Genetics13: 469–472.

Chandler VL, Eggleston WB, Dorweiler JE (2000) Paramutation in maize. Plant Molecular Biology 43: 121–145.

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157–161.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19.1: 11–15.

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, et al. (2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics 7: e1002372.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics 9: 397–405.

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proceedings of the National Academy of Sciences of the United States of America 93: 8449–8454.

Franklin GC, Adam GI, Ohlsson R (1996) Genomic imprinting and mammalian development. Placenta 17: 3–14.

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

Girard L, Freeling M (1999) Regulatory changes as a consequence of transposon insertion. Developmental Genetics 25: 291–296.

Haun WJ, Springer NM (2008) Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. The Plant Journal 56: 903–912.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491.

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451–1454.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. Science 277: 1100–1103.

Kohler C, Weinhofer-Molisch I (2010) Mechanisms and evolution of genomic imprinting in plants. Heredity 105: 57–63.

Lauria M, Rupe M, Guo M, Kranz E, Pirona R, et al. (2004) Extensive maternal DNA hypomethylation in the endosperm of Zea mays. The Plant Cell 16: 510–522.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, et al. (2010) The developmental dynamics of the maize leaf transcriptome. Nature Genetics 42: 1060–1067.

Li W, Liu H, Cheng ZJ, Su YH, Han HN, et al. (2011) DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. PLoS Genetics 7: e1002243.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948–952.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. Plant Molecular Biology 16: 753–770.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23: 314–318.

Ngernprasirtsiri J, Chollet R, Kobayashi H, Sugiyama T, Akazawa T (1989) DNA methylation and the differential expression of C4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. The Journal of Biological Chemistry 264: 8241–8248.

Ohtani K, Dimmeler S (2011) Epigenetic regulation of cardiovascular differentiation. Cardiovascular Research 90: 404–412.

Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, et al. (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proceedings of the National Academy of Sciences of the United States of America 100: 2538–2543.

Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425–432.

Richards EJ (1997) DNA methylation and plant development. Trends in Genetics13: 319–323.

Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews Genetics 7: 395–401.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nature Biotechnology 29: 24–26.

Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. Science 273: 654–657.

Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends in Plant Science 17: 149–154.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Skinner MK (2011) Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Research93: 51–55.

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, et al. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3: 3.

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in Arabidopsis disease resistance. Genes & Development 16: 171–182.

Suter CM, Martin DI, Ward RL (2004) Germline epimutation of MLH1 in individuals with multiple cancers. Nature Genetics 36: 497–501.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology 5: e174.

Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23: 657–663.

Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Current Opinion in Genetics & Development 18: 188–192.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539–551.

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, et al. (2010) Local DNA hypomethylation activates genes in rice endosperm. Proceedings of the National Academy of Sciences of the United States of America 107: 18729–18734.

Zhang M, Xu C, von Wettstein D, Liu B (2011) Tissue-specific differences in cytosine methylation and their association with differential gene expression in sorghum. Plant Physiology 156: 1955–1966.

Table 1: *de novo* discovery of DMRs

| Contrast Type | Contrast | First member of contrast high methylation | First member of contrast low methylation | Total | Total DMRs in contrast type |
|---|---|---|---|---|---|
| Genotype | Endosperm Mo17 vs. B73 | 159 | 344 | 503 | 2384 |
| | Embryo Mo17 vs. B73 | 315 | 262 | 577 | |
| | Leaf Mo17 vs. B73 | 378 | 358 | 736 | |
| | Tassel Mo17 vs. B73 | 362 | 206 | 568 | |
| Tissue | B73 embryo vs. leaf | 16 | 0 | 16 | 600 |
| | B73 embryo vs. tassel | 27 | 51 | 78 | |
| | B73 endosperm vs. embryo | 86 | 59 | 145 | |
| | B73 endosperm vs. leaf | 42 | 27 | 69 | |
| | B73 endosperm vs. tassel | 33 | 63 | 96 | |
| | B73 leaf vs. tassel | 13 | 28 | 41 | |
| | Mo17 embryo vs. leaf | 17 | 4 | 21 | |
| | Mo17 embryo vs. tassel | 2 | 1 | 3 | |
| | Mo17 endosperm vs. embryo | 5 | 40 | 45 | |
| | Mo17 endosperm vs. leaf | 44 | 35 | 79 | |
| | Mo17 endosperm vs. tassel | 2 | 0 | 2 | |
| | Mo17 leaf vs. tassel | 3 | 2 | 5 | |

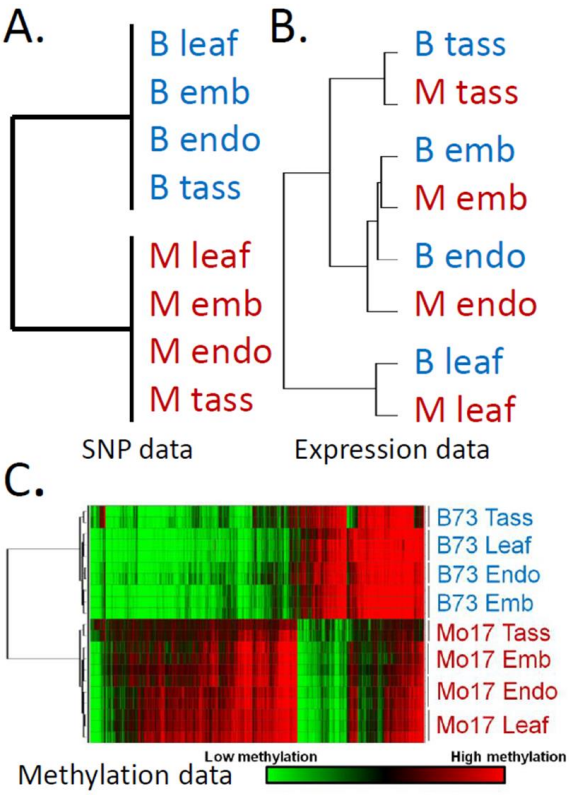



Figure 1: Hierarchical clustering of (A) SNP, (B) RNA-seq and (C) DNA methylation levels in four tissues of B73 and Mo17. The clustering in (A) is an artificial plot that is based on the assumption that different tissues of the same plants will contain the same genotype. The clustering in (B) is based on 2-3 biological replicates of RNA-seq data for each of the eight tissue-genotype combinations. In (C), the DNA methylation profiles from meDIP-chip profiling of the three biological replicates of each of the eight tissue-genotype are used for hierarchical clustering (Ward's, Euclidean distance). Only probes with a significant difference in at least one of the tissue or genotype contrasts were used for clustering. The heatmap indicates high (red) or low (green) DNA methylation levels.

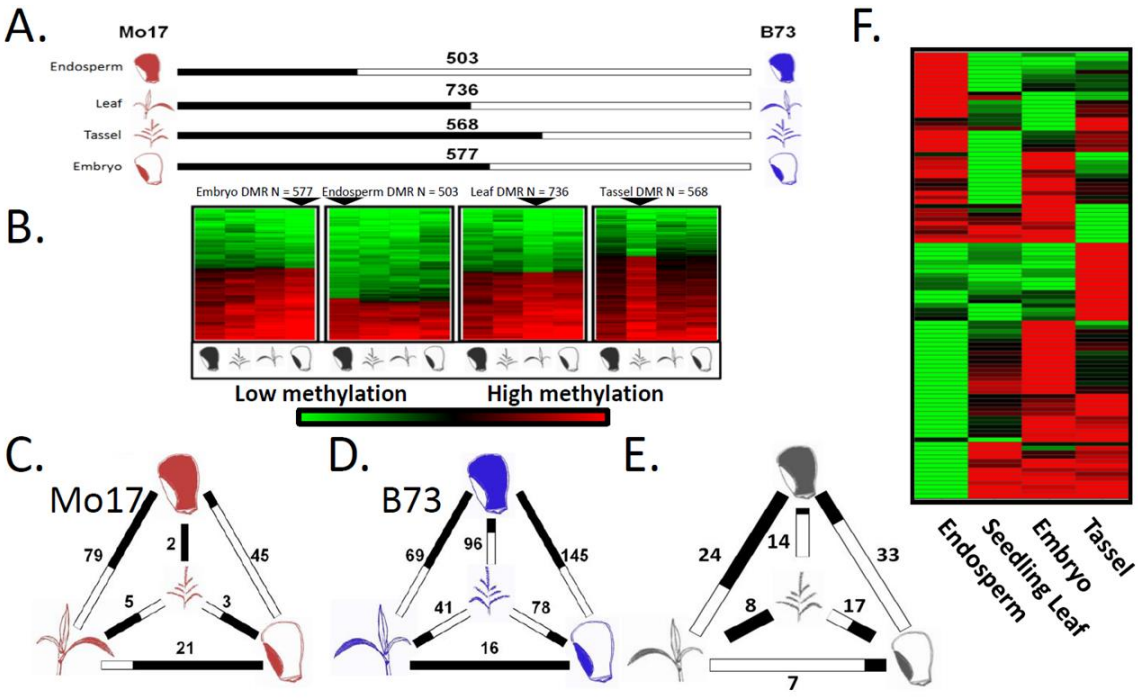



Figure 2: Discovery of differentially methylated regions (DMRs) among genotypes or tissues. (A) Genotype DMRs (gDMRs) were identified by contrasting each tissue of B73 and Mo17. The number above each bar indicates the total number of gDMRs and the shading of the bar indicates the proportion that are more highly methylated in each direction (the dark portion of the bar indicates the portion that are more highly methylated in Mo17). (B) Comparison of DNA methylation variation between genotypes in all four tissues. The gDMRs discovered in each of the four tissues (tissue used for discovery indicated above each plot) were used to perform hierarchical clustering (Ward's, Euclidean distance) of the DNA methylation levels of the same regions in all four genotypes. These plots provide evidence that regions of variable DNA methylation discovered in any one tissue exhibit reproducible genotype variation in the other three tissues. (C) The tissue DMRs (tDMRs) were separately identified in B73, Mo17 (D), and those conserved between the two genotypes (E). The numbers by each contrast indicate the total number and the shading of bars indicates the portion with higher methylation in each tissue. (F) Hierarchical clustering (Ward's, Euclidean distance, colored -2 ==2) of normalized tissue values for 103 non-redundant, genotype-conserved, tissue-specific DMRs.

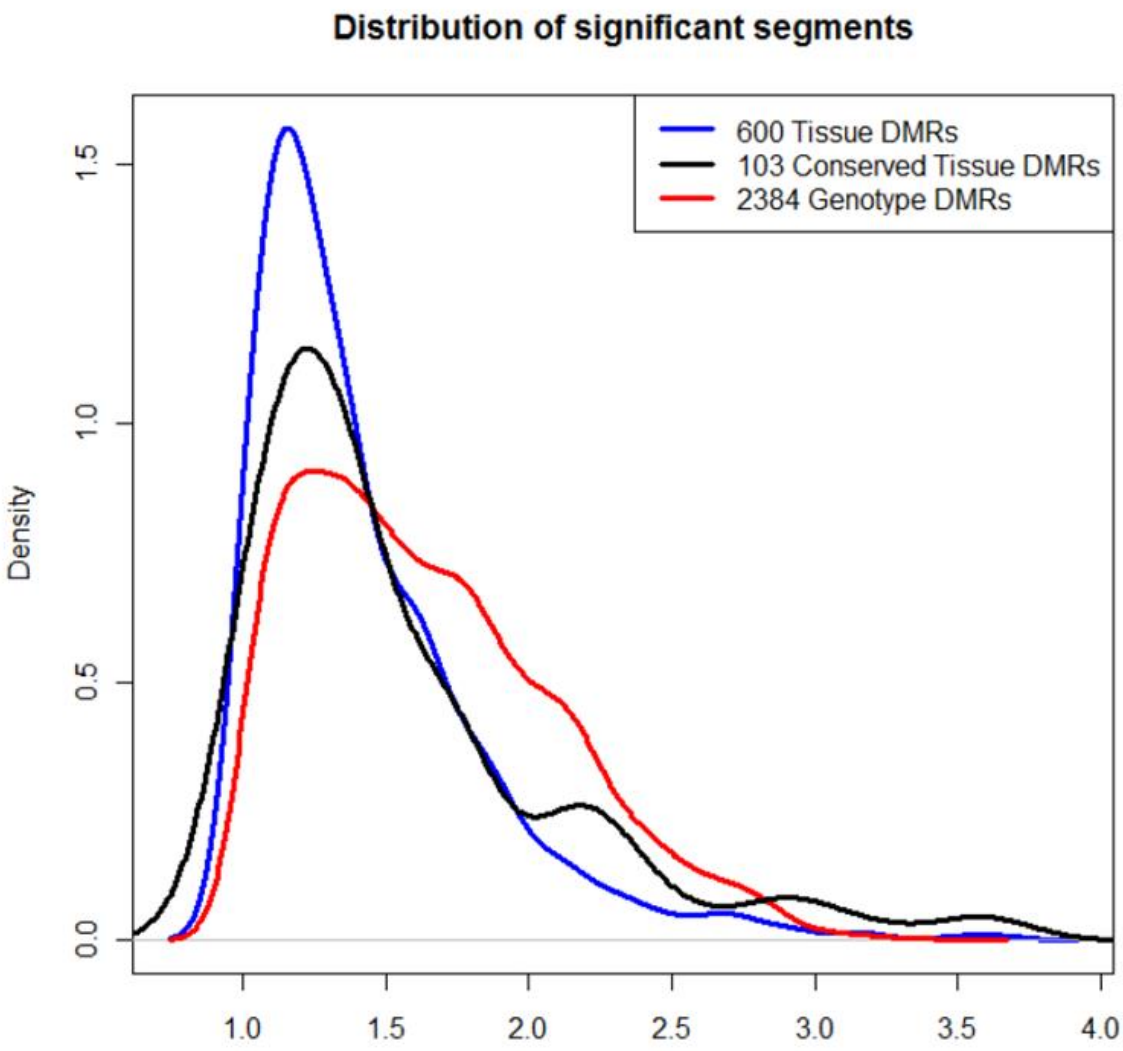



Figure 3: Density plots of the absolute value of segment means for genotype contrasts (black), all tissue contrasts (blue) and a subset of all tissue contrast segments that show conserved DMR status in the other inbred line (red). The x-axis indicates the relative methylation value between tissues. All density distributions end at 1.0 given the cutoffs used for defining DMRs.

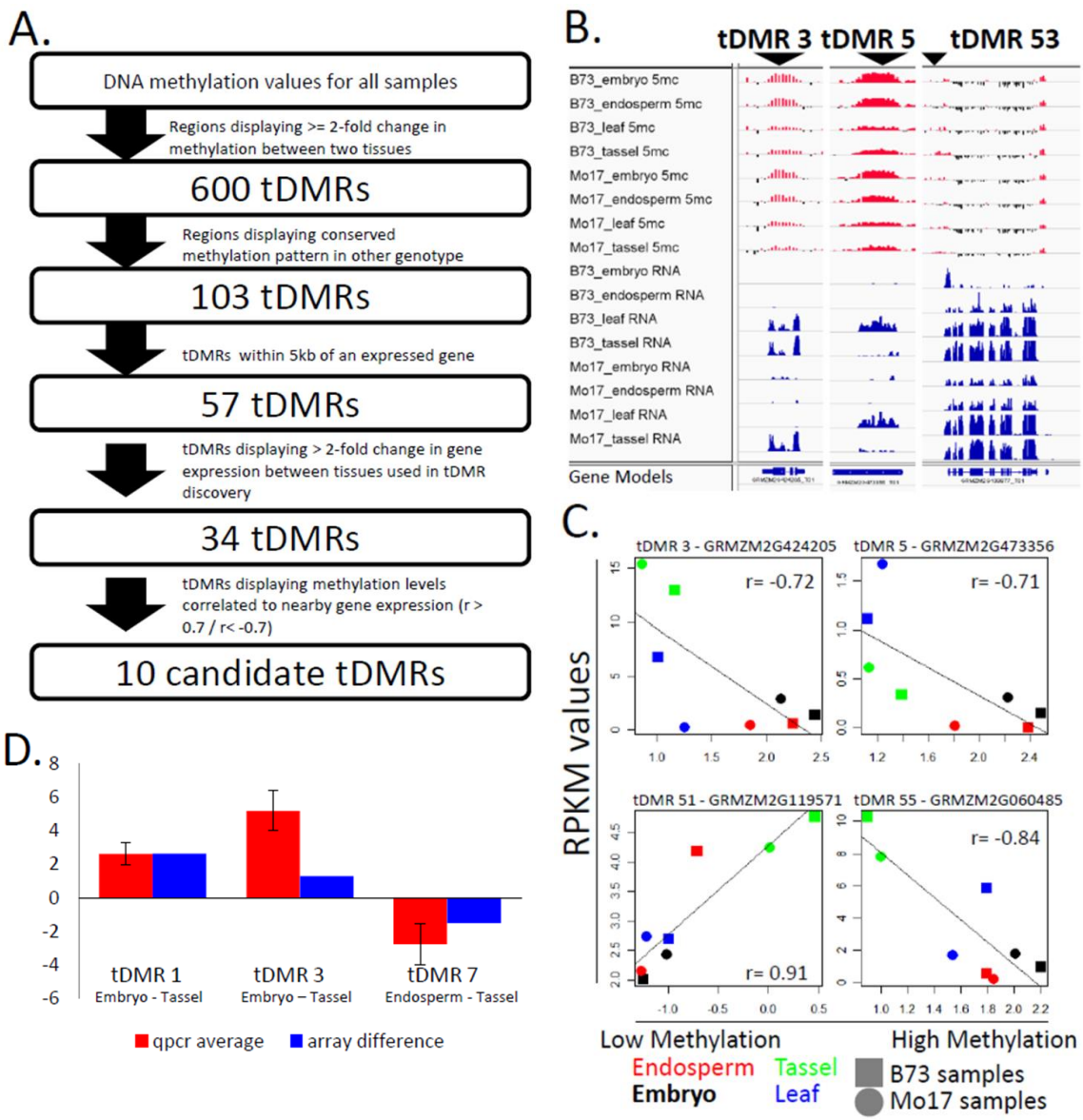



Figure 4: Relationship between tissue-specific DNA methylation and transcript abundance. (A) Flowchart describing sequential filtering of tDMRs. (B) The level of DNA methylation (red) from meDIP-chip profiling and transcript abundance (blue) from RNA-seq are shown for three tDMRs located near genes. (C) The correlation between DNA methylation and transcript abundance in the four tissues of both genotypes is shown for four tDMRs. (D) Example qPCR validations of tDMRs. Array-based methylation values (blue) are the difference between the first and second tissues used in the tDMR discovery. qPCR values (red) are calculated as the difference between the first tissue (C(t)Digest – C(t)Mock) and the second C(t)Digest – C(t)Mock) .

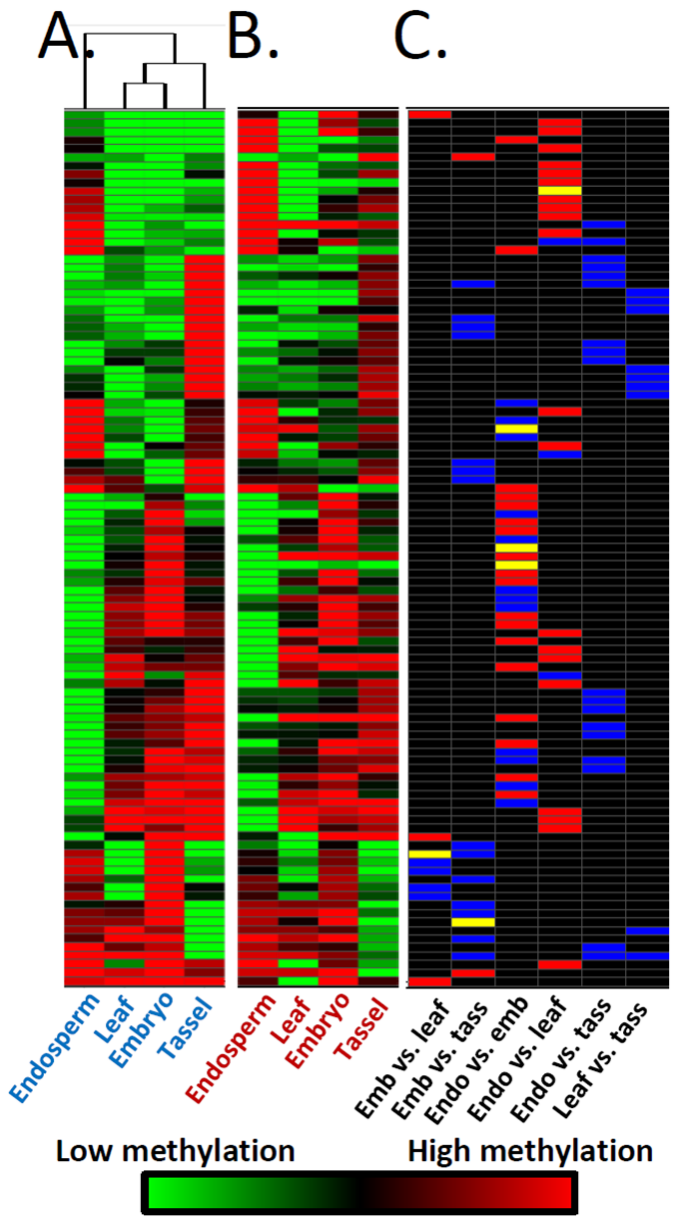



Figure 5**:** (A) Hierarchical clustering of DNA methylation levels in B73 was performed using the 103 non-redundant tDMRs. In (B) the same ordering was used to visualize the Mo17 DNA methylation levels for the same 103 tDMRs. (C) Indicates the tissue contrast(s) and genotype in which the tDMR was discovered; B73 only (blue), Mo17 only (red) or in both genotypes (yellow).

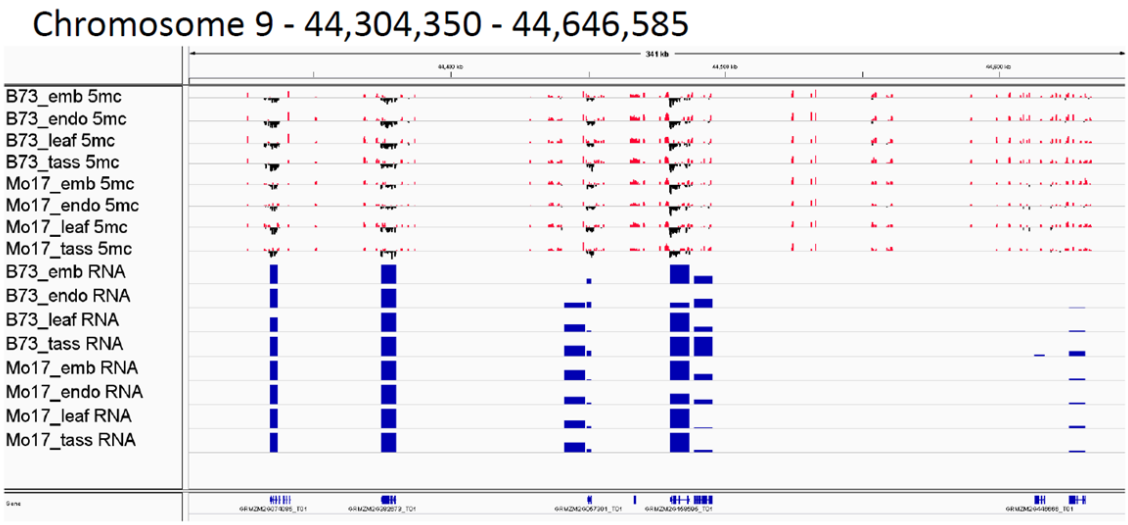



Figure 6: View of genomic region for all 8 meDIP-chip samples as well as per-gene RNA-seq counts for each sample on chromosome 9. Similar patterns of methylation are generally present for all samples. Variable expression is prominent between tissues and/or genotypes.

Chapter IV – **CONTEXT STATEMENT**

DNA methylation is often viewed solely as an epigenetic mark that acts independent of underlying genetic sequence. However, not all DNA methylation is truly epigenetic as underlying sequence polymorphisms may lead to variable patterns of DNA methylation observed across different genotypes. In this context, DNA methylation is not epigenetic but instead viewed as heterochromatin variation caused by genetic variation. This chapter details a mechanism of 'spreading' DNA methylation from a specific subset of highly methylated retrotransposable elements into the surrounding low-copy genomic sequence. This spreading of DNA methylation from certain retrotransposons can lead to variation in DNA methylation patterns when the insertion of the retrotransposon is polymorphic in different genotypes. In this way, examples of 'obligatory' DNA methylation variation have been identified within maize.

Chapter IV entitled 'Spreading of heterochromatin is limited to specific families of maize retrotransposons' has been adapted from the publication:

*Eichten SR, Ellis NA, Makarevitch I, Yeh C-T, Gent IJ, Guo L, McGinnis KM, Zhang X, Schnable PS, Vaughn MW, Dawe RK, Springer NM. (2012) Spreading of Heterochromatin Is Limited to Specific Families of Maize Retrotransposons. PLoS Genet 8(12): e1003127. doi:10.1371/journal.pgen.1003127*

In the development of this manuscript, other researchers played roles in the creation of data and assistance to my analysis. The experimental design was developed by SRE and NS. Microarray methylation profiling and local PCR validations of transposable elements were performed by PJH (non-author). Genome-wide bisulfite sequence data creation and analysis was performed by NE, JG, and KD. Methylation mutant samples were provided by KM and profiled by PJH and SRE. MV assisted SRE in proper normalization of microarray data to facilitate analysis. Sequencing data from Mo17 used to identify the presence or absence of specific retrotransposable elements was provided by CY, XZ, and PS. Text and figures of this chapter were developed by SRE with assistance from NS.

Many of the journal specific sections as well as a supplemental data table defining PCR primers used in presence-absence detection have been omitted from this chapter. All omitted data is available through the initial scientific publication.

# Chapter IV
# Spreading of heterochromatin is limited to specific families of maize retrotransposons

Transposable elements (TEs) have the potential to act as controlling elements to influence the expression of genes and are often subject to heterochromatic silencing. The current paradigm suggests that heterochromatic silencing can spread beyond the borders of TEs and influence the chromatin state of neighboring low-copy sequences. This would allow TEs to condition obligatory or facilitated epialleles and act as controlling elements. The maize genome contains numerous families of class I TEs (retrotransposons) that are present in moderate to high copy numbers and many are found in regions near genes which provides an opportunity to test whether the spreading of heterochromatin from retrotransposons is prevalent. We have invesTrends in Geneticsated the extent of heterochromatin spreading into DNA flanking each family of retrotransposons by profiling DNA methylation and di-methylation of lysine 9 of histone 3 (H3K9me2) in low-copy regions of the maize genome. The effects of different retrotransposon families on local chromatin are highly variable. Some retrotransposon families exhibit enrichment of heterochromatic marks within 800-1200 base pairs of insertion sites while other families exhibit very little evidence for the spreading of heterochromatic marks. The analysis of chromatin state in genotypes that lack specific insertions suggests that the heterochromatin in low-copy DNA flanking retrotransposons often results from the spreading of silencing marks rather than insertion-site preferences. Genes located near TEs that exhibit spreading of heterochromatin tend to be expressed at lower levels than other genes. Our findings suggest that a subset of retrotransposon families may act as controlling elements influencing neighboring sequences while the majority of retrotransposons have little effect on flanking sequences.

**INTRODUCTION**

A substantial fraction of most eukaryotic genomes is composed of transposable elements (TEs) (Biemont & Vieira, 2006; Wicker et al., 2007; Levin & Moran, 2011; Lisch & Bennetzen, 2011). While these TEs are sometimes referred to as “junk” DNA, there is evidence for potential functional roles in some instances (Biemont 2010). Indeed, Barbara McClintock used the term

"controlling elements" to describe the potential for these sequences to affect the regulation of endogenous genes (McClintock 1984; Comfort 2001). Mobile genetic elements include class I retrotransposons and class II DNA transposons (Wicket et al., 2007). The class I TEs transpose via an RNA intermediate while class II TEs utilize a DNA intermediate for transposition. There are a variety of sub-families of both types of TEs (Wicket et al., 2007) that differ in structure, activity, and integration patterns.

TEs could influence neighboring genes by providing regulatory elements or promoters that would alter expression levels or patterns (Girard & Freeling, 1999; Feschotte 2008). Alternatively, TEs may be targeted for silencing and this silencing could spread to affect neighboring sequences potentially including endogenous genes or regulatory elements (Richards 2006; Weil & Martienssen, 2008; Lisch 2009). There are several examples in which heterochromatic silencing of TEs can influence expression of nearby genes, including the *agouti* and *Axin* locus in mouse (Michaud et al., 1994; Morgan et al., 1999; Rakyan & Whitelaw, 2003), *FLC* (Liu et al., 2004), *FWA* (Soppe et al., 2000) and BNS (Saze & Kakutani, 2007) in Arabidopsis and sex-determination in melons (Martin et al., 2009). While there are examples of heterochromatin spreading from retrotransposons to neighboring sequences, it is unclear how general this phenomenon is. Whole genome profiling of DNA methylation in Arabidopsis (Cokus et al., 2008) found that the level of DNA methylation often had sharp boundaries at the edge of repeats although some inverted repeats did exhibit spreading. Another study (Ahmed et al., 2011) found limited (200-500bp) spreading of DNA methylation surrounding some TEs in Arabidopsis. There is evidence that highly methylated TEs are under-represented near genes in Arabidopsis and it has been suggested that the silencing of TEs located near genes might have deleterious consequences (Ahmed et al., 2011; Hollister & Gaut, 2009; Hollister et al., 2011). There is evidence for variation in the spreading of heterochromatin for different families of TEs in mouse (Rebollo et al., 2011) and evidence that differences in TE insertions contribute to gene expression variation in other rodents (Pereira et al., 2009).

The complex organization of the maize genome, with interspersed genes and TEs (Bennetzen et al., 1994; Schnable et al., 2009; SanMiguel & Bitte, 2009), provides an excellent system in which to study the effects of retrotransposons on neighboring DNA. Many model organisms have relatively small, compact genomes with relatively few retrotransposons. Since these genomes do not have a number of moderate-high copy retrotransposon families it can be difficult to

assess the variation in spreading of heterochromatin to neighboring low-copy sequences. The maize genome is more representative of the organization of sequences observed within most flowering plants and is similar to the organization of many mammalian genomes as well. There are a large number of distinct families of retrotransposons within the maize genome and many of these families are moderate to high copy number (SanMiguel & Vitte, 2009; Meyers et al., 2001; Baucom et al., 2009; Du et al., 2009). In addition, haplotypes differ substantially with regard to the presence or absence of specific retrotransposon insertions (Du et al., 2009; Fu & Dooner, 2002; Brunner et al., 2005; Wang & Dooner, 2006). The majority of repetitive sequences, including retrotransposons, in the maize genome are highly methylated (Bennetzen et al., 1994; Rabinowicz et al., 1999; Palmer et al., 2003; Whitelaw et al., 2003; Emberton et al., 2005).

The existence of heavily silenced retrotransposons interspersed with genes throughout the maize genome provides ample opportunities for TEs to exert epigenetic regulation on surrounding sequences. We were interested in further documenting the extent of heterochromatin spreading from maize retrotransposons to neighboring sequencings. Genomic profiling of DNA methylation and H3K9me2 found that heterochromatic spreading is only observed for a small number of specific retrotransposon families. These families tend to be enriched in pericentromeric regions of chromosomes. The analysis of haplotypes lacking specific retrotransposon insertions provides evidence that the adjacent heterochromatin is the result of spreading rather than insertion site bias.

## RESULTS

### *Heterochromatin spreads from some retrotransposons within the maize genome*

DNA methylation and chromatin modifications were profiled for low-copy sequences in the B73 genome using methylated DNA immunoprecipitation (meDIP) and chromatin-immunoprecipitation (ChIP) with antibodies specific for H3K9me2 or H3K27me3, respectively. The fractions of the genome enriched for DNA or histone modifications were hybridized to a high-density microarray containing ~2.1 million long oligonucleotide probes derived from the unmasked, non-repetitive fraction of the maize genome. The probes are spaced every 200 bp in the low-copy portions of the maize genome and can provide a profile for the chromatin state in

these regions (Eichten et al., 2011). Our analyses focused on a subset of ~1.4 million probes that are single-copy (no other sequences with at least 90% identity within maize genome sequence). While this approach does not provide information on the chromatin state within repetitive sequences it can assess how retrotransposons impact neighboring sequences (Eichten et al., 2011). An independent whole-genome bisulfite sequencing dataset (~7X coverage) was used to further confirm the patterns that we observed in the meDIP-chip experiments. This independent approach was able to assess DNA methylation within retrotransposons as well as low-copy sequences. The enrichment for sequences associated with H3K9me2 was validated using a set of known sequences (Figure 1A) and several sequences identified by the profiling experiments (Figure 1B).

A large number of class I TEs (retrotransposons) have been identified within the maize genome (Baucom et al., 2009). These retrotransposons tend to be highly methylated in CG and CHG sequence contexts (Figure 2). We assessed whether heterochromatic chromatin modifications would be enriched in the single-copy regions that flank these retrotransposons. The chromatin state of sequences adjacent to any specific insertion of a retrotransposon is influenced by regulatory and insulator sequences as well as any potential effects of nearby retrotransposons. By assessing the average level of chromatin modifications near all of the retrotransposons of the same family it is possible to identify whether retrotransposon families vary in their influence on local chromatin state. Single-copy probes that are located within 4 kb of all retrotransposons were identified and used to assess the level of chromatin modifications in 200 bp bins of low-copy sequences adjacent to superfamilies, such as gypsy or copia (Figure 3) and individual families of retrotransposons (Figure 4). Many of the retrotransposon families exhibit elevated levels of DNA methylation and H3K9me2 in the 200 bp immediately adjacent to their insertion sites (Figure 3). Because the meDIP-chip profiling of DNA methylation has a resolution of 300-500 bp it is likely that some of the apparent increase in DNA methylation levels very close to retrotransposons represents DNA methylation within the repeats themselves.

A subset of the retrotransposon families also exhibit elevated levels of DNA methylation and H3K9me2 in regions more than 200 bp away from their insertion sites. In general, levels of H3K9me2 and DNA methylation were well correlated, but there were some families with different enrichment for these two marks. As expected, there was no evidence for enrichment (or depletion) of the facultative heterochromatin mark, H3K27me3, in regions flanking the

retrotransposons (Figure 4C). To identify retrotransposon families associated with significant levels of spreading of heterochromatic chromatin modifications in adjacent low-copy sequences we compared the distribution of methylation levels in each 200 bp bin with a set of randomly permuted data (10,000 randomly assigned "insertions") and defined whether each 200 bp bin had significantly higher levels of a chromatin modification than random genomic sequences. Retrotransposon families that exhibit significant ($p<0.001$) enrichment for a chromatin modification for each bin up through at least 800 bp were classified as spreading families. There are 39 retrotransposon families that exhibit significant enrichments of DNA methylation and H3K9me2 within each of the first four 200 bp bins adjacent to their insertion sites. These families will hereafter classified as "spreading (both)" families (Figure 4A-B,E-F and Figure 5). Another 10 retrotransposon families had significant levels of H3K9me2 but did not have at least 800 bp of significant enrichment for DNA methylation. These families will hereafter be classified as "spreading (H3K9)" (Table 1; Figure 4A-B, G and Figure 6). Many of these H3K9 only spreading families have elevated levels of DNA methylation in these same regions (Figure 5), but do not pass the significance threshold for all bins within the adjacent 800 base pairs. The remaining 95 retrotransposon families did not exhibit significant enrichment for either DNA methylation or H3K9me2 (example in Figure 4H). There was no evidence for significant enrichment of H3K27me3 in regions near any retrotransposon families (Figure 4C). The initial classification of retrotransposon families was based upon chromatin profiles from B73 seedling tissue. However, very similar patterns were observed for other genotypes and tissues. Specifically, the same families have significant enrichments of DNA methylation in Mo17 seedling, B73 endosperm and B73 embryo tissue (Figure 7). The H3K9me2 patterns are quite similar in both B73 and Mo17 seedlings (Figure 8A-B) and there was no evidence for enrichment for H3K27me3 in any of the tissues or genotypes assessed (Figure 8C-E).

The analysis of the whole-genome bisulfite sequencing data supports the classifications of different retrotransposon families (Figure 4D and Figure 2). Both CG and CHG DNA methylation levels are higher in low-copy regions flanking spreading (both) and spreading (H3K9) families (Figure 4D). The level of DNA methylation is higher in sequences flanking spreading (both) retrotransposon families than for sequences flanking spreading (H3K9) retrotransposons. The sequences flanking the non-spreading families have DNA methylation levels that are similar to randomly selected genomic regions (Figure 4D). The analysis of internal (within the repeat itself) DNA methylation levels (Figure 2) reveals that the levels of CG methylation within

retrotransposons with, or without spreading are similar. However, the spreading (both) and spreading (H3K9) retrotransposon families have slightly elevated levels of CHG methylation at internal sequences. Interestingly, the non-spreading retrotransposon families tend to have higher levels of internal CHH methylation than do spreading families (Figure 2). The relative levels of H3K9me2 within retrotransposons was assessed by qPCR for 10 of the families, including six spreading (both) and four non-spreading families (Figure 9). There was no evidence for higher levels of H3K9me2 within the families that exhibit heterochromatic spreading than for those that do not (Figure 9). The elevated levels of DNA methylation and/or H3K9me2 in low copy sequences flanking the insertion sites observed for a subset of the retrotransposon families are largely confined to the region within 800 – 1,600 bp of the insertion site (Figure 4A-B). A closer examination of the levels of DNA methylation and H3K9me2 near each spreading family indicates a fairly sharp drop to non-significant levels of the modifications within 2kb of the insertion site (Figures 4E-G, 5 and 6) for spreading families. The visualization of individual spreading families (Figures 5 and 6) reveals that the distance of heterochromatin spreading varies for different retrotransposon families. This analysis provides clear evidence for diversity in the prevalence of heterochromatin found in low-copy regions flanking different families of retrotransposons in the maize genome.

***Spreading of heterochromatin does not require CMT or Mop1.***

The mechanistic basis for the spreading of heterochromatin is not well defined. It is possible that the interplay between DNA methylation and histone modifications (Lippman et al., 2004; Bernatavichute et al., 2008) would result in spreading of chromatin modifications beyond the specific target. To probe the mechanistic basis of spreading we profiled DNA methylation levels in several maize mutants that are known, or expected, to affect DNA methylation patterns. In plants, one pathway that impacts DNA methylation is RNA-directed, and requires the activity of multiple RNA polymerases (RNA PolIV and PolV), an RNA dependent RNA polymerase (RDR2), a dicer like protein, and multiple chromatin modifiers (Haag & Pikaard, 2011). The *mop1* mutant of the maize *Rdr2* gene (Dorweiler et al., 2000; Lisch et al., 2002; Alleman et al., 2006) exhibits variable expression of specific retrotransposon families in mutant relative to wild-type tissue (Jia et al., 2009). However, we found no evidence for a consistent effect of the *mop1* mutation on the expression levels of spreading or non-spreading retrotransposon families. Indeed, spreading retrotransposon families include examples of both up- and down-regulation in mop1 mutant

individuals relative to wild-type (Table 1). In addition, there were examples of non-spreading retrotransposon families that do, and do not, exhibit altered expression in mop1 plants. The levels of DNA methylation in low-copy sequences neighboring retrotransposon families was analyzed in the *mop1* mutant to assess whether the spreading of heterochromatin might be affected (Figure 10). There was no evidence for a reduction in the distance or magnitude of the spreading of DNA methylation in the *mop1* mutants relative to wild-type plants. The small RNA profile of spreading and non-spreading retrotransposon families was assessed using a recently published small RNA profile based on B73 shoot tissue (Barber et al., 2012). The average count of small RNAs per retrotransposon and coverage of retrotransposon did not vary between spreading (both), spreading (H3K9) or non-spreading retrotransposon families (Figure 11).

Spreading retrotransposons exhibit higher levels of CHG methylation within the retrotransposon themselves (Figure 2). Spreading levels were assessed in plants that were homozygous for mutations in the maize chromomethylase *zmet2* (GRMZM2G025592) gene, which contributes substantially to CHG methylation (Papa et al., 2001; Makarevitch et al., 2007). While there were examples of locus-specific alterations in DNA methylation levels in this mutant, there was no evidence for a reduction in the spreading of DNA methylation in low copy sequences flanking spreading retrotransposon families (Figure 10).

***Analysis of empty sites***

The observation that certain families of retrotransposons have high levels of heterochromatic modifications in adjacent regions could reflect insertion site biases for these families or indicate that these families cause local spreading of heterochromatin. Examples of "empty" sites in the Mo17 haplotypes were identified and used to assess whether the high levels of DNA methylation would be observed in these regions when the retrotransposon was absent. Mo17 whole-genome shotgun WGS) sequences (generated by the DOE's Joint Genome Institute (JGI) and downloaded from ftp://ftp.jgipsf.org/pub/JGI_data/Zea_mays_Mo17/) were aligned to the B73 reference genome sequence. Empty sites were defined as being those as which at least three Mo17 sequence reads cover a low-copy sequence flanking an insertion but do not align to the retrotransposon itself and for which no Mo17 reads cover the junction between the low-copy sequence and the retrotransposon. In total, 668 empty sites were identified for the spreading (both) retrotransposon families and 29 empty sites for the spreading (H3K9) retrotransposon families for which we had DNA methylation data in the unique regions flanking

the insertion. The lack of the specific insertion in Mo17 was confirmed at 13 of the 14 empty sites that were tested using site-specific PCR primers to confirm the presence/absence of specific insertions. This suggests that there is a low false-positive rate in the identification of empty sites in Mo17. However, given the low coverage of the WGS data and challenges associated with aligning polymorphic sequences it is likely that many of the true empty sites were not identified in this analysis.

The level of DNA methylation at the probe nearest to the empty site was used to assess relative DNA methylation levels with (B73) and without (Mo17) each insertion (Figure 12). The low-copy DNA flanking many of the empty sites showed differences in DNA methylation levels between B73 and Mo17 in 34.7% of the empty sites flanking spreading (both) retrotransposons and in 43.5% of the empty sites flanking spreading (H3K9) retrotransposons (Figure 12A). Over 95% of the empty sites with differential methylation had higher DNA methylation levels in B73 (the genotype with the insertion) than in Mo17. While 35-43% of the probes flanking the empty sites for spreading retrotransposons had variable DNA methylation in B73 and Mo17, only 3% of genome-wide probes assayed show significantly different levels of DNA methylation in B73 and Mo17 and these differences include equal frequencies of higher methylation levels in each genotype. This suggests that the insertion of the retrotransposon conditioned higher levels of DNA methylation and was responsible for the observed DNA methylation polymorphisms. In contrast, DNA methylation levels were similar (and frequently quite high) between B73 and Mo17 when the retrotransposon insertion was present in both genotypes (Figure 12A). Closer inspection of several of the empty sites provides evidence for enrichment of DNA methylation or H3K9me2 in regions flanking the sites in B73 but these modifications were not observed in the Mo17 haplotype that lacks the retrotransposon (Figure 12B). The presence of the insertion as well as the enrichment for DNA methylation was also assessed in five other inbred genotypes of maize (Figure 12B). The presence of insertions was strongly correlated with the presence of high levels of DNA methylation in these other genotypes as well. These results suggest that the high level of heterochromatin observed around these spreading retrotransposon families is an outcome of TE insertion rather than insertion site bias.

***Characterization of retrotransposon families that induce local spreading of heterochromatin***

The finding that only a subset of maize class I retrotransposon families are associated with local spreading of heterochromatin suggested that there might be intrinsic differences among different retrotransposon families that would explain this variation. We proceeded to characterize these families to ascertain whether there were specific common attributes of spreading families. None of the LINE families exhibit evidence for spreading of heterochromatic marks. RLG (gypsy) families are over-represented among spreading (both) retrotransposon families, while the spreading (H3K9) retrotransposons have more RLC (copia) families than expected (Figure 13A). Spreading (both or H3K9) retrotransposons exhibit significantly higher copy number and comprise a greater fraction of the genome (Table 1, attributes from (Meyers et al., 2001)) than do non-spreading retrotransposon families (Figure 14A-B). While there are significant differences in copy number and total Mb within the genome there are examples of families with spreading that have lower copy numbers (Figure 14A). In addition, spreading (both) retrotransposon families have significantly higher average fragment lengths than do non-spreading families (Table 1). Spreading families do not have a significant difference in their mean insertion date relative to non-spreading families (Table 1). However, the analysis of average insertion date for each family (Figure 14C) shows that while non-spreading retrotransposon families include both old and young families the spreading (both) retrotransposon families only include younger families. The analysis of several characteristics of the retrotransposon families with and without spreading provides evidence for some significant differences but none of these factors are sufficient for predicting whether or not spreading occurs. Previous studies that had assessed expression of some retrotransposons in maize tissues (Ohtsu et al., 2007; Vicient 2010) did not find unusually high or low abundance for transcripts of the families with heterochromatin spreading relative to other families.

The relative abundance of spreading (both) retrotransposons is higher in the middle of the chromosome than the other families suggesting that these retrotransposons may be enriched in pericentromeric regions (Figure 3B). However, it should be noted that there are other retrotransposon families also preferentially located in pericentromeric regions (Meyers et al., 2001) but that do not show spreading of heterochromatin to low-copy adjacent regions. Hence, the pericentromeric enrichment is insufficient for heterochromatin spreading. The observation that the spreading (both) retrotransposon families are enriched in pericentromeric regions

suggested the possibility that the higher levels of DNA methylation in flanking sequences may be due to sampling bias. Because pericentromeric regions tend to have higher levels of DNA methylation (Eichten et al., 2011) it is possible that higher sampling of these regions led to the observation of spreading. However, an analysis of the levels of DNA methylation in low-copy flanking regions relative to chromosome position provides evidence that low-copy sequences flanking spreading (both) retrotransposons is substantially higher than the corresponding regions flanking non-spreading families throughout the chromosome in both CG and CHG contexts (Figure 15). The levels of CG and CHG DNA methylation in spreading (H3K9) retrotransposon families are intermediate (Figure 15).

***Genes located near retrotransposon with spreading of heterochromatic marks tend to have lower expression levels***

The finding that some retrotransposon families exhibit spreading of heterochromatic marks to surrounding sequences while others do not led us to hypothesize that these families may influence expression of nearby genes. RNA-seq was used to estimate transcript abundance in three tissues of B73 and Mo17 including the identical leaf tissue samples used for profiling DNA methylation levels. All maize genes were annotated to identify the first retrotransposon 5' of the transcription start site and to determine the distance between the retrotransposon and the transcription start site. Genes that are located near retrotransposons that exhibit spreading (both or H3K9) have significantly ($p<0.001$) lower expression levels in all genotypes and tissue examined (Figures 16 and 17). This reduction in expression is most severe when we examine genes with retrotransposons inserted within 500bp of the transcription start site. As the distance between the insertion site and the transcription start site increases there is less evidence for an effect on expression levels, suggesting a limited range within which retrotransposons can influence gene expression. The genes located near spreading (both) and spreading (H3K9) retrotransposons frequently have no detectable expression (Figure 17B). However, even if we exclude genes with no expression, the mean expression of genes near spreading retrotransposons is lower ($p<0.001$) (Figure 17C).

**DISCUSSION**

Epigenetic variation in low-copy sequences can be the result of pure epigenetic changes (no correlation with DNA sequence polymorphisms) or occur in a facilitated or obligatory fashion such that DNA sequence differences contribute to the epigenetic changes (Richards 2006). A handful of examples in which epigenetic differences that impact phenotype has been shown to involve TEs inserted near genes (Michaud et al., 2004; Morgan et al 1999; Rakyan & Whitelaw, 2003; Liu et al., 2004; Soppe et al., 2000; Saze & Kakutani, 2007; Martin et al., 2009; Martienssen et al., 1990) and genomic profiling of DNA methylation in Arabidopsis has revealed some examples of heterochromatin spreading from TEs (Cokus et al., 2008; Ahmed et al., 2011). However, it has not been clear whether all TEs have similar effects on neighboring chromatin or whether there are family-specific attributes that affect the spreading of heterochromatin. A recent study analyzed several families of retrotransposons in mouse and found that there is variation in the level of heterochromatin spreading (Rebollo et al., 2011) and there have been suggestions of variation in the effects of different repetitive elements on nearby gene expression in Arabidopsis (Hollister & Gaut, 2009; Hollister et al., 2011). The complex organization of the maize genome with interspersed TEs and genes provides the opportunity to examine differences among class I retrotransposon families. The chromatin state of any low-copy region of a genome is likely influenced by nearby sequences including regulatory elements and insulator elements. In addition, it is quite likely that TEs will exert an influence on the chromatin state. By examining the average level of chromatin modifications in low-copy sequences neighboring families of retrotransposons we found evidence for heterochromatic spreading from a subset of the moderate to high-copy retrotransposon families in maize. Even in these families the heterochromatic marks spread only 600 – 1,000 base pairs from the retrotransposon. It is worth noting that there may be other mechanisms through which retrotransposons influence flanking regions. Our assessment is based upon only two chromatin marks, H3K9me2 and DNA methylation. These marks are frequently associated with heterochromatin, but there may be other specific types of chromatin marks that spread from these and transposon families.

There is also evidence that differences in interspecific variation in transposon insertions contributes to gene expression diversity between related species (Hollister & Gaut, 2009; Hollister et al., 2011). Here we provide evidence that transposon insertions can also contribute

to differences in DNA methylation patterns and gene expression levels within a species. Many TE insertions exhibit presence/absence variation among maize haplotypes (Du et al., 2009; Fu & Dooner, 2002; Brunner et al., 2005; Wang & Dooner, 2006). The retrotransposons that cause spreading of heterochromatin are expected to result in obligatory epigenetic variation in the low-copy sequences that flank insertions. Indeed, we found that the levels of DNA methylation and H3K9me2 were quite different in B73 and Mo17 at regions that exhibit presence/absence variation for an insertion of a retrotransposon from one of the spreading families. Specifically, these retrotransposons with spreading of heterochromatin may contribute to obligatory and facilitated epialleles, as defined by Richards (Richards 2006), among different genotypes. Genomic resequencing is often used to identify SNPs as a means to explain phenotypic variation. However, it might be important to also use resequencing data to identify retrotransposon insertion polymorphisms, especially for the retrotransposon families that exhibit spreading of heterochromatic marks. The polymorphism for these insertions may lead to functional variation in the expression of nearby genes.

Barbara McClintock proposed the concept that transposons could serve as "controlling" elements that would influence nearby genes (McClintock 1984; Comfort 2001) and this could be extended to include the potential for retrotransposons to influence nearby genes as well. There are examples in which transposons contain regulatory elements or cryptic promoters that can influence the expression of nearby genes (Feschotte 2008; Kashkush et al., 2003). There is also evidence that some transposons can act as controlling elements by "seeding" heterochromatin that spreads to adjacent low copy sequences (Richards 2006; Weil & Martienssen, 2008; Lisch 2009). Here we have shown that this activity is not a generic feature of all retrotransposons but is instead limited to a subset of retrotransposons. Hollister and Gaut (2009) provide evidence that the presence of heavily silenced TEs near genes may lead to reduced expression and result in fitness consequences. This would suggest that many TEs would evolve to have minimal effects on neighboring genes to reduce their fitness costs. There is evidence that some *Drosophila* retrotransposons contain insulator elements that reduce the spreading of chromatin states (Gdula et al., 1996). Alternatively, studies at the *bns* locus in Arabidopsis have suggested the presence of an active mechanism to prevent the spreading of heterochromatin from retrotransposons (Saze et al., 2008). It might be expected that different families of TEs would vary in their ability to limit potential spreading of heterochromatin through the presence of insulators or the recruitment of factors that limit spreading. Hollister and Gaut (2009) noted

heterogeneity among families of Arabidopsis class I retrotransposons for their distance to the nearest gene and suggested that this may reflect family specific differences in heterochromatin spreading. The analysis of the large families of retrotransposons in maize permitted us to identify several families of retrotransposons with high levels of spreading. These retrotransposon families may be considered as bad "neighbors" for genes. Indeed we find that many genes located near retrotransposons with spreading tend to be silenced or expressed at lower levels. We might predict that insertions of retrotransposons from these families will be more strongly selected against when inserted near genes, especially if they affect gene expression. Therefore, our observed expression differences will only report effects that have been tolerated during natural and artificial selection of maize lines. Consistent with this possibility, our observation that these retrotransposon families are enriched in relatively gene-poor pericentromeric regions may reflect selection against insertions of these retrotransposons when they are near genes. Further research efforts to understand the basis of this difference will be important in providing the ability to predict which retrotransposon families are likely to condition spreading of heterochromatin and understanding the consequences of the spreading of heterochromatin.

**EXPERIMENTAL METHODS & PROCEDURES**

**Epigenomic profiling:** DNA methylation profiling on three replicates of 3$^{rd}$ leaf tissue of B73 and Mo17 was performed as described (Eichten et al., 2011; GEO Accession – GSE29099). Briefly, methylated DNA was immunoprecipitated with an anti-5-methylcytosine monoclonal antibody from 400ng sonicated DNA using the Methylated DNA IP Kit (Zymo Research, Orange, CA; Cat # D5101). For each replication and genotype, whole genome amplification was conducted on 50-100ng IP DNA and also 50-100ng of sonicated DNA (input control) using the Whole Genome Amplification kit (Sigma Aldrich, St. Louis, MO, Cat # WGA2-50RXN). For each amplified IP input sample, 3ug amplified DNA were labeled using the Dual-Color Labeling Kit (Roche NimbleGen, Cat # 05223547001) according to the array manufacturer's protocol (Roche NimbleGen Methylation UserGuide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA was labeled with Cy3. H3K9me2 and H3K27me3 profiling were performed on three replicates of B73 and Mo17 seedlings using antibodies specific for H3K27me3 (#07-449) and H3K9me2 (#07-441) purchased from Millipore (Billerica, USA). For each replicate, 1 g of

plant material was harvested on ice, rinsed with water, and crosslinked with 1% formaldehyde for 10 minutes under vacuum. Cross-linking was quenched by adding glycine solution to a final concentration of 0.125M under vacuum infiltration for 5 minutes. Treated tissue was frozen in liquid nitrogen and stored at -800C until chromatin extraction. Chromatin extractions were performed using EpiQuik Plant ChIP Kit (Epigentek, Brooklyn, USA) according to manufacturer's recommendations. Extracted chromatin was sheared in 600 µl of the EpiQuik buffer CP3F with 5 10-second pulses on a sonicator. To test and optimize sonication conditions, cross-linking was reversed in a sample of sheared chromatin and the resulting products were analyzed on agarose gel. Sonication conditions were optimized to yield predominantly 200 – 500 bp DNA samples. Chromatin immunoprecipitations, reverse cross-linking, and DNA cleanup was performed using EpiQuik Plant ChIP Kit (Epigentek) according to manufacturer's recommendations. For each genotype, antibody, and replicate, 50 – 100 ng of input and immunoprecipitated (IP) DNA was amplified with a whole genome amplification kit (WGA2, Sigma, St. Louis, USA). The amplification of no antibody control (negative control) was always 5 – 10 fold less efficient confirming specificity of immunoprecipitation. For each amplified IP and input sample, 3 ug of amplified DNA were labeled using the Dual-Color Labeling Kit (Roche NimbleGen, Cat # 05223547001) according to the array manufacturer's protocol (Roche NimbleGen Methylation User- Guide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA was labeled with Cy3. Samples were hybridized to the custom 2.1 M probe array (GEO Platform GPL13499) for 16–20 hrs at 42C. Slides were washed and scanned according to NimbleGen's protocol. Images were aligned and quantified using NimbleScan software (Roche NimbleGen) producing raw data reports for each probe on the array. The histone modification and methylation mutants array data can be obtained from GEO accession (GSE39460). The resulting microarray data were imported into the Bioconductor statistical environment (http://bioconductor.org/). Microarray data channels were assigned the following factors: B73 immunoprecipitation, Mo17 immunoprecipitation, B73 input, or Mo17 input depending on sample derivation. Non-maize probes and vendor-supplied process control probes were configured to have analytical weights of zero. Variance-stabilizing normalization was used to account for array-specific effects. Factor-specific hybridization coefficients were estimated by fitting fixed linear model accounting for dye and sample effects to the data using the limma package (Smyth 2005). The probes were each annotated with respect to their location relative to repeats from the ZmB73_5a_MTEC_repeats file available from www.maizesequence.org.

Each probe was only associated with the closest repeat and all probes located within 5kb of a repeat were retained for further analyses. The probes were assigned based on distance to the retrotransposon and include both upstream (5') and downstream (3') sequences together. The distribution of retrotransposons along the length of the chromosome was performed as described in (Gent et al., 2012). Data formatted for the Integrative Genomics Viewer (IGV) can be downloaded from http://genomics.tacc.utexas.edu/data/rte_methylation_spreading/ .

**Bisulfite sequencing**: DNA was extracted from the outer tissues of B73 ears whose silks had emerged but had not been fertilized. Sodium bisulfite-treated Illumina sequencing libraries were prepared using a method similar to that of Lister et al (Lister et al., 2009). Alignment to the genome (AGPv2) and identification of methylated cytosines was performed using BS Seeker (Chen et al., 2010). A total of 198,333,982 single-end reads with unique alignments specifically on the ten chromosomes were obtained, with an average genome-matching read length of 72.8 bases (7.0X coverage, SRA accession SRA050144.1). The level of methylation in CG, CHG and CHH contexts and the total proportion of DNA methylation was calculated for non-repeat masked sequences (as annotated within ZmB73_5a_MTEC_repeats) located within 1kb of each retrotransposon family. Percent methylation is defined as the number of methylated Cs per total number of Cs for a region. BEDTools (Quinlan & Hall, 2010) was used to identify low-copy sequences flanking retrotransposons.

**Identification and analysis of empty sites:** Approximately 63M Mo17 454 whole-genome shotgun sequencing reads generated by the DOE's Joint Genome Institute (JGI) were trimmed and aligned to Maize B73 reference genome (AGPv2) and reads aligned uniquely (single loci) were filtered for subsequent analysis. A retrotransposon insertion site was classified as empty if we identified at least 3 WGS reads supporting the site that aligned to the insertion site that included>50 bp of aligned sequence outside of the repeat region in B73 with similarity of ≥94%, relatively short unaligned tails (≤20 bp), and contained a long overhang of >20 bp that begins ±3 bp from the annotated retrotransposon insertion site. PCR primers were designed to amplify the sequence at the "empty" sites using the B73 sequence (which contained the insertion) and the Mo17 sequence (which lacks the insertion). These same primers were also used to assess the presence or absence of the insertion in several other maize genotypes including CML228, CML277, Hp301, Tx303 and Oh7b. Seeds for these genotypes were obtained from the USDA

North Central Regional Plant Introduction Station. PCR and gel electrophoresis was conducted as described (Swanson-Wagner et al., 2010).

**RNA-seq and expression analysis:** RNA-seq was performed on three biological replicates of four tissues (3$^{rd}$ leaf, embryo, endosperm, and immature ear) for both B73 and Mo17. Samples were prepared at the University of Minnesota BioMedical Genomics Center in accordance with the TruSeq library creation protocol (Illumina). Samples were sequenced on the HiSeq 2000 developing 6-17 million reads per replicate. Raw reads were filtered to eliminate poor quality reads using CASAVA (Illumina). Transcript abundance was calculated by mapping reads to the maize reference genome (AGPv2) using TopHat under standard parameters (Trapnell et al., 2010). Counts of mapped reads across the exon space of the maize genome reference working gene set (ZmB73_5a) were developed using 'BAM to Counts' within the iPlant Discovery Environment (www.iplantcollaborative.org). RPKM values were calculated per gene. All genes within 500, 1000, 2500, and 5000 bases of the closest upstream annotated transposable element (ZmB73_5a) using BEDtools (Quinlan & Hall, 2010) were grouped by the spreading class of the nearest TE: spreading (5mc / H3K9), spreading (H3K9 only), non-spreading, and no TE within distance. Genes were also classified as expressed for any RPKM value > 0. The proportions of genes showing expression for each distance and spreading class combination were calculated. Average RPKM values for each distance and spreading class combination were also calculated. Significance testing was performed non-parametrically through Wilcox rank-sum tests. Sequencing data is available from the NCBI short read archive under studies SRP013432 and SRP009313.

**REFERENCES**


Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H (2011) Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Research 39: 6919–6931.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, et al. (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442: 295–298.

Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, et al. (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proceedings of the National Academy of Sciences of the United States of America 109: 10444-10449

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genetics 5: e1000732.

Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome / National Research Council Canada 37: 565–576.

Bernatavichute Y V, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana. PloS ONE 3: e3156.

Biémont C (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186: 1085–1093.

Biémont C, Vieira C (2006) Junk DNA as an evolutionary force. Nature 443: 521–524.

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. The Plant Cell 17: 343–360.

Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11: 203.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.

Comfort NC (2001) From controlling elements to transposons: Barbara McClintock and the Nobel Prize. Trends in Genetics17: 475–478.

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, et al. (2000) Mediator of Paramutation1 is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. The Plant Cell 12: 2101–2118.

Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. Proceedings of the National Academy of Sciences of the United States of America 106: 19916–19921.

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, et al. (2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics 7: e1002372.

Emberton J, Ma J, Yuan Y, SanMiguel P, Bennetzen JL (2005) Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. Genome Research 15: 1441–1446.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics 9: 397–405.

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proceedings of the National Academy of Sciences of the United States of America 99: 9573–9578.

Gdula DA, Gerasimova TI, Corces VG (1996) Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. Proceedings of the National Academy of Sciences of the United States of America 93: 9378–9383.

Gent JI, 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 Research 40: 1550–1560.

Girard L, Freeling M (1999) Regulatory Changes as a Consequence of Transposon Insertion EXPRESSION QUALITATIVELY. 296: 291–296.

Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature Reviews Molecular Cell Biology 12: 483–492.

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research 19: 1419–1428.

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, et al. (2011) Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proceedings of the National Academy of Sciences of the United States of America 108: 2322–2327.

Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, 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 Genetics 5: e1000737.

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature Genetics 33: 102–106.

Kent WJ (2002) BLAT---The BLAST-Like Alignment Tool. Genome Research 12: 656–664.

Levin HL, Moran J V (2011) Dynamic interactions between transposable elements and their hosts. Nature Reviews Genetics 12: 615–627.

Lippman Z, Gendrel A V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annual Review of Plant Biology 60: 43–66.

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Current Opinion in Plant Biology 14: 156–161.

Lisch D, Carey CC, Dorweiler JE, Chandler VL (2002) A mutation that prevents paramutation in maize also reverses Mutator transposon methylation and silencing. Proceedings of the National Academy of Sciences of the United States of America 99: 6130–6135.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.

Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes & Development 18: 2873–2878.

Makarevitch I, Stupar RM, Iniguez AL, Haun WJ, Barbazuk WB, et al. (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase zmet2. Genetics 177: 749–760.

Martienssen R, Barkan A, Taylor WC, Freeling M (1990) Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. Genes & Development 4: 331–343.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135–1138.

McClintock B (1983) The significance of responses of the genome to challenge. Science 226: 792–801.

Meyers B, Tingey S, Morgante M (2001) Abundance, Distribution, and Transcriptional Activity of Repetitive Elements in the Maize Genome. Genome Research 11: 1660–1676.

Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, et al. (1994) Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes & Development 8: 1463–1472.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23: 314–318.

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, et al. (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). The Plant Journal 52: 391–404.

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, et al. (2003) Maize genome sequencing by methylation filtration. Science 302: 2115–2117.

Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM (2001) Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. The Plant Cell 13: 1919–1928.

Pereira V, Enard D, Eyre-Walker A (2009) The effect of transposable element insertions on gene expression evolution in rodents. PloS ONE 4: e4321.

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

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, et al. (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genetics 23: 305–308.

Rakyan V, Whitelaw E. (2003) Transgenerational epigenetic inheritance. Current Biology 13: 6.

Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, et al. (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genetics 7: e1002301.

Richards EJ (2006) revisiting soft inheritance. 7: 395–402.

SanMiguel P, C. V (2009) The LTR-retrotransposons of maize. Handbook of maize: Genetics and genomcs.

Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. The EMBO Journal 26: 3641–3652.

Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. Science 319: 462–465.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Smyth GK (2005) Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor . New York: Springer. 397–420.

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, et al. (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular Cell 6: 791–802.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research 20: 1689–1699.

Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28: 511–515.

Vicient CM (2010) Transcriptional activity of transposable elements in maize. BMC genomics 11: 601.

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proceedings of the National Academy of Sciences of the United States of America 103: 17644–17649.

Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Current Opinion in Genetics & Development 18: 188–192.

Whitelaw C a, Barbazuk WB, Pertea G, Chan a P, Cheung F, et al. (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302: 2118–2120.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. (2007) A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8: 973–982.

Table 1. Attributes and classification of retrotransposon families. A list of 145 retrotransposon families with information and statistics about each family in the B73 genotype.

| Family | Heterochromatin spreading? | Spreading distance (bp) | Order[a] | Superfamily[a] | # of genomic fragments | # of unique array probes flanking this family | Mb in B73 genome | # fragments, homology search | Avg length, homology search | # FL elements, structural search[a] | Avg length, structural search[a] | Avg insertion date (mya), FL elements[a] | Jia et al (2009) expression in Mop1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| opie | 5mC and H3K9 | 1400 | LTR | [RLC] Copia | 183843 | 48105 | 178.171 | 159512 | 1117 | 3530 | 8888 | 0.78 | Both |
| ji | 5mC and H3K9 | 1400 | LTR | [RLC] Copia | 160225 | 49848 | 225.818 | 127484 | 1771 | 4093 | 9523 | 0.77 | up |
| iteki | 5mC and H3K9 | 1200 | LTR | [RLC] Copia | 23681 | 8727 | | | | | | | up |
| ruda | 5mC and H3K9 | 800 | LTR | [RLC] Copia | 35766 | 14924 | 19.152 | 42455 | 451 | 568 | 6485 | 0.74 | up |
| bygum | 5mC and H3K9 | 800 | LTR | [RLC] Copia | 4395 | 2962 | 2.15 | 3642 | 590 | 49 | 9245 | 2.44 | |
| gori | 5mC and H3K9 | 1200 | LTR | [RLC] Copia | 11060 | 3840 | | | | | | | |
| leviathan | 5mC and H3K9 | 800 | LTR | [RLC] Copia | 7618 | 2432 | | | | | | | |
| vuijon | 5mC and H3K9 | 1200 | LTR | [RLC] Copia | 11478 | 5462 | 0.062 | 116 | 538 | 1 | 9200 | 0.91 | |
| dagaf | 5mC and H3K9 | 1200 | LTR | [RLG] Gypsy | 20866 | 5554 | 15.785 | 13991 | 1128 | 185 | 10955 | 0.95 | Both |
| grande | 5mC and H3K9 | 1200 | LTR | [RLG] Gypsy | 37170 | 13204 | 62.269 | 19303 | 3226 | 1338 | 13796 | 0.56 | Both |
| gyma | 5mC and H3K9 | 1600 | LTR | [RLG] Gypsy | 59653 | 30114 | 64.419 | 39405 | 1635 | 436 | 12797 | 0.92 | Both |
| flip | 5mC and H3K9 | 1600 | LTR | [RLG] Gypsy | 39452 | 13592 | 96.263 | 29485 | 3265 | 716 | 14847 | 0.86 | up |
| doke | 5mC and H3K9 | 1000 | LTR | [RLG] Gypsy | 19889 | 8285 | 43.276 | 19523 | 2217 | 697 | 10630 | 0.74 | up |
| ansuya | 5mC and H3K9 | 800 | LTR | [RLG] Gypsy | 5823 | 1052 | | | | | | | up |
| huck | 5mC and H3K9 | 1200 | LTR | [RLG] Gypsy | 108639 | 56001 | 233.485 | 59208 | 3943 | 3341 | 13407 | 1.1 | up |
| nihep | 5mC and H3K9 | 1000 | LTR | [RLG] Gypsy | 5666 | 2927 | | | | | | | up |
| riiryl | 5mC and H3K9 | 800 | LTR | [RLG] Gypsy | 5820 | 3070 | 0.091 | 100 | 906 | 2 | 5504 | 0.24 | up |
| prem1 | 5mC and H3K9 | 1000 | LTR | [RLG] Gypsy | 109037 | 74508 | 76.954 | 75605 | 1018 | 1479 | 8958 | 0.57 | down |
| cinful-zeon | 5mC and H3K9 | 1000 | LTR | [RLG] Gypsy | 110920 | 49939 | 188.308 | 82429 | 2284 | 9844 | 8202 | 0.6 | down |
| neha | 5mC and H3K9 | 800 | LTR | [RLG] Gypsy | 8913 | 5137 | 2.523 | 8463 | 298 | 39 | 9610 | 2.51 | down |
| xilon-diguus | 5mC and H3K9 | 1400 | LTR | [RLG] Gypsy | 66869 | 36317 | 83.558 | 48297 | 1730 | 197 | 10964 | 0.77 | down |
| gunu | 5mC and H3K9 | 1200 | LTR | [RLX] Unknown | 40878 | 6490 | | | | | | | Both |
| uwum | 5mC and H3K9 | 1600 | LTR | [RLX] Unknown | 17842 | 15461 | 15.802 | 13271 | 1191 | 238 | 8495 | 0.8 | Both |
| bobeg | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 2106 | 1895 | 0.013 | 49 | 270 | 3 | 3077 | 0.19 | up |

| Family | Heterochromatin spreading? | Spreading distance (bp) | Order[a] | Superfamily[a] | # of genomic fragments | # of unique array probes flanking this family | Mb in B73 genome | # fragments, homology search | Avg length, homology search | # FL elements, structural search[a] | Avg length, structural search[a] | Avg insertion date (mya), FL elements[a] | Jia et al (2009) expression in Mop1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| kahoba | 5mC and H3K9 | 1400 | LTR | [RLX] Unknown | 1617 | 3479 | 0.157 | 1295 | 121 | 1 | 2635 | n.d. | up |
| hutu | 5mC and H3K9 | 1000 | LTR | [RLX] Unknown | 5957 | 4255 | 2.368 | 10561 | 224 | 1 | 20168 | 3.14 | up |
| iwik | 5mC and H3K9 | 1000 | LTR | [RLX] Unknown | 10435 | 6132 | 8.451 | 18024 | 469 | 32 | 13874 | 2.29 | up |
| naadira | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 6969 | 5041 | | | | | | | up |
| ojah | 5mC and H3K9 | 1200 | LTR | [RLX] Unknown | 3660 | 12136 | 0.774 | 4272 | 181 | 1 | 9869 | n.d. | up |
| baso | 5mC and H3K9 | 1000 | LTR | [RLX] Unknown | 4890 | 3715 | 1.117 | 4401 | 254 | 65 | 12051 | 2.41 | up |
| mada | 5mC and H3K9 | 1000 | LTR | [RLX] Unknown | 30526 | 15055 | | | | | | | down |
| egisu | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 8975 | 9968 | | | | | | | down |
| odoj | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 9733 | 5401 | | | | | | | down |
| tekay | 5mC and H3K9 | 1600 | LTR | [RLX] Unknown | 20852 | 15801 | 15.864 | 15387 | 1031 | 102 | 12102 | 0.74 | down |
| ubow | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 6625 | 4026 | 4.12 | 6225 | 662 | 57 | 8935 | 2.7 | down |
| dapuuv | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 11065 | 5205 | | | | | | | |
| naiba | 5mC and H3K9 | 1600 | LTR | [RLX] Unknown | 7426 | 11779 | 1.394 | 6872 | 203 | 1 | 21781 | 0.92 | |
| odip | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 1432 | 2117 | 0.238 | 1267 | 188 | 8 | 3906 | 0.56 | |
| ugymos | 5mC and H3K9 | 800 | LTR | [RLX] Unknown | 996 | 1432 | 0.264 | 1221 | 216 | 1 | 5243 | n.d. | |
| tufe | H3K9 | 1400 | LTR | [RLC] Copia | 4625 | 2185 | 0.033 | 316 | 104 | 2 | 13464 | 6.43 | up |
| nida | H3K9 | 800 | LTR | [RLC] Copia | 4106 | 1907 | | | | | | | up |
| giepum | H3K9 | 800 | LTR | [RLC] Copia | 30580 | 22476 | 27.823 | 28737 | 968 | 186 | 12387 | 0.71 | down |
| dijap | H3K9 | 1000 | LTR | [RLC] Copia | 7461 | 11134 | 3.667 | 5288 | 693 | 5 | 11505 | 1.85 | |
| tiwewi | H3K9 | 800 | LTR | [RLC] Copia | 2167 | 1015 | 0.039 | 66 | 593 | 2 | 14210 | 0.57 | |
| ywyt | H3K9 | 800 | LTR | [RLG] Gypsy | 2979 | 1724 | 2.643 | 3015 | 876 | 2 | 11087 | 3 | up |
| wihov | H3K9 | 1000 | LTR | [RLX] Unknown | 2465 | 1334 | 0.617 | 548 | 1127 | 2 | 15043 | 1.4 | down |
| demo | H3K9 | 1000 | LTR | [RLX] Unknown | 270 | 1712 | 0.031 | 185 | 168 | 1 | 1742 | 1.6 | down |
| vusu | H3K9 | 800 | LTR | [RLX] Unknown | 318 | 1217 | 0.113 | 400 | 284 | 1 | 7071 | n.d. | down |
| emuh | H3K9 | 800 | LTR | [RLX] Unknown | 231 | 1440 | 0.022 | 167 | 133 | 1 | 1281 | n.d. | |
| etiti | Non-spreading | | LINE | [RIL] L1 | 941 | 2491 | | | | | | | |
| afeda | Non-spreading | | LINE | [RIX] Unknown | 1960 | 6873 | | | | | | | down |
| pyelum | Non-spreading | | LINE | [RIX] Unknown | 1636 | 5740 | | | | | | | down |
| keneat | Non-spreading | | LINE | [RIX] Unknown | 307 | 1195 | | | | | | | down |
| loneok | Non-spreading | | LINE | [RIX] Unknown | 1572 | 6064 | | | | | | | down |
| totyru | Non-spreading | | LINE | [RIX] Unknown | 365 | 1234 | | | | | | | down |
| edaej | Non-spreading | | LINE | [RIX] Unknown | 836 | 2234 | | | | | | | down |

| Family | Heterochromatin spreading? | Spreading distance (bp) | Order[a] | Superfamily[a] | # of genomic fragments | # of unique array probes flanking this family | Mb in B73 genome | # fragments, homology search | Avg length, homology search | # FL elements, structural search[a] | Avg length, structural search[a] | Avg insertion date (mya), FL elements[a] | Jia et al (2009) expression in Mop1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| leijoh | Non-spreading | | LINE | [RIX] Unknown | 1563 | 4366 | | | | | | | down |
| ekeje | Non-spreading | | LINE | [RIX] Unknown | 1965 | 5166 | | | | | | | down |
| koajav | Non-spreading | | LINE | [RIX] Unknown | 1127 | 2853 | | | | | | | down |
| nugimu | Non-spreading | | LINE | [RIX] Unknown | 2787 | 9747 | | | | | | | down |
| ejoet | Non-spreading | | LINE | [RIX] Unknown | 972 | 2832 | | | | | | | down |
| dager | Non-spreading | | LINE | [RIX] Unknown | 717 | 2288 | | | | | | | |
| ijij | Non-spreading | | LINE | [RIX] Unknown | 774 | 2773 | | | | | | | |
| ogepy | Non-spreading | | LINE | [RIX] Unknown | 466 | 1373 | | | | | | | |
| okor | Non-spreading | | LINE | [RIX] Unknown | 1347 | 3903 | | | | | | | |
| udag | Non-spreading | | LINE | [RIX] Unknown | 376 | 1380 | | | | | | | |
| anar | Non-spreading | | LTR | [RLC] Copia | 877 | 1403 | 0.669 | 1010 | 662 | 5 | 5198 | 0.76 | down |
| eninu | Non-spreading | | LTR | [RLC] Copia | 1379 | 2383 | 1.811 | 1084 | 1670 | 64 | 7127 | 0.61 | up |
| sawujo | Non-spreading | | LTR | [RLC] Copia | 5074 | 6636 | 0.347 | 2395 | 145 | 2 | 19972 | 3.07 | up |
| ekoj | Non-spreading | | LTR | [RLC] Copia | 2258 | 8611 | 0.501 | 1728 | 290 | 2 | 3700 | 0.44 | up |
| gudyeg | Non-spreading | | LTR | [RLC] Copia | 4564 | 5263 | 3.091 | 3151 | 981 | 9 | 6191 | 2 | down |
| guvi | Non-spreading | | LTR | [RLC] Copia | 4494 | 7464 | 0.04 | 58 | 692 | 3 | 5808 | 0.75 | down |
| raider | Non-spreading | | LTR | [RLC] Copia | 1043 | 1046 | 1.972 | 1090 | 1809 | 146 | 5869 | 0.73 | down |
| stonor | Non-spreading | | LTR | [RLC] Copia | 1779 | 3855 | 0.706 | 1453 | 486 | 27 | 6195 | 0.36 | down |
| victim | Non-spreading | | LTR | [RLC] Copia | 991 | 1170 | 0.708 | 949 | 746 | 4 | 4849 | 2.2 | down |
| wamenu | Non-spreading | | LTR | [RLC] Copia | 653 | 1209 | 0.351 | 690 | 509 | 13 | 5265 | 0.26 | down |
| udav | Non-spreading | | LTR | [RLC] Copia | 4578 | 20120 | 0.52 | 3977 | 131 | 4 | 9964 | 0.5 | down |
| wiwa | Non-spreading | | LTR | [RLC] Copia | 5059 | 2643 | 6.782 | 4049 | 1675 | 162 | 7935 | 0.56 | down |
| lusi | Non-spreading | | LTR | [RLC] Copia | 360 | 1056 | 0.136 | 307 | 442 | 1 | 4993 | 0.62 | down |
| debeh | Non-spreading | | LTR | [RLC] Copia | 2090 | 2342 | 0.77 | 2050 | 376 | 27 | 6711 | 0.83 | |
| donuil | Non-spreading | | LTR | [RLC] Copia | 1831 | 2279 | 0.558 | 1303 | 428 | 3 | 10759 | 1.67 | |
| ibulaf | Non-spreading | | LTR | [RLC] Copia | 970 | 1707 | 0.8 | 813 | 984 | 1 | 4625 | 1.51 | |
| iwim | Non-spreading | | LTR | [RLC] Copia | 1064 | 1453 | 0.074 | 73 | 1016 | 2 | 6637 | 0.11 | |
| labe | Non-spreading | | LTR | [RLC] Copia | 382 | 1040 | 0.171 | 332 | 516 | 1 | 3005 | 2.49 | |
| machiavelli | Non-spreading | | LTR | [RLC] Copia | 4420 | 7554 | 3.638 | 3451 | 1054 | 1 | 5701 | 2.12 | |
| totu | Non-spreading | | LTR | [RLC] Copia | 663 | 1470 | 0.258 | 688 | 375 | 2 | 4965 | 5.77 | |
| ajajog | Non-spreading | | LTR | [RLG] Gypsy | 2529 | 1027 | 0.05 | 64 | 782 | 1 | 5469 | n.d. | up |
| fourf | Non-spreading | | LTR | [RLG] Gypsy | 5263 | 6706 | 3.707 | 4039 | 918 | 19 | 6259 | 0.62 | down |

| Family | Heterochromatin spreading? | Spreading distance (bp) | Order[a] | Superfamily[a] | # of genomic fragments | # of unique array probes flanking this family | Mb in B73 genome | # fragments, homology search | Avg length, homology search | # FL elements, structural search[a] | Avg length, structural search[a] | Avg insertion date (mya), FL elements[a] | Jia et al (2009) expression in Mop1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lata | Non-spreading | | LTR | [RLG] Gypsy | 5168 | 4950 | 1.88 | 3996 | 470 | 1 | 9068 | 4.57 | down |
| guhis | Non-spreading | | LTR | [RLG] Gypsy | 2283 | 1017 | 2.142 | 1742 | 1230 | 90 | 7582 | 1.05 | down |
| gyte | Non-spreading | | LTR | [RLG] Gypsy | 3975 | 20245 | 0.507 | 3556 | 142 | 5 | 4224 | 2.6 | down |
| bosohe | Non-spreading | | LTR | [RLG] Gypsy | 1891 | 3096 | 1.507 | 1461 | 1032 | 106 | 5182 | 0.59 | |
| tuteh | Non-spreading | | LTR | [RLX] Unknown | 1586 | 5106 | 0.451 | 1086 | 415 | 10 | 1911 | 1.63 | up |
| leso | Non-spreading | | LTR | [RLX] Unknown | 6092 | 22892 | 0.602 | 4380 | 137 | 2 | 1373 | 3.28 | up |
| kawivo | Non-spreading | | LTR | [RLX] Unknown | 3031 | 6449 | 0.542 | 3422 | 158 | 1 | 11243 | n.d. | up |
| japov | Non-spreading | | LTR | [RLX] Unknown | 869 | 2144 | 0.375 | 713 | 526 | 2 | 7452 | 1.01 | up |
| dugiab | Non-spreading | | LTR | [RLX] Unknown | 7666 | 8682 | 2.809 | 10178 | 276 | 6 | 5548 | 3.37 | up |
| naseup | Non-spreading | | LTR | [RLX] Unknown | 7438 | 6686 | 3.123 | 7202 | 434 | 9 | 12057 | 2.2 | up |
| milt | Non-spreading | | LTR | [RLX] Unknown | 32278 | 34300 | 21.552 | 16341 | 1319 | 599 | 6308 | 1.18 | down |
| nuhan | Non-spreading | | LTR | [RLX] Unknown | 11634 | 30071 | 4.64 | 17550 | 264 | 1 | 14219 | 6.41 | down |
| uwub | Non-spreading | | LTR | [RLX] Unknown | 5926 | 36722 | 0.565 | 5910 | 96 | 1 | 11064 | 0.72 | down |
| yreud | Non-spreading | | LTR | [RLX] Unknown | 5100 | 5975 | 1.726 | 5927 | 291 | 24 | 2832 | 2.76 | down |
| bumy | Non-spreading | | LTR | [RLX] Unknown | 319 | 1288 | 0.057 | 244 | 233 | 3 | 5386 | 1.63 | down |
| panen | Non-spreading | | LTR | [RLX] Unknown | 805 | 3277 | 0.265 | 1549 | 171 | 1 | 2925 | n.d. | down |
| vedi | Non-spreading | | LTR | [RLX] Unknown | 1097 | 4923 | 0.293 | 2444 | 120 | 4 | 3446 | 0.33 | down |
| lyna | Non-spreading | | LTR | [RLX] Unknown | 521 | 1323 | 0.249 | 454 | 549 | 1 | 7435 | 0.21 | down |
| pute | Non-spreading | | LTR | [RLX] Unknown | 1107 | 2339 | 0.478 | 982 | 487 | 2 | 5108 | 1.6 | down |
| ebeg | Non-spreading | | LTR | [RLX] Unknown | 4161 | 4900 | | | | | | | down |
| votaed | Non-spreading | | LTR | [RLX] Unknown | 4185 | 7414 | 0.021 | 137 | 157 | 1 | 16653 | 1.85 | down |
| ebel | Non-spreading | | LTR | [RLX] Unknown | 5455 | 7912 | 4.27 | 5436 | 786 | 150 | 5301 | 0.57 | down |
| ywely | Non-spreading | | LTR | [RLX] Unknown | 689 | 4679 | 0.149 | 1222 | 122 | 1 | 3130 | n.d. | down |
| hiimam | Non-spreading | | LTR | [RLX] Unknown | 3330 | 9526 | 1.612 | 6528 | 247 | 3 | 4583 | 2.75 | down |
| pibo | Non-spreading | | LTR | [RLX] Unknown | 1227 | 2712 | 0.258 | 1488 | 173 | 1 | 2318 | n.d. | down |
| jelat | Non-spreading | | LTR | [RLX] Unknown | 451 | 1154 | 0.098 | 461 | 213 | 1 | 1602 | n.d. | down |
| arar | Non-spreading | | LTR | [RLX] Unknown | 99 | 1007 | 0.039 | 75 | 524 | 1 | 4148 | 0.52 | down |
| hoda | Non-spreading | | LTR | [RLX] Unknown | 213 | 1770 | 0.032 | 147 | 217 | 1 | 2955 | 1.15 | down |
| lamyab | Non-spreading | | LTR | [RLX] Unknown | 496 | 1408 | 0.181 | 620 | 291 | 88 | 2551 | 0.67 | down |
| afeke | Non-spreading | | LTR | [RLX] Unknown | 882 | 3627 | 0.303 | 2984 | 102 | 1 | 1857 | n.d. | |
| afuv | Non-spreading | | LTR | [RLX] Unknown | 5371 | 22210 | 0.965 | 6027 | 160 | 1 | 14195 | n.d. | |
| avahi | Non-spreading | | LTR | [RLX] Unknown | 3693 | 15972 | 0.545 | 3341 | 163 | 10 | 6230 | 1.58 | |

| Family | Heterochromatin spreading? | Spreading distance (bp) | Order[a] | Superfamily[a] | # of genomic fragments | # of unique array probes flanking this family | Mb in B73 genome | # fragments, homology search | Avg length, homology search | # FL elements, structural search[a] | Avg length, structural search[a] | Avg insertion date (mya), FL elements[a] | Jia et al (2009) expression in Mop1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| bene | Non-spreading | | LTR | [RLX] Unknown | 1558 | 1146 | 0.952 | 1425 | 668 | 3 | 9383 | 3.03 | |
| bogu | Non-spreading | | LTR | [RLX] Unknown | 985 | 2589 | 0.492 | 729 | 675 | 27 | 6733 | 0.66 | |
| buire | Non-spreading | | LTR | [RLX] Unknown | 3513 | 18637 | 0.361 | 3604 | 100 | 1 | 2632 | 7.75 | |
| ewib | Non-spreading | | LTR | [RLX] Unknown | 1245 | 2993 | 0.512 | 794 | 645 | 94 | 2083 | 1.64 | |
| fageri | Non-spreading | | LTR | [RLX] Unknown | 1782 | 2920 | | | | | | | |
| fajy | Non-spreading | | LTR | [RLX] Unknown | 497 | 1563 | 0.097 | 483 | 201 | 3 | 1932 | 3.38 | |
| habu | Non-spreading | | LTR | [RLX] Unknown | 4013 | 5290 | 0.232 | 2865 | 81 | 1 | 1554 | 4.89 | |
| ikukob | Non-spreading | | LTR | [RLX] Unknown | 6987 | 1199 | | | | | | | |
| jeli | Non-spreading | | LTR | [RLX] Unknown | 432 | 1372 | 0.079 | 327 | 243 | 22 | 3097 | 0.42 | |
| kake | Non-spreading | | LTR | [RLX] Unknown | 610 | 1197 | 0.314 | 518 | 606 | 3 | 7372 | 0.22 | |
| loukuv | Non-spreading | | LTR | [RLX] Unknown | 6958 | 22123 | 2.933 | 13645 | 215 | 1 | 9909 | 3.36 | |
| name | Non-spreading | | LTR | [RLX] Unknown | 2868 | 6210 | 0.57 | 2206 | 258 | 212 | 6024 | 0.85 | |
| ogiv | Non-spreading | | LTR | [RLX] Unknown | 277 | 1240 | 0.057 | 356 | 160 | 1 | 5212 | 2.12 | |
| pibi | Non-spreading | | LTR | [RLX] Unknown | 2093 | 1322 | | | | | | | |
| rulo | Non-spreading | | LTR | [RLX] Unknown | 371 | 1690 | 0.046 | 334 | 139 | 1 | 3353 | n.d. | |
| sagyfy | Non-spreading | | LTR | [RLX] Unknown | 1394 | 1862 | | | | | | | |
| small | Non-spreading | | LTR | [RLX] Unknown | 747 | 1137 | 0.25 | 456 | 548 | 4 | 4037 | 0.82 | |
| teuta | Non-spreading | | LTR | [RLX] Unknown | 169 | 1480 | 0.021 | 136 | 154 | 3 | 2216 | 1.55 | |
| ujinas | Non-spreading | | LTR | [RLX] Unknown | 420 | 1020 | 0.125 | 333 | 374 | 1 | 7136 | 0.65 | |
| utar | Non-spreading | | LTR | [RLX] Unknown | 2590 | 12205 | 0.326 | 2728 | 119 | 1 | 1925 | n.d. | |
| vegu | Non-spreading | | LTR | [RLX] Unknown | 2963 | 2001 | 2.728 | 2113 | 1291 | 1 | 6245 | 4.75 | |
| vora | Non-spreading | | LTR | [RLX] Unknown | 2693 | 8134 | 0.263 | 2250 | 117 | 2 | 4285 | 1.62 | |
| wuwe | Non-spreading | | LTR | [RLX] Unknown | 787 | 2453 | 0.109 | 609 | 179 | 3 | 6550 | 0.7 | |
| yraj | Non-spreading | | LTR | [RLX] Unknown | 4323 | 15429 | 0.796 | 3530 | 225 | 32 | 6758 | 0.86 | |

[a]Information for these columns was obtained from Baucom et al., 2009

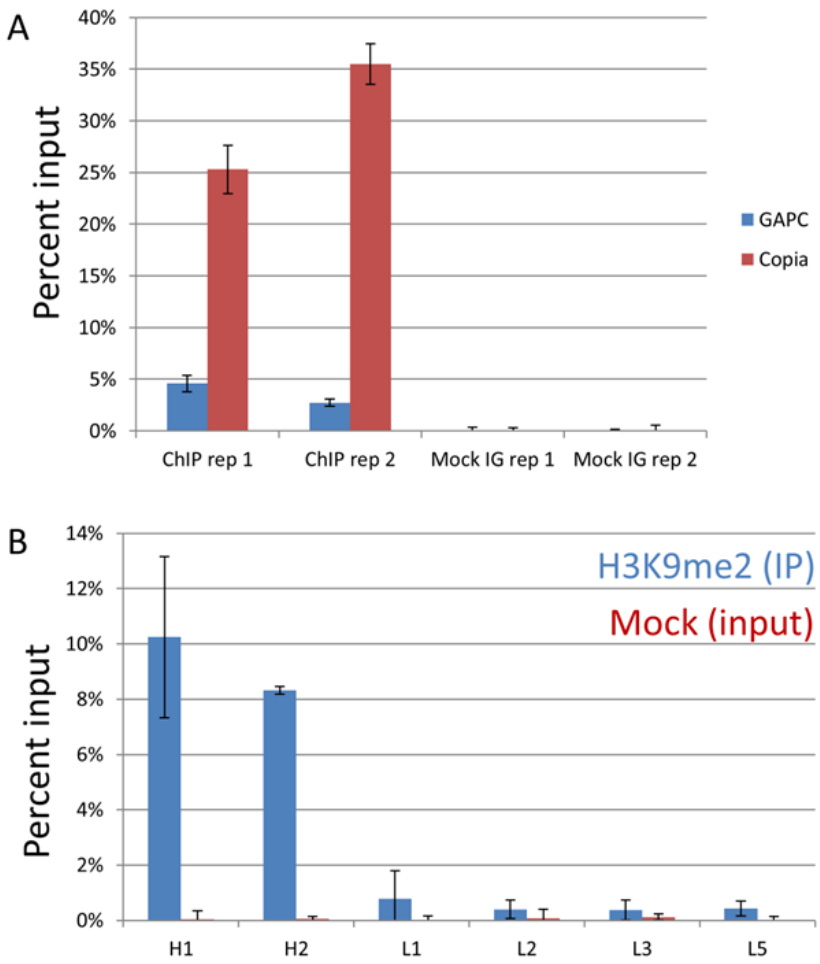



Figure 1: Validation of H3K9me2 ChIP-chip. (A) The efficient enrichment of DNA associated with H3K9me2 was assessed using qPCR. The copia sequence is known to be enriched for H3K9me2 while the GAPC sequence is not associated with H3K9me2 (Haring et al., 2007). Primers for these regions were used to perform qPCR using three technical replicates. The percent of input DNA recovered after IP with the H3K9me2 antibody or a noIG control was determined for both sequences. (B) Several regions were selected for validation based on ChIP-chip profiling. Two regions enriched for H3K9me2 (H1 and H2) and four regions with no evidence for H3K9me2 (L1, L2, L3, L5) were used to design primers for qPCR. The percent of input DNA recovered by ChIP using an H3K9me2 antibody or a noIG control was determined for three replicates of B73 using these primers. The H1 and H2 sequences were enriched by ChIP while the L1, L2, L3 and L5 sequences showed much lower levels of recovery.

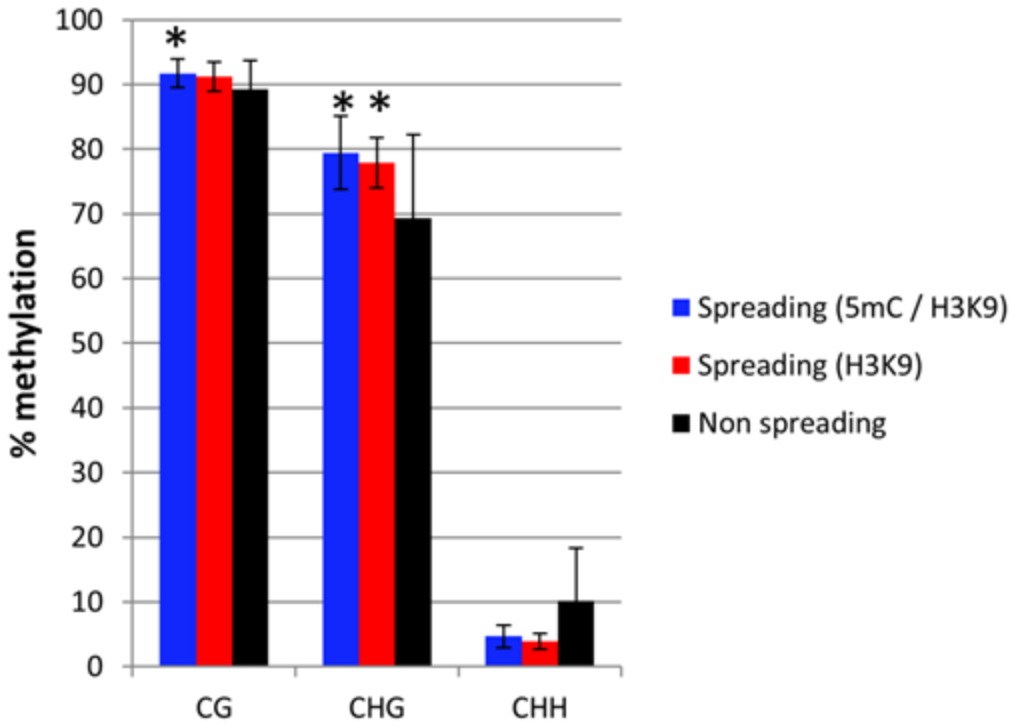



Figure 2: Levels of DNA methylation within retrotransposons. Whole-genome bisulfite sequencing data was used to assess the average level of methylation within retrotransposons. DNA methylation levels within each sequence context (CG, CHG and CHH) were determined for each family of retrotransposon. The average levels of methylation for elements classified as having spreading of both 5mC and H3K9me2, spreading of H3K9me2 only and non-spreading were determined and plotted. The error bars indicate standard deviation among retrotransposon families and "*" indicate significant ($p < 0.001$) differences for a group relative to the non-spreading families. The level of internal methylation at CG sites is similar for retrotransposons with and without spreading of heterochromatin although there is a significant difference in spreading (both) relative to non-spreading families. CHG methylation is slightly lower in non-spreading families. The non-spreading families have slightly elevated levels of CHH methylation relative to the other families.

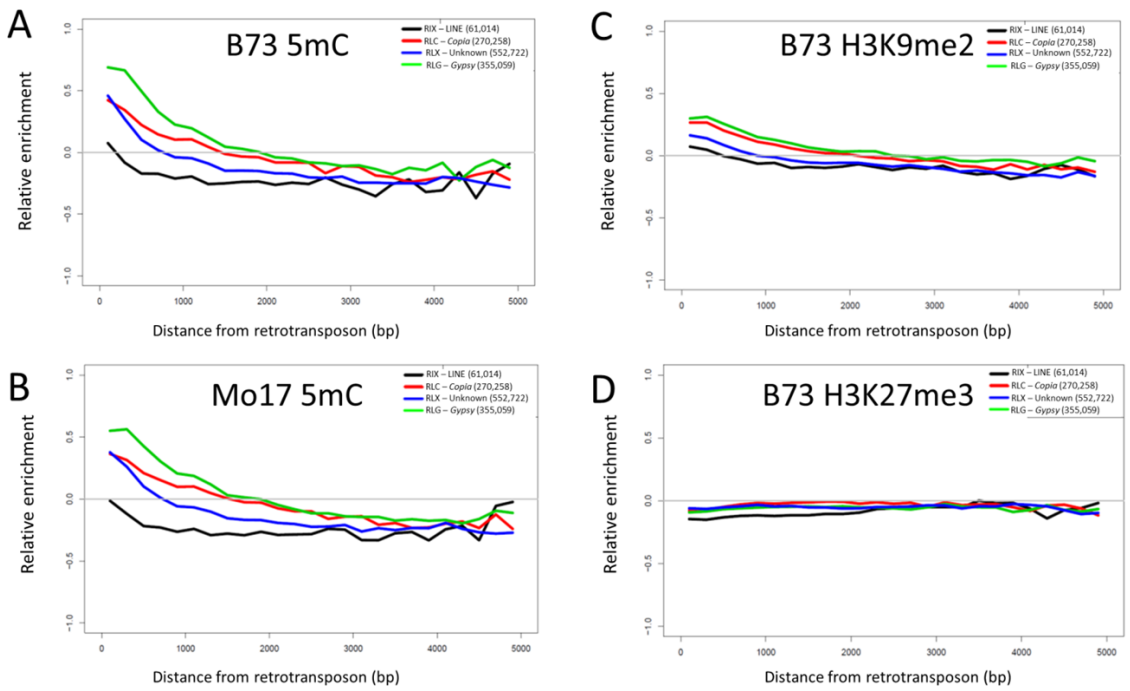



Figure 3: Chromatin modifications in regions flanking maize retrotransposon superfamilies including RIX – LINE (black), RLC – copia (red), RLX – unknown LTR (blue) and RLG – gypsy (green). We identified probes located in low-copy DNA flanking retrotransposons in maize. The number of probes for each class is indicated within the Figure legend. The average level of DNA methylation (A-B), H3K9me2 (C) or H3K27me3 (D) is shown for the 5,000bp adjacent to each superfamily. The level of chromatin modifications are based on ChIP-chip experiments and the y-axis represents the average log ratio for the immunoprecipitated samples relative to genomic input DNA.

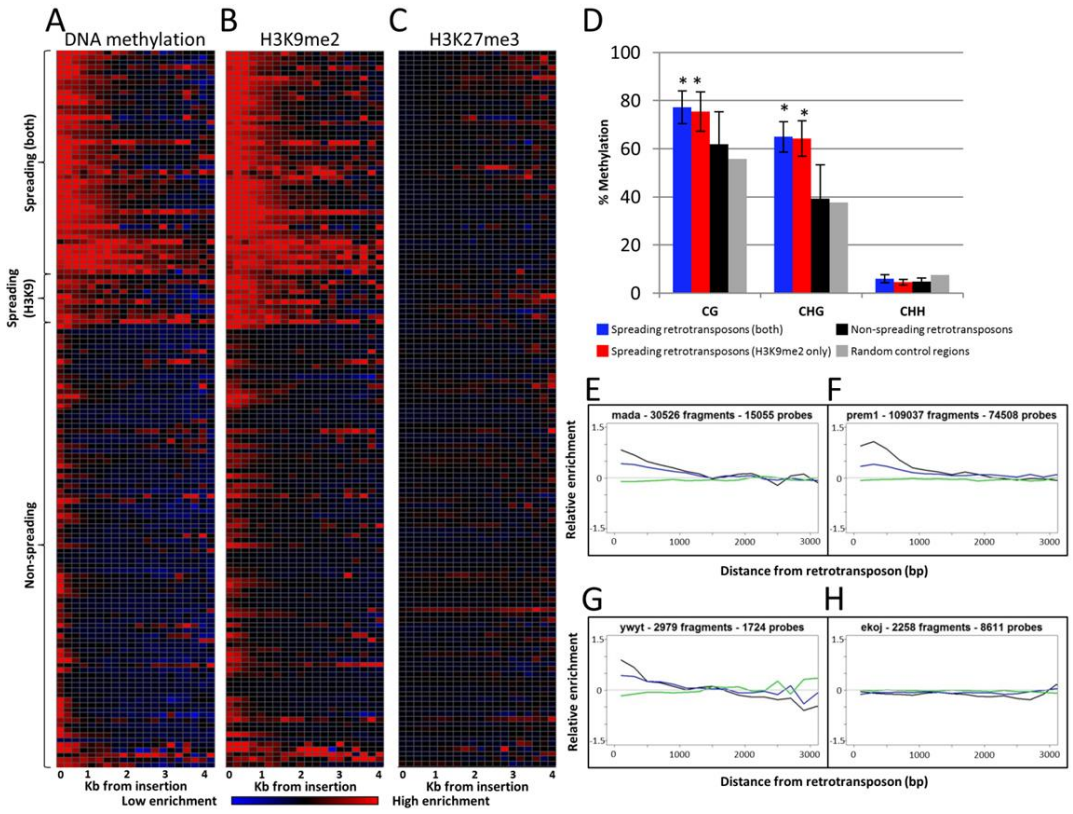



Figure 4: Heterochromatin spreading is restricted to some retrotransposon families. 144 families of class I retrotransposons (RTEs) represented by at least 1000 probes within the adjacent 4 kb were identified and the average level of H3K9me2, 5-methylcytosine and K3K27me3 enrichment was determined for 200 bp bins. The average values are calculated from both sides of annotated elements collapsed. (A) The relative level of DNA methylation in each 200bp bin is shown for each of the 144 families. Red and blue indicate enrichment and depletion of the mark respectively. Black indicates levels of the modification similar to genome-wide average values. The color intensity is based on the average log ratio of immunoprecipitated DNA relative to input DNA. The RTEs are grouped according to whether they show spreading for DNA methylation and H3K9me2, H3K9me2 only or neither of the marks. Similar plots are shown for H3K9me2 (B) and H3K27me3 (C). (D) Whole-genome bisulfite sequencing data was used to assess the methylation level in different cytosine contexts in the low-copy (based on the absence of repetitive sequences) 1 kilobase regions flanking spreading (both), spreading (H3K9), non-spreading RTEs and for 10,000 random genomic regions. The error bars indicate standard deviation among retrotransposon families and "*" indicate significant ($p < 0.001$) differences for a group relative to the non-spreading families. The level of CG and CHG methylation is higher for spreading RTEs than for non-spreading RTEs. (E-H) show plots of representative family(s) for each class. The relative abundance of DNA methylation (black), H3K9me2 (blue) and H3K27me3 (green) is shown for the first 3 kb of low-copy DNA flanking the retrotransposon family on both sides. The y-axis indicates enrichment relative to genome-wide average (value of 0 is equal to average of permuted data).


Figure 5: Profiles of chromatin surrounding spreading (both) retrotransposons. Black lines indicate DNA methylation. Blue lines indicate H3K9. Green lines indicate H3K27 levels. Chromatin values calculated include probes flanking both ends of retrotransposable elements. Copy number of repeat fragments in the B73 reference genome as well as the total number of probes flanking each repeat are displayed. The y-axis provides the distance (in bp) from the retrotransposon insertion.

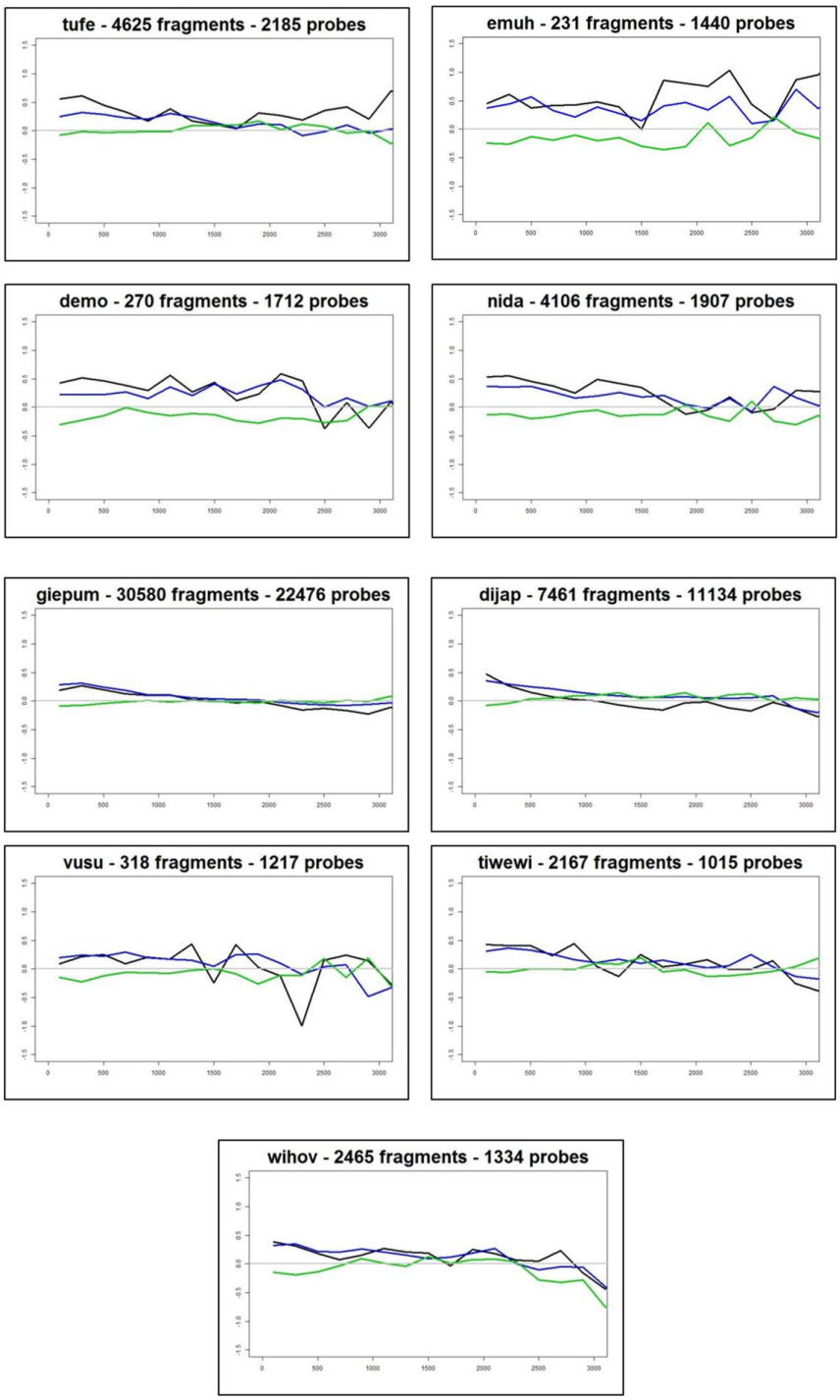



Figure 6: Profiles of chromatin surrounding spreading (H3K9) retrotransposons. Black lines indicate DNA methylation. Blue lines indicate H3K9. Green lines indicate H3K27 levels. Chromatin values calculated include probes flanking both ends of retrotransposable elements. Copy number of repeat fragments in the B73 reference genome as well as the total number of probes flanking each repeat are displayed. The y-axis provides the distance (in bp) from the retrotransposon insertion.

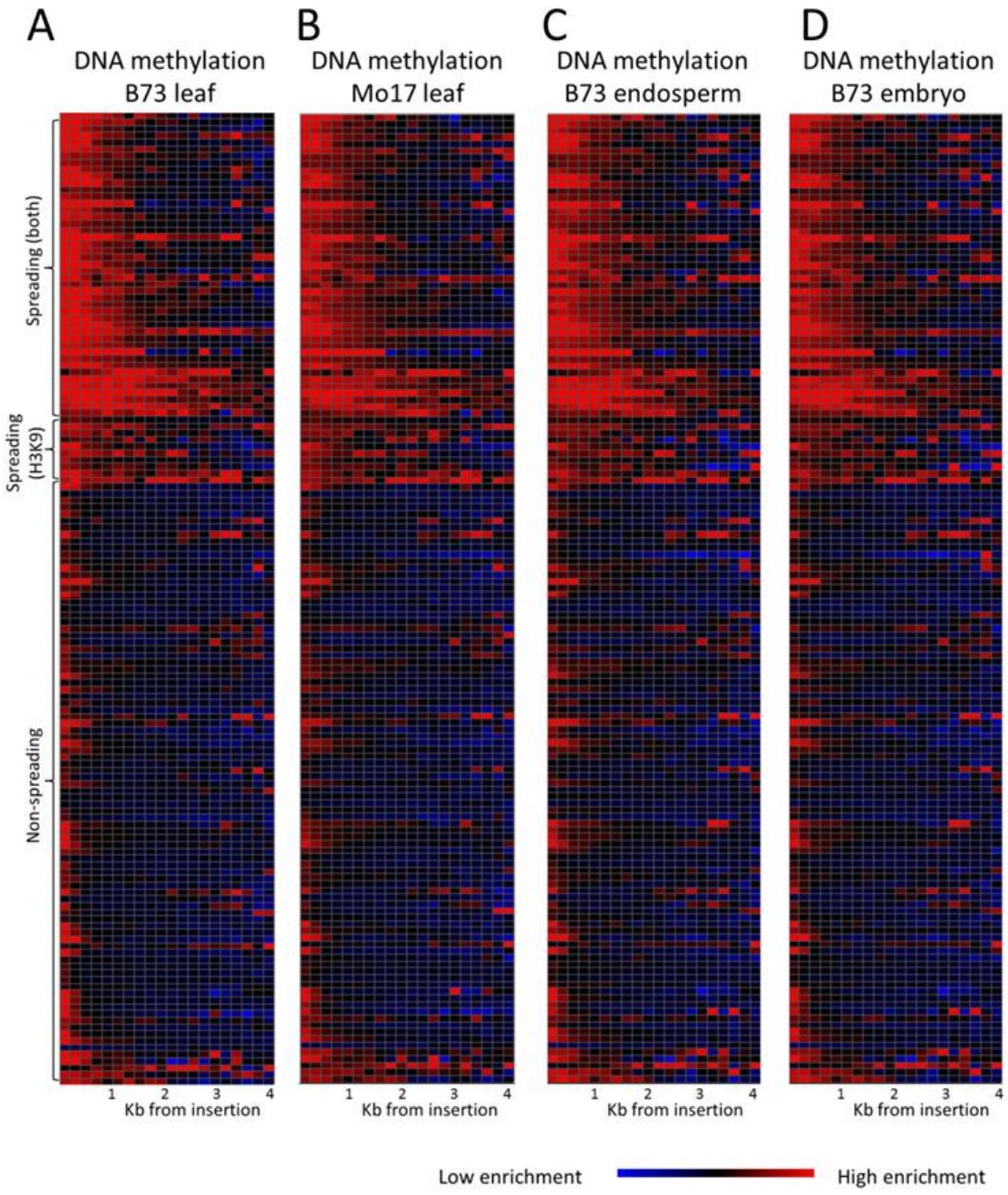



Figure 7: Similar profiles of DNA methylation enrichment adjacent to retrotransposon families in different tissues and genotypes of maize. The relative level of DNA methylation in each 200bp bin is shown for each of the 144 retrotransposon families. The red color indicates enrichment for the modification while blue indicates depletion of the mark. Black indicates levels of the modification similar to genome-wide average values. The color intensity is based on the average log ratio of immunoprecipitated DNA relative to input DNA. The retrotransposons are grouped according to whether they show spreading for DNA methylation and H3K9me2, just H3K9me2 or neither of the marks. The profiles are shown for B73 leaf (A), Mo17 leaf (B), B73 endosperm (C) and B73 embryo (D).

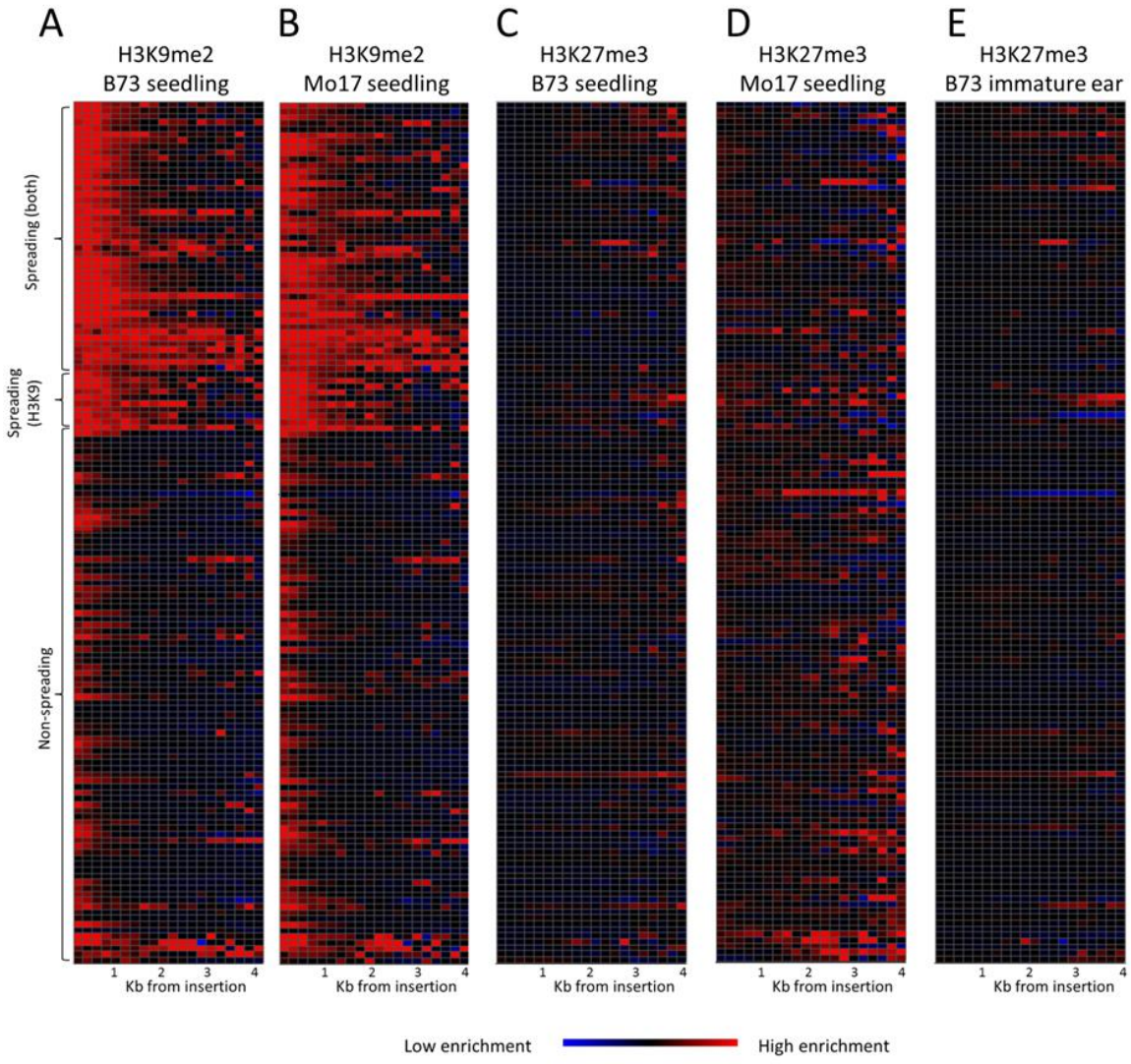



Figure 8: Histone modification patterns in different genotypes. The profile of several histone modifications is shown for both B73 and Mo17. The relative level of H3K9(di)- or H3K27(tri)-methylation in each 200bp bin is shown for each of the 144 retrotransposon families. Modification levels calculated include probes flanking both ends of retrotransposable elements. The red color indicates enrichment for the modification while blue indicates depletion of the mark. Black indicates levels of the modification similar to genome-wide average values. The color intensity is based on the average log ratio of immunoprecipitated DNA relative to input DNA. The retrotransposons are grouped according to whether they show spreading for DNA methylation and H3K9me2, just H3K9me2 or neither of the marks. (A) and (B) display the profiles for H3K9me2 in B73 and Mo17 seedlings, respectively. (C) and (D) display the profiles for H3K27me3 in B73 and Mo17 seedlings, respectively. (E) shows the H3K27me3 profile for immature ear tissue from B73.

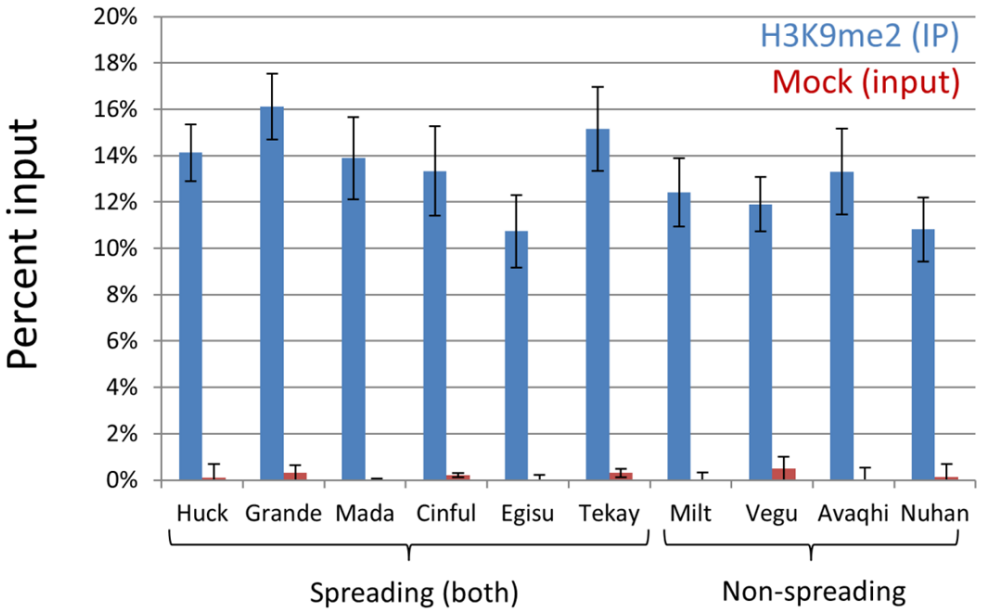



Figure 9: Similar levels of H3K9me2 within spreading and non-spreading retrotransposons. The ChIP-chip assay does not provide information on the abundance of H3K9me2 within repetitive regions. In order to assess whether the level of H3K9me2 was similar within retrotransposons with, and without, spreading we designed primers for internal sequences of 10 retrotransposon families including six with spreading (both) and four that were classified as non-spreading. The qPCR protocol described by Haring et al. (2007) was used to assess the percent input DNA recovered by immunoprecipitation. The percent of input DNA recovered after IP with the H3K9me2 antibody or a noIG control was determined for both sequences and the standard deviation is indicated with error bars. There were high levels of H3K9me2 in each of these retrotransposons but there was no significant difference between the spreading and non-spreading retrotransposons.

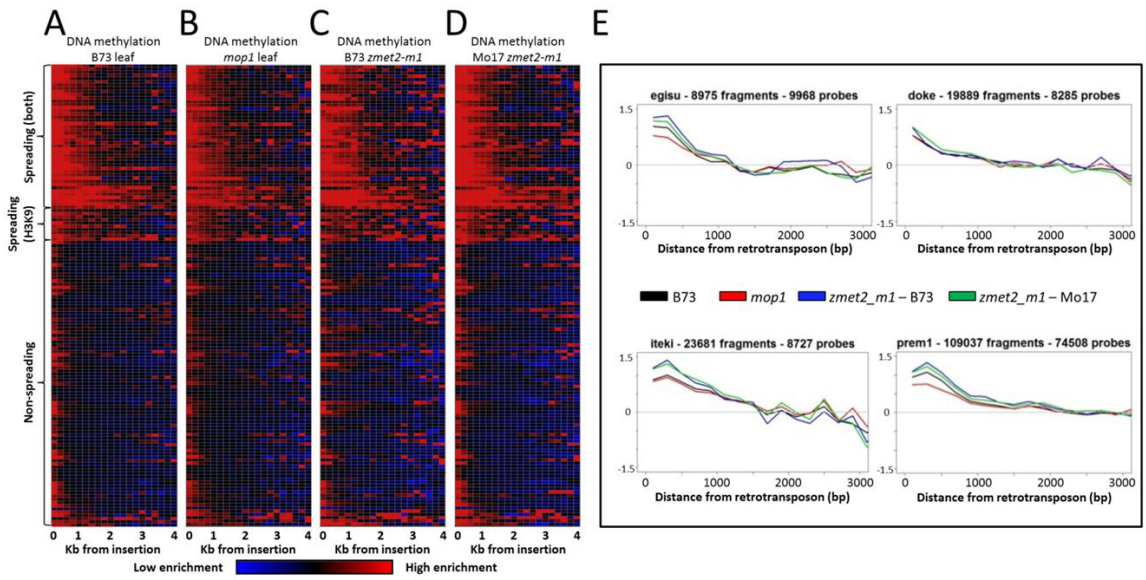



Figure 10: DNA methylation enrichment near retrotransposons is not affected by mop1 or zmet2-m1 mutations. The relative level of DNA methylation in each 200bp bin is shown for each of the 144 retrotransposon families. Modification levels calculated include probes flanking both ends of retrotransposns. The red color indicates enrichment for the modification while blue indicates depletion of the mark. Black indicates levels of the modification similar to genome-wide average values. The retrotransposons are grouped according to whether they show spreading for DNA methylation and H3K9me2, just H3K9me2 or neither of the marks. The profiles are shown for B73 leaf (A), homozygous *mop1* mutant (B), homozygous mutant *zmet2-m1* in B73 (C) and Mo17 (D) genetic backgrounds. Example plots of methylation levels in the same backgrounds (E) indicate minimal differences between mutant backgrounds on spreading.

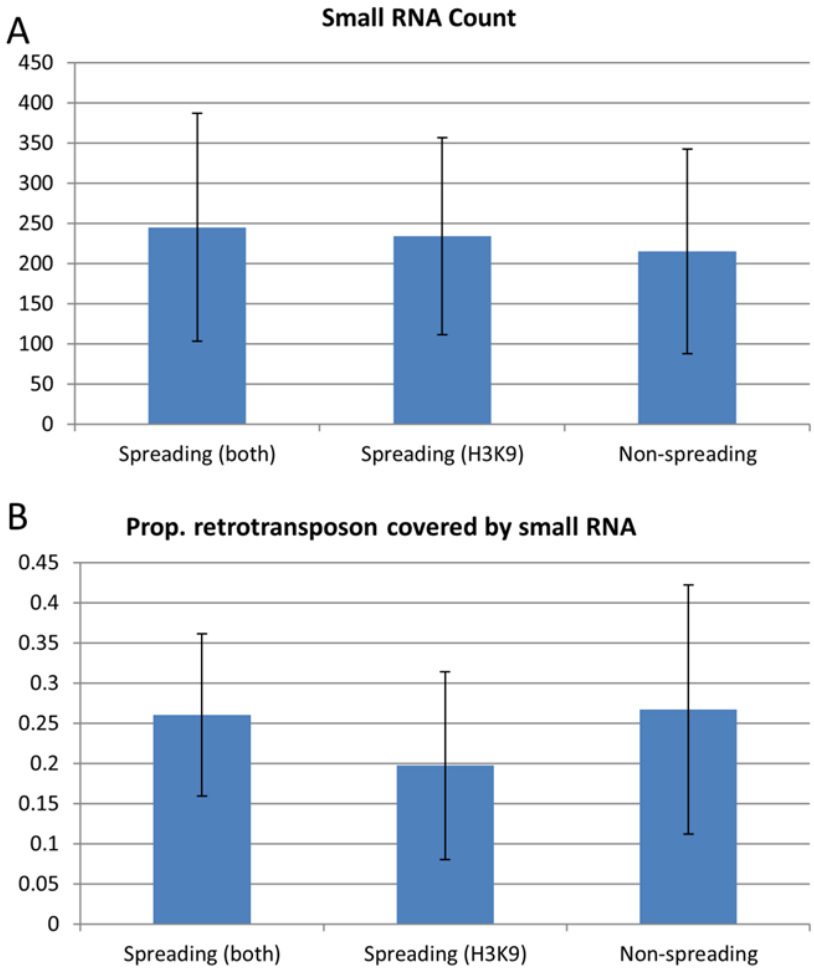



Figure 11: Small RNA coverage of retrotransposon families. B73 shoot small RNAs (Barber et al., 2012). Small RNA reads for all size classes were mapped to the maize reference genome (AGPv2) using BLAT (Kent 2002) under standard parameters. Coverage of small RNA reads over annotated maize transposons (ZmB73_5a) were calculated using BEDtools coverageBED (Quinlan & Hall, 2010). (A) The average small RNA count per retrotransposon family was determined for each of the 144 retrotransposon families. The average count (and standard deviation) for families classified as spreading (both), spreading (H3K9) and non-spreading was then determined. There is no evidence for significant differences in the average small RNA count. (B) The proportion of each retrotransposon covered by small RNAs was then determined in a similar fashion. There is no significant difference in the proportion of coverage for the spreading and non-spreading families of retrotransposons. Small RNA data were downloaded from GEO as samples GSM918108 (Barber et al., 2012).

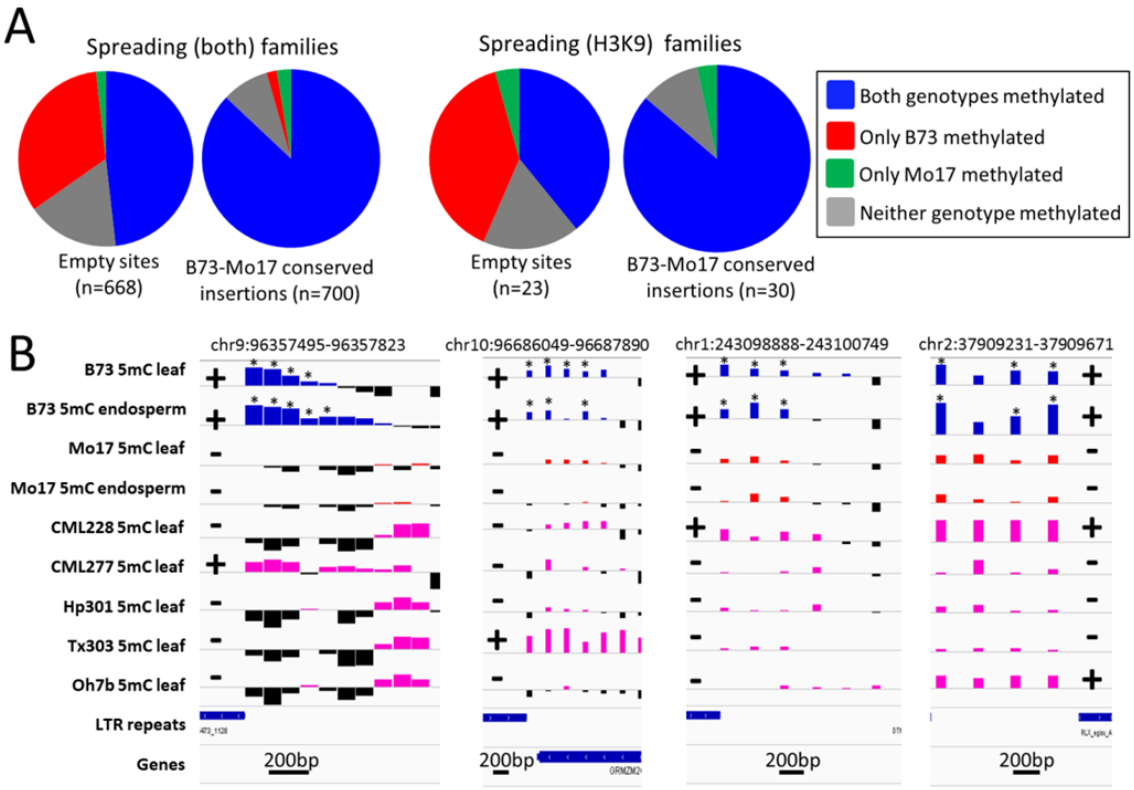



Figure 12: Heterochromatic marks are associated with presence of retrotransposons. (A) Whole genome shotgun sequence data for Mo17 was used to identify retrotransposon insertions that are present within the B73 haplotype but missing in Mo17. The relative level of DNA methylation at the probe nearest the empty site was assessed in B73 and Mo17 for empty sites of spreading (both) and for spreading (H3K9) retrotransposon families. For each set of families we compared the distribution of methylation patterns to a similar number of insertion sites that are conserved in both B73 and Mo17. (B) The DNA methylation levels for four of these "empty" sites (the coordinates specify the retrotransposon present in B73) are shown for two tissues of B73 (blue bars) and Mo17 (red bars) as well as for a single tissue of five other maize genotypes; CML228, CML277, Hp301, Tx303 and Oh7b (pink bars). Black bars for all genotypes indicate depletion of methylation signal. The location of the retrotransposon and its presence or absence are indicated by + or – symbols, respectively. The genotypes containing the insertion of the retrotransposon all exhibit enrichment for DNA methylation while the genotypes without the insertion do not. The "*" indicates probes with significantly ($P<0.01$) higher DNA methylation levels in B73 relative to the same tissue in Mo17. The horizontal gray line indicates the genome-wide average for each of the modifications. Higher values indicate higher levels of DNA methylation and all plots are show on the same scale for the y-axis. The scale for black line near the bottom of each plot indicates the base pair scale for the x-axis.

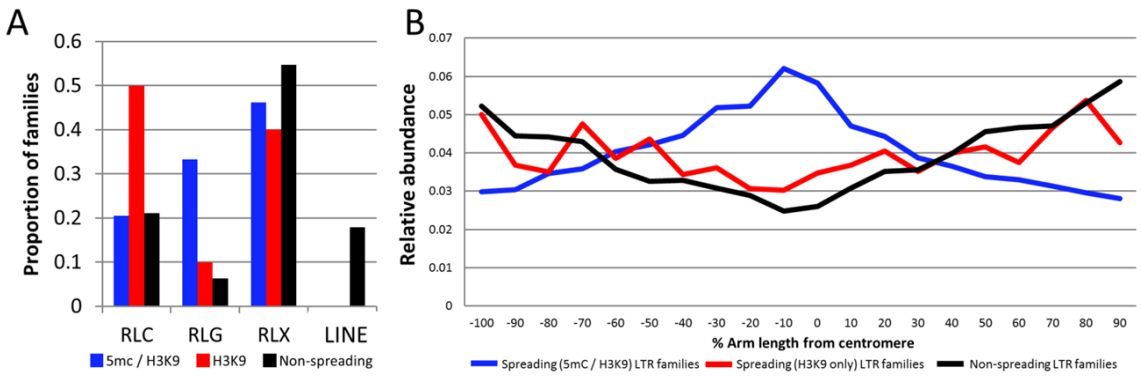



Figure 13: Characterization of retrotransposons that exhibit heterochromatin spreading. (A) The proportion of families within each of the superfamily designations [RLC – copia; RLG – gypsy; RLX – unknown retrotransposon; LINE – LINE elements] is shown for retrotransposons with spreading of both DNA methylation and H3K9me2 (blue), families with H3K9 spreading only (red) and families without spreading (black). (B) The relative abundance of the retrotransposons within each category was determined according to chromosomal position. The retrotransposons in families with spreading (both) are found throughout the maize chromosomes but are enriched in pericentromeric regions relative to the other families. The y-axis provides a normalized estimate of TE abundance along the chromosome (normalized relative to total copy number for each family).

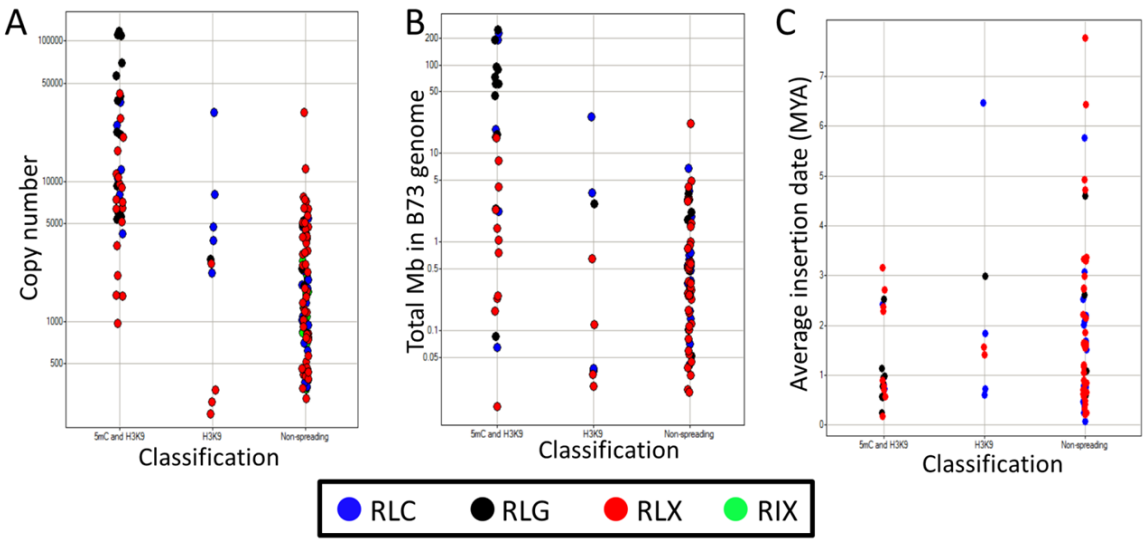



Figure 14: Characteristics of retrotransposon families with spreading of heterochromatic marks. In each of the plots the retrotransposon families are grouped into both (5mC and H3K9), H3K9 only and non-spreading columns and the superfamily is indicated by the color. The data points are jittered to allow visualization of all families. (A) The genomic copy number of each family is shown using a log-scale. The families with spreading of both marks tend to have higher copy numbers. However, there is an overlap in the range of copy number for families with and without spreading. (B) The total Mb of the B73 genome comprised by each family is shown. (C) The average insertion date for each family is plotted. While the non-spreading families include both young and old retrotransposons the families with spreading of both marks tend to be younger families.

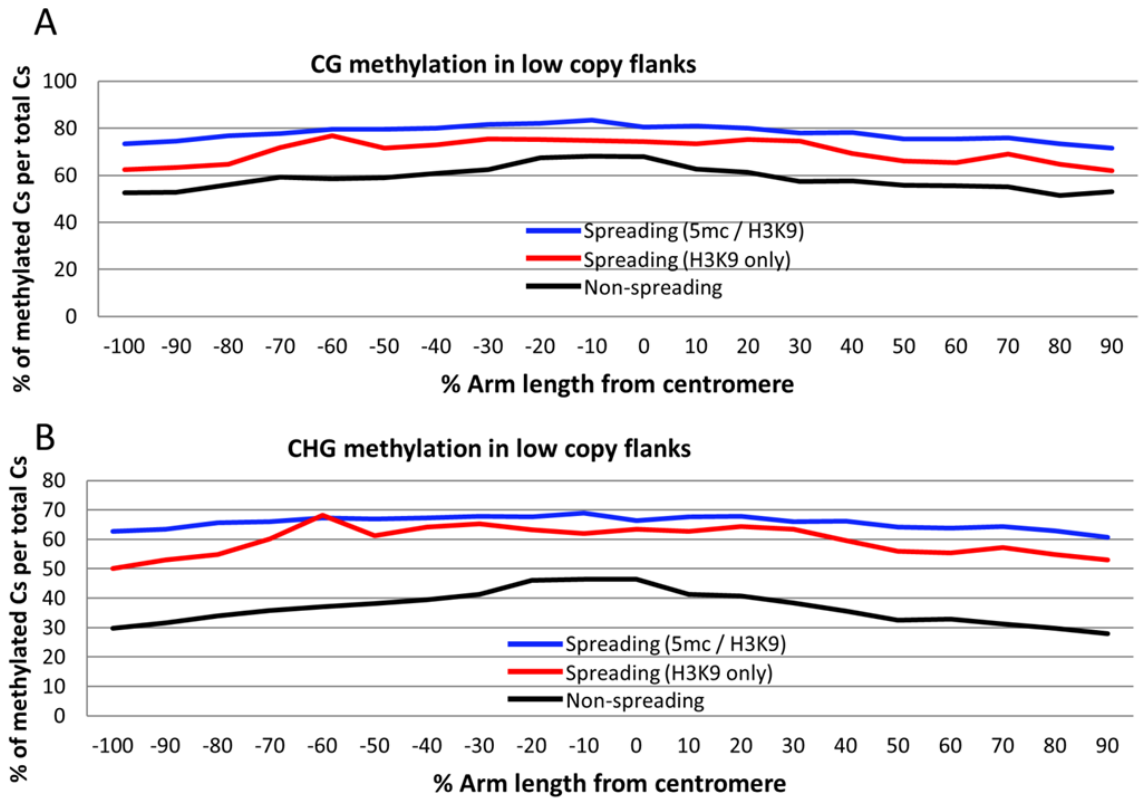



Figure 15: DNA methylation levels in flanking sequences are similar throughout the chromosome. The level of DNA methylation in sequences flanking retrotransposons was determined from bisulfite sequencing data. Only flanking sequences that did not contain any repetitive sequences within 1kb of the retrotransposon were used. The proportion of methylated cytosines in CG (A) or CHG (B) contexts was determined for 1kb of low copy sequences flanking spreading (both), spreading (H3K9) and non-spreading retrotransposon families.

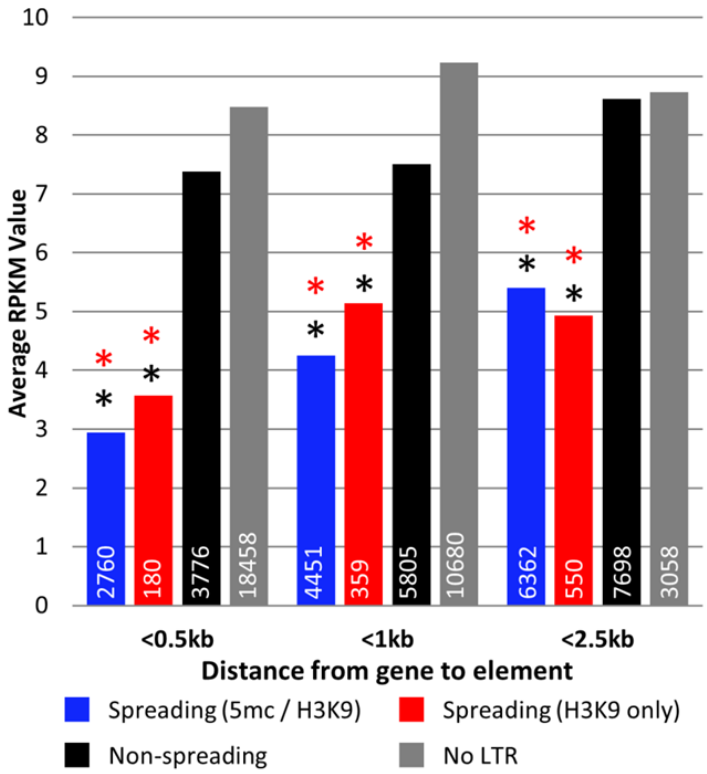



Figure 16: Genes near spreading retrotransposons show lower expression than genes near non-spreading retrotransposons. Average RPKM values (from B73 leaf) for all genes falling within 0.5kb, 1kb, and 2.5kb from the respective class of retrotransposons were developed. White numbers within bars indicate the total number of genes within each category. Red asterisks indicate highly significant ($p<0.001$) difference from the non-spreading retrotransposons (black) within the distance classification. Black asterisks indicate significant ($p<0.001$) difference from genes not near any retrotransposons (grey).

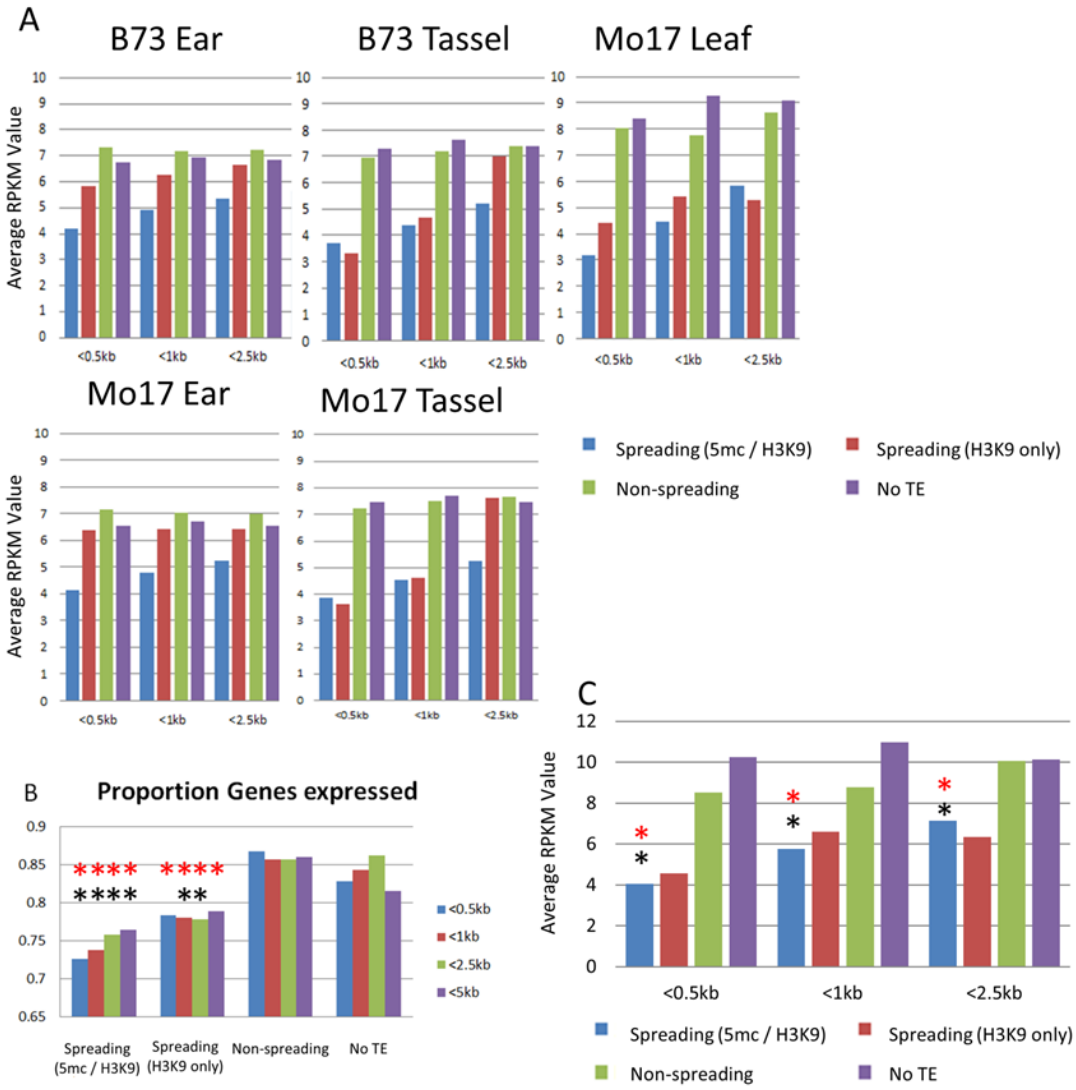



Figure 17: Genes near spreading retrotransposons show lower expression than genes near non-spreading retrotransposons. (A) Average RPKM values (from B73 Ear, B73 Tassel, Mo17 leaf, Mo17 Ear, and Mo17 Tassel tissues) for all genes falling within 0.5kb, 1kb, and 2.5kb from the respective class of retrotransposons were determined. (B) Proportion of genes expressed (RPKM>0) in B73 leaf tissue for genes near 5mc and H3K9 spreading elements, H3k9 spreading only, non-spreading, and no TE nearby. Red asterisks indicate highly significant ($p<0.001$) difference from the non-spreading elements within the distance classification. Black asterisks indicate significant ($p<0.001$) difference from genes not near any TE. (C) Average RPKM values in B73 leaf tissue for expressed genes (all genes with RPKM = 0 were omitted) falling within 0.5kb, 1kb, and 2.5kb from the respective class of retrotransposons were developed. Red asterisks indicate significant ($p<0.05$) difference from the non-spreading elements (green) within the distance classification. Black asterisks indicate significant ($p<0.05$) difference from genes not near any retrotransposons (purple).

Chapter V – **CONTEXT STATEMENT**

DNA methylation is known to show variable states across a population leading to unique regulation of repetitive elements and in some cases genes. Although DNA methylation is often viewed solely as an epigenetic mark, it is still unclear how often DNA methylation state is connected to genetic variation. This study expands on the previous chapters to profile DNA methylation in 51 diverse maize inbred lines and identifies thousands of differentially methylated regions (DMRs). These DMRs are then tested to determine their stability across generations, their association to expression levels for nearby genes, and their association with local genetic polymorphisms. This chapter provides evidence that a substantial proportion of DMRs appear to be associated with the local genetic variation of the individual, and provide possible causes for these associations.

Chapter V entitled 'Epigenetic and genetic influences on DNA methylation variation in maize populations' has been adapted from submitted manuscript:

*Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, Myers CL, Vaughn MW, Nathan M Springer. Epigenetic and genetic influences on DNA methylation variation in maize populations.*

In the development of this chapter, other researchers played roles in the creation of data and assistance to my analysis. SRE , MV, and NS designed the experiment. SRE, QL, RSW, AJW, PJH, and ES conducted the methylation profiling. MV assisted SRE in proper normalization of microarray data to facilitate analysis. SRE, PJH, and the University of Minnesota Genomics Center (UMGC) performed bisulfite sequencing. ES and PTW assisted SRE in computational analysis and data organization. PT, MV, and JS assisted SRE in local genetic association analyses. Text and figures of this chapter were developed by SRE with assistance from NS.

# Chapter V
# Epigenetic and genetic influences on DNA methylation variation in maize populations

DNA methylation is a chromatin modification that is frequently associated with epigenetic regulation in plants and mammals. However, other genetic changes such as transposon insertions also can lead to changes in DNA methylation levels. Genome-wide profiles of DNA methylation levels for 20 maize inbreds were used to discover differentially methylated regions (DMRs). The methylation level for each of these DMRs was also assayed in 31 additional maize genotypes resulting in the discovery of 1,966 common DMRs and 1,754 rare DMRs. Analysis of recombinant inbred lines provides evidence that the majority of DMRs are heritable. A local association scan found that nearly half of the DMRs with common variation are significantly associated with SNPs found within or near the DMR. Many of the DMRs that are significantly associated with local genetic variation are found near transposable elements that may contribute to the DNA methylation variation. The analysis of gene expression in the same samples used for DNA methylation profiling identified over 300 genes with expression patterns that are significantly associated with the DNA methylation variation among genotypes. Collectively, our results suggest that DNA methylation variation is influenced by genetic and epigenetic variation, is often stably inherited and can influence expression level of genes in the population.

**INTRODUCTION**

Heritable information within a species is most commonly found as genetic variation between individuals and populations. Recent studies have restored interest in heritable variation that is not directly connected to DNA sequence polymorphisms, deemed epigenetic variation (Bird 2007). Epigenetic variation is often associated with a variety of chromatin marks such as histone tail modifications, small RNAs, as well as DNA methylation of cytosine within genomic DNA (Jablonka & Raz, 2009). The methylation of cytosines within genomic DNA sequences is one of the most well studied examples of chromatin variation with a large body of work invesTrends in Geneticsating the mechanisms of heritability as well as the actions of particular chromatin states (Stroud et al., 2013).

DNA methylation often acts to suppress the activity of transposable elements, pseudogenes, repetitive sequences, and genes within the genome (Chan et al., 2005; Slotkin & Martienssen 2007). DNA methylation variation has been implicated in processes leading to stable phenotypic variation within plant and animal species (Rasmusson et al., 1997; Richards 2006; Feinberg 2007; Jirtle & Skinner 2007; Vaughn et al., 2007; Richards 2008; Johannes et al., 2008). Well characterized examples of natural variants displaying differing DNA methylation patterns, deemed epialleles, include *Lcyc,* a *cycloidea* homolog causing phenotypic changes in flower morphology of *Linaria vulgaris* (Cubas et al., 1999), the Cnr locus in tomato (Manning et al., 2006), *CmWIP1* in melon (Martin et al., 2009), and FOLT1 (Durand et al., 2012) and PAI (Luff et al., 1999; Melquist et al., 1999) in *Arabidopsis thaliana*. There are also examples of natural variation for epigenetic state associated with paramutation at several loci in maize (Chandler et al., 2000). Many examples of natural variation for DNA methylation exhibit occasional changes of methylation state in the natural environment suggesting the possibility of semi-stable heritability for this epigenetic information (Vaughn et al., 2007; Eichten et al., 2011). There is increased interest in epigenetic phenomenon within plant systems as a possible source for previously unassessed heritable variation for improving phenotypic predictions and breeding methods. In order to fully evaluate the potential of epigenetics for understanding phenotypic variation it is important both to characterize epigenetic variation and to understand the association between genetic and epigenetic changes at particular loci.

Genome-wide profiling of DNA methylation abundance in several plant species has documented variation among genotypes or mutant strains. In *Arabidopsis thaliana,* DNA methylation is frequently associated with repetitive DNA sequences such as transposable elements and heterochromatic regions (Lippman et al., 2004; Zilberman et al., 2007; Miura et al., 2009; Vaughn et al., 2007) although there is also evidence for intermediate levels of CG methylation within some gene-bodies as well. Mutations in genes involved in chromatin modification or siRNA generation and processing can have global or locus-specific effects on DNA methylation (Stroud et al., 2013). Several studies that generated epiRILs, genotypes that segregate for differences in DNA methylation but not for genetic changes, provide evidence that alterations in DNA methylation patterns have the potential to influence quantitative traits in Arabidopsis (Johannes et al., 2009; Reinders et al., 2009). The analysis of spontaneous mutation accumulation lines that were propagated for 30 generations with minimal selective pressure provided evidence that DNA methylation variation at single nucleotides is much more common

than genetic change but that changes in regional methylation levels (DMRs) occur at frequencies roughly similar to the rate of genetic change (Becker et al., 2011; Schmitz et al., 2011).

The comparison of DNA methylation levels in different ecotypes reveals relatively stable inheritance of DNA methylation variation that is associated with repetitive sequences but less stable inheritance of gene-body methylation (Vaughn et al., 2007). Detailed profiling of DNA methylation in 152 Arabidopsis ecotypes reveals many examples of single nucleotide differences in DNA methylation as well as examples of regions that are highly methylated in some ecotypes but lack methylation in other ecotypes (Schmitz et al., 2013). Association mapping of these differentially methylated regions (DMRs) shows that many of these are locally controlled with some examples of trans-acting effects on DNA methylation (Schmitz et al., 2013). Similar results were found by assessing DNA methylation variation in rice and looking at allele-specific DNA methylation and expression in the F1 hybrid (Chodavarapu et al., 2012). In maize, the comparison of DNA methylation in two genotypes revealed examples of DNA methylation variation and suggested that these include examples that are likely to be purely epigenetic (Eichten et al., 2011) as well as examples that are likely the result of transposon insertion variation (Eichten et al., 2012).

DNA methylation is often considered to be an epigenetic mark. However, DNA methylation is specifically a chromatin modification that can be the result of genetic or epigenetic influences. Richards (2006) provided a classification scheme to consider the relative influence of genetics and epigenetics upon chromatin state at epialleles. Obligatory epialleles represent examples in which the chromatin state is directly correlated with a genetic change, such as a transposon insertion or a structural variant. Facilitated epialleles occur when a genetic change leads to a poised allelic state that could exist in either a silenced or active form. Pure epialleles describe instances in which there are no genetic changes that influence chromatin state. There is evidence that each of these types of epialleles exist in plant populations. Many of the known loci showing DNA methylation variation (Martin et al., 2009; Durand et al., 2012; Luff et al., 1999; Melquist et al., 1999) are located near sequence variation or transposable element insertions that could be causal factors determining the methylation state. Indeed the hypermethylation of certain retrotransposon elements has been shown to spread DNA methylation to low copy regions flanking the element (Hollister et al., 2011; Eichten et al., 2012) and could lead to obligatory epiallele formation. Other methylation variants have been

connected to sequence variation for related sequences located in other areas of the genome in which siRNAs may play a role in directing the methylation state at the observed loci (Durand et al., 2012; Luff et al., 1999; Melquist et al., 1999). In contrast, the *Lcyc* locus in *Linaria vulgaris*, the *CNR* locus in tomato and the *SUP/clark kent* alleles in *Arabidopsis*, do not display any local sequence variation within 10kb of the gene when comparing lines displaying altering methylation states (Jacobsen & Meverowitz, 1997; Cubas et al., 1999; Manning et al., 2006). Similarly, some examples of differential DNA methylation in maize exist in genomic regions without genetic differences and are not influenced by other genomic regions (Eichten et al., 2011).

Understanding the relative contribution of these three classes of epialleles across our observed DMRs is important for several reasons. First, if the majority of DMRs are the result of obligatory variation then it is quite possible that SNPs will be in linkage disequilibrium with the genetic change and could be used to predict or infer DNA methylation state. In contrast, facilitated and pure epialleles have information content that would not be captured by DNA genotyping approaches. Second, these different types of epialleles are expected to exhibit differences in stability and heritability. If the chromatin state is "programmed" by genetic features then the chromatin state should be quite stable and reproducible across generations. In contrast, facilitated or pure epialleles might be much less stable and exhibit reversions in chromatin state at much higher frequencies. In this study we invesTrends in Geneticsate the diversity of DNA methylation states and their association to genotype and gene expression in 51 diverse maize inbred lines. Maize (Mays ssp. mays) is a highly diverse species (Buckler et al., 2006; Messing & Dooner 2006; Chia et al., 2012) that provides a useful model to study the role of epigenetic variation due to the interspersed nature of genes and repetitive sequences which may create abundant opportunities for epigenetic variation (Rabinowicz & Bennetzen 2006; Schnable et al., 2009; Baucom et al., 2009). Thousands of DMRs were found in these diverse maize lines and many of these can be confirmed using MethylC-Seq (Lister et al., 2008). The comparison of DNA methylation levels and genotypes provides evidence that a substantial portion of DMRs are associated with local genotype. This heritability of the DNA methylation patterns for many of the DMRs was confirmed by analysis of methylation levels in RILs. The genes located near DMRs include a number of genes with very tissue-specific expression patterns and there are over 300 genes located near DMRs have expression patterns that are strongly correlated with the methylation state among genotypes.

**RESULTS**

***Thousands of differentially methylated regions are detected in diverse maize inbreds***

The process for discovering and characterizing differentially methylated regions (DMRs) is outlined in Figure 1A. Genome-wide profiling of DNA methylation abundance using meDIP-chip profiling was performed on the B73 reference genome along with 19 diverse inbred genotypes of maize selected from the NAM mapping to represent maize diversity (McMullen et al., 2009). Immunoprecipitation of methylated DNA (meDIP) was performed on DNA samples of each genotype from three biological replicates using a 5-methylcytosine antibody that allows for enrichment of DNA fragments that contain DNA methylation in any sequence context (CG, CHG or CHH). The immunoprecipitated DNA and input total DNA were hybridized to a custom long-oligonucleotide microarray platform containing 2.1M probes spaced every 200bp across the low-copy portion of the maize genome sequence (Eichten et al., 2011). This meDIP-chip method allows for the analysis of DNA methylation abundance across the genome but does not provide information about the sequence context of DNA methylation.

Maize inbreds have high levels of structural variation including CNV and PAV (Swanson-Wagner et al., 2010). A two-fold approach was used to miTrends in Geneticsate the influence of structural variation on the analysis of DNA methylation levels (see Methods for details). Any probes with >2-fold difference in hybridization intensity for genomic DNA of a genotype relative to the genomic DNA of the reference genotype (B73) were omitted from DNA methylation analysis for that genotype. For the remaining probes the estimate of DNA methylation levels were normalized to account for slight differences in the hybridization intensities of genomic DNA. The DNA methylation profile for each of the 19 genotypes was compared to the B73 DNA methylation profile in order to identify DMRs. The DNAcopy algorithm (Venkatraman & Olshen, 2007) was used to find regions consisting of multiple probes that exhibit at least two-fold variation in the level of DNA methylation relative to B73 and resulted in 415-804 DMRs in each genotype relative to B73. The DMRs discovered for all 19 samples were condensed to a non-redundant list of 9,899 DMRs that are detected in at least one genotype relative to B73.

In order to better survey the frequency of DNA methylation variation and to assess potential causes or consequences of this variation, the level of DNA methylation at these DMRs was

assayed in a single biological replicate of 31 additional genotypes. These additional genotypes were part of a recent resequencing study (Chia et al., 2012) and include eight inbred lines derived from teosinte or landraces (Table 1). The average of per-probe DNA methylation level for each of the DMRs was calculated for each of these genotypes. As only a single replicate was profiled for these 31 genotypes we did not attempt to discover additional novel DMRs represented in these genotypes relative to the original 20 genotypes. DMRs were validated in independent biological replicates by qPCR (Figure 2), bisulfite sequencing, and profiling in a recombinant inbred population (see below).

A series of filtering criteria were applied to the 9,899 DMRs in order to identify a robust set of common and rare variants (Figure 1A; see Methods for details). The 1,966 common variants include DMRs that have high/low levels of DNA methylation in at least three of the 51 genotypes surveyed (example in Figure 1B). The 1,754 rare variants only exhibit altered DNA methylation levels in one or two of the genotypes (example in Figure 1C-D). The criteria used to identify common or rare variants are different (see methods) and therefore it was improper to directly compare the rate of rare and common DMRs across the entire panel of genotypes. However, it was possible to assess whether the minority state for common or rare DMRs was more often hypermethylated (as in Figure 1C) or hypomethylated (as in Figure 1D). There was a significant ($P<0.001$) enrichment for DMRs in which most genotypes are methylated and a small number exhibit hypomethylation (Figure 3). This suggests that it is more common for methylated regions to lose DNA methylation in a small number of genotypes than for generally unmethylated regions to gain methylation in a small number of genotypes. The common and rare DMR variants are distributed throughout the maize genome (Figure 1E).

***DMRs are frequently the result of CG and CHG methylation***

MethylC-Seq (Lister et al., 2008) was performed on independently grown samples of B73 and Mo17 in order to validate the DNA methylation differences observed by array-based profiling and to characterize the sequence context of DNA methylation at DMRs. There were 248 DMRs with difference between B73 and Mo17 methylation levels based on meDIP-array estimates that were present in the common or rare set of DMRs and had at least 80% coverage in both the B73 and Mo17 MethylC-seq data. The majority (91%) of these B73-Mo17 DMRs had substantial

(>50%) difference in the level of CG and/or CHG methylation within the DMR (Figure 4). Similar proportions of common (92%) and rare (89%) DMRs were confirmed by the bisulphite sequencing data. In most (84%) of the DMRs that were validated by bisulphite sequencing data there was a substantial difference in DNA methylation in both the CG and CHG contexts. A small proportion of DMR only exhibit differences in levels of CG (9%) or CHG (7%) DNA methylation. None of the DMRs exhibit substantial variation in CHH DNA methylation.

***DNA methylation variation is frequently associated with local genetic variation***

Although DNA methylation is often considered to be an epigenetic mark there is evidence that it can be highly influenced by genetic changes, such as transposon insertions (Law & Jacobsen, 2010; Hollister et al., 2011; Eichten et al., 2012). A "local" association scan was implemented to determine what proportion of DMRs might be the result of local genetic variation and represent information content that was already "captured" by SNP-based analyses. We did not attempt a genome-wide association scan due to the limited number of genotypes (51) that were sampled in this study. All 51 genotypes with DNA methylation data have been re-sequenced and used to identify a set of 56 million SNPs (Chia et al., 2012). For each of the common DMRs the SNPs within 1kb of each DMR were used to perform an association analysis relative to the DNA methylation level. Per-SNP significance was estimated using permutation to associate the same DNA methylation phenotype with random genomic regions (Figure 5A-B) and the per-DMR significance was determined by comparing the frequency of significant SNPs in the DMRs compared to randomly selected genomic regions. This analysis was applied only to the common DMRs as the rare DMRs did not have statistical power for associating rare phenotypic variation in DNA methylation with SNPs. A substantial proportion (1003/1966) of the common DMRs were significantly associated (see Methods) with local SNPs (Figure 5C). In some cases the highly associated SNPs were located within the DMRs but in many cases the associated regions were outside or partially overlapped the DMR (Figure 6). The remaining 963 common DMRs did not exhibit significant associations between the DNA methylation level and local SNPs. There are several potential reasons for a lack of a local association including lack of SNPs in LD with local causative genetic changes, potential trans-acting factors that influence DNA methylation level or purely epigenetic influences on DNA methylation.

The significant associations of local SNPs with DNA methylation level for 1,003 of the common DMRs suggests that the DNA methylation levels can be predicted based upon SNPs for these

regions. The ability to accurately predict DNA methylation level based upon SNPs was tested by determining DNA methylation levels in 12 additional genotypes (Table S1) that had re-sequencing based SNP calls (Chia et al., 2012). These genotypes were hybridized to a microarray format that only contained probes to survey DMRs (methods). There were 535 DMRs that exhibit allelic variation for the most significantly associated SNP and were assessed by this microarray platform (Figure 7). In many cases (316 of 535 DMRs) only one or two of these genotypes contained the alternate methylation state based on the most significantly associated SNP for each DMR (Figure 8A). However, we compared the genotype and DNA methylation phenotype at each of these DMRs to determine whether the associated SNP could predict DNA methylation levels (Figure 8A). For the majority (77%) of the DMRs the SNP genotype was associated with DNA methylation levels in these 12 genotypes in the expected direction (Figure 8B). The agreement between genotype and DNA methylation was higher for DMRs with at least three observations for each haplotype (Figure 8B).

While there were a number of common DMRs that exhibit associations with SNPs, it is not likely that SNPs themselves are causative for differential DNA methylation. Instead, it is more likely that the SNPs are in linkage disequilibrium with nearby transposon insertions or structural variation. A subset of retrotransposable elements have been identified in maize that display a 'spreading' of DNA methylation into the low-copy sequence adjacent to the insertions (Eichten et al., 2012) and may contribute to DNA methylation variation if the specific insertion is polymorphic among genotypes. Each of the DMRs was mapped relative to nearby transposons in order to invesTrends in Geneticsate the relationship between transposons insertions and DMRs. Each of the common or rare DMRs, as well as nearly 10,000 randomly selected control regions was classified according to the presence and type of transposable element within 500bp (Figure 5D). The common and rare DMRs are more likely to be located near a spreading element relative to control regions and are depleted for having no element. The proximity to transposons was tested based on analysis of the B73 reference genome. However, it is likely that some of these insertions are polymorphic among the genotypes that were assessed. We would hypothesize that DMRs found near spreading retrotransposable elements, which where annotated in the B73 reference background, should display a high level of methylation in the B73 sample. To test this, the B73 methylation level across all DMRs was calculated. Indeed, B73 tends to have much higher levels of DNA methylation for DMRs that are near transposons as compared to control regions (Figure 5E). This increase is more pronounced when looking

specifically at the spreading elements. As the presence of a transposon insertion may be in LD with local SNPs, the proportion of common DMRs displaying an association to local SNPs in each group (no element, spreading, non-spreading) was calculated(Figure 5D, shown in common DMR bar). The DMRs that are located near transposons are more likely (>50%) to have a significant association with SNPs than DMRs that are not located near transposons (37%; Figure 5D). Collectively, these analyses suggest that some DMRs are associated with genetic changes and that many of these may be the result of transposon insertions that influence nearby low-copy sequences.

***Many DMRs are heritable***

While association scans invesTrends in Geneticsate the genetic influences on DMRs, an alternative approach to study the heritability and control of DNA methylation at DMRs was performed in a bi-parental recombinant inbred line (RIL) population derived from B73 and Mo17 (Lee et al., 2002). A custom 12-plex microarray that includes a subset of the probes from the full array was used to profile DNA methylation in 17 RILs (Table 1) as well as independently grown replicates of B73 and Mo17. DNA methylation states were confirmed for 401 of the B73-Mo17 DMRs assessed by this platform. The relative levels of DNA methylation at the confirmed 401 DMRs were analyzed relative to the genotype of the chromosomal region containing the DMR in each recombinant inbred line. DMRs that are locally controlled by genetic or epigenetic change are expected to have DNA methylation levels that are correlated with the genotype of the region surrounding the DMR (Figure 8C - cis; Figure 9). If there is some instability in the memory of epigenetic information then it is possible that inheritance will appear mostly associated with local genotype but some "switches" in epigenotype relative to genotype might be observed (Figure 8C – unstable cis). Trans-acting regulation of DNA methylation will result in two discrete levels of DNA methylation that are not correlated with local genotype but instead are correlated with the genotype of a locus elsewhere in the genome (Figure 8C - trans). Paramutation would be expected to result in all RILs exhibiting DNA methylation states that are similar to the levels observed in one parent (Figure 8C - para). The vast majority of DMRs exhibit local control of DNA methylation (Figure 8C). Very few examples of trans-acting control (24) or paramutation (5) were observed.

The DMRs surveyed in RILs included 201 common DMRs with significant associations with local SNPs and 101 common DMRs without significant local association as well as 99 rare DMRs. The

common DMRs that exhibit significant association with local SNPs would be expected to show cis-acting inheritance of DNA methylation levels. The common DMRs without significant local SNP associations might be expected to include more examples of trans- or complex genetic control. While there is some increase in the proportion of trans- or complex acting genetic control the majority of the DMRs without a significantly associated nearby SNP still exhibit cis-acting control (Figure 8D). This may suggest that these do not have nearby SNPs in LD with the causative change or that they are purely epigenetic local variation.

***Functional consequences of DMRs***

Genes located adjacent to DMRs (the closest gene in each direction from the DMR) were identified and assayed to understand the potential consequences of variable DNA methylation levels. Comparing the 2,375 genes that are located within 10kb of a DMR to all maize genes suggests that DMRs occur at equal frequency in both sub-genomes of maize (the products of the recent whole-genome duplication event) and are not enriched for inserted or syntenic genes (Figure 10A-B) (based on annotations from Schnable et al., 2012). In addition, the DMRs occur at similar frequency in the high-recombination, gene-rich arms of maize chromosomes and in the central, gene-poor low recombination portions of maize chromosomes (Figure 10C). Schmitz et al, (2013) recently noted that many of the genes near DMRs in Arabidopsis tend to exhibit pollen-specific expression patterns. Data from an RNA-seq based expression atlas of 51 tissues of B73 (Sekhon et al., 2013) was used to assess the tissue-specific expression pattern for each of the genes located within 10kb of DMRs was assessed in 51 tissues of B73 (Figure 10D). Although pollen tissue was not included in this experiment there are 85 genes that show high levels of expression (average FPKM = 1,775) only in meiotic tassel or anthers (Figure 10D).

The DMRs may influence phenotypic variation by affecting expression of nearby genes. In order to determine possible consequences of variable methylation state, RNAseq was performed on the same tissue samples that were used for DNA methylation profiling. Transcript abundance of 277 genes in the annotated filtered maize gene set that were closest to a classified common DMR were significantly correlated to DNA methylation levels (both rank-sum and Kendall's tau $q < 0.05$) (example shown in Figure 11A). For rare DMRs, significance was calculated by z-score and Kendall's tau identifying a total of 111 significant correlations (example shown Figure 11B). The use of different statistical tests allowed for the identification of either qualitative or quantitative associations of expression and DNA methylation in both the common and rare

DMRs. As expected, there was an enrichment for significant negative (as opposed to positive) associations of DNA methylation with gene expression for both common (70%) and rare (73%) DMRs. DNA methylation appears to be correlated with on/off states in gene expression for 26% of the DMRs that are negatively associated with DNA methylation levels (as exemplified in Figure 11B). In other cases, the DNA methylation state is associated with quantitative differences in gene expression (as in Figure 11A). The examples of DMRs that are significantly associated with gene expression are enriched for examples of DMRs that are located near, or overlap with, genes (Figure 11C) as opposed to DMRs that are located >5kb from genes. The genes that have a negative association between gene expression levels and DNA methylation state are slightly depleted for syntenic genes and show some enrichment for inserted sequences or genes without homologs in other grass species (Figure 10B). Over 45% of these genes are conserved in other grasses and located in syntenic positions which would suggest that many of the genes associated with variable DNA methylation and expression are not mis-annotated transposons.

## DISCUSSION

### *Characterization of DMRs in diverse maize lines*

The profiling of DNA methylation in 51 diverse maize inbred lines identified 1966 common DMRs and another 1,754 rare DMRs. There are likely additional DMRs that did not meet our filtering criteria, were only present in one of the 30 lines not used for the DMR discovery, or that are present in regions of the genome not targeted by the microarray platform. However, the number of DMRs found among maize genotypes are similar to the numbers identified in recent studies of *Arabidopsis* (Schmitz et al., 2013) or rice (Chodavarapu et al., 2012). In *Arabidopsis* it was noted that hypermethylation is the rare state for most DMRs (Schmitz et al., 2013). In contrast, we observed an enrichment for hypomethylation as the rare state more commonly for maize DMRs (Figure 3). We found that the methylation levels for DMRs were highly consistent in different samples of the same genotypes. All three biological replicates of a genotype that were profiled for the 20 NAM parents exhibit very similar methylation profiles. In addition, the bisulphite sequencing of an independent replicate of B73 and Mo17 (seeds from a different

plant) confirmed the majority of DMRs. In the majority of genomic regions , patterns of methylation were extremely similar and reproducible.

In Arabidopsis (Schmitz et al., 2013) and rice (Chodavarapu et al., 2012) the DMRs were classified based on the sequence context of DNA methylation (CG, CHG or CHH). The DMRs that only exhibit differences in CG DNA methylation in Arabidopsis often occurred in genes, showed high levels of variation and rarely were associated with low levels of gene expression (Vaughn et al., 2007, Schmitz et al., 2013). In contrast, DMRs associated with methylation in contexts beyond CG (termed C-DMRs) were more stable and were often associated with lower levels of expression. The protocol that we used for profiling should have the potential to discover either CG or C-DMRs. However, the bisulphite sequence analysis of B73 and Mo17 (Figure 4) reveals that the majority of maize DMRs exhibit both CG and CHG differences. There were very few examples of CG or CHG specific differences in DNA methylation. It might be that the partial differences in DNA methylation (ie 20% versus 60% DNA methylation) often found at CG-DMRs were not strong enough to be detected by our approach. The bisulphite sequencing approach for DNA methylation profiling is often used to identify single base cytosine methylation differences (Becker et al., 2011, Schmitz et al., 2011, Chodavarapu et al., 2012, Schmitz et al., 2013). Our microarray-based approach cannot detect these single base methylation polymorphisms, however there is limited evidence for functional impacts or heritability of these single nucleotide changes in DNA methylation.

The most likely mechanism for differential DNA methylation levels to impact phenotype is via effects on gene expression levels. The general view is that DNA methylation within gene promoters or regulatory elements can result in repression of gene expression. However, there is evidence that gene body methylation is associated with moderately expressed genes (Zilberman et al., 2007; Law & Jacobsen, 2010). We found that DMRs were associated with variable gene expression in matched seedling tissue for 11.5% of the common DMRs and 6.3% of the rare DMRs. The majority (70%) of associated genes exhibit expression levels that are negatively associated with the DNA methylation levels. It is possible that there are additional examples of gene expression that are influenced by the variable DNA methylation that might have been detected by analysis of gene expression in other tissues or using more permissive criteria. This analysis of the contribution of DNA methylation variation to gene expression differences suggests that only a small portion of gene expression variation among maize

genotypes is attributable to differences in DNA methylation. Schmitz et al. (2013) noted that many of the genes near DMRs that were frequently methylated exhibit pollen-specific gene expression. Similarly, we found that a number of the genes located near DMRs show high levels of tissue-specific expression in anthers or root tissue (Figure 10D).

***Potential causes of DNA methylation***

DNA methylation is often considered to be an epigenetic mark. There is certainly evidence that DNA methylation can be a major contributor to epigenetic regulation in some examples (Cubas et al., 1999; Manning et al., 2006; Jacobsen & Meyerowitz, 1997). However, there is also strong evidence to show that certain genetic features can lead to DNA methylation. We are interested in understanding the causes of variable DNA methylation in maize plants in order to understand how to capture this information in breeding programs. Richards et al (2006) classified three major types of epialleles based on the interaction of genetic and epigenetic influences. We propose to add further nuance to these classifications and define six sub-types of DMRs (Figure 12).

The first two types of DMRs are both classified as purely epigenetic are distinguished based on whether they are highly stable or relatively unstable. The next three groups are classified as genetic as they all show some level of influence from the genetic content of the individual. Genetic-local DMRs are influenced by local genetic changes, such as structural rearrangements or transposon insertions, and could result in stochastic DNA methylation variation in the presence of the genetic change or could show complete agreement between DNA methylation and genetic state. Genetic-remote DMRs exhibit changes in DNA methylation that are influenced by a trans-acting region of the genome. It is most likely that this trans-acting signal is based on small RNAs that are generated by related sequences elsewhere in the genome, such as in the well characterized example of PAI silencing (Bender and Fink 1995). The third group type of genetically influenced DMR has polygenic influences on the DNA methylation state. This could be due to influences from multiple trans-acting sites or a combination of a cis-acting change and a trans-acting feature. The final type of DMR could result from paramutation, an allelic interaction in which one allele could influence the methylation state of the other when present together in a heterozygote (Erhard and Hollick 2011). This type of DNA methylation pattern is likely a sub-type of one of the other classes but is expected to show unique behavior in populations and is therefore separately classified. It is useful to separate these potential types

of DMRs in order to address which types appear to be most prevalent and to consider how to most efficiently capture this type of variation when predicting phenotype in large populations.

We evaluated the heritability of DNA methylation variation both in an association panel of 51 genotypes as well as in a bi-parental RIL population. A major goal was to identify the proportion of DMRs that are associated to the local genetic content of an individual. In order to detect a significant association in the panel of 51 genotypes it would be necessary to have quite stable inheritance of DNA methylation levels and to have a SNP in linkage disequilibrium with either the causative genetic change or the altered DNA methylation level. The finding that approximately half of the common DMRs had significant associations with local SNPs (Figure 5) suggests either stable epigenetic DMRs or genetic-local (*cis)* DMRs are relatively common. The analysis of DNA methylation levels in the RILs provides further support for these two classes. The genetic-remote (*trans*) and paramutation types of DMR patterns are expected to have unique patterns in the RILs compared to genetic-local or epigenetic DMRs. However, there were few examples of DMRs that exhibit patterns that would reflect paramutation or simple trans-acting control (Figure 8). There is evidence for paramutation-like patterns or trans-chromosomal DNA methylation in rice (Chodavarapu et al., 2012) and Arabidopsis (Greaves et al., 2012) but these seem to be rare relative to the number of DNA methylation differences among the parental genotypes. Many of the DMRs with, or without, significant local associations still exhibit evidence for cis-acting control of DNA methylation in the RILs (Figure 8).

The finding that some RILs have altered DNA methylation levels relative to their genotype (unstable cis) suggests some stochastic behavior for DNA methylation levels relative to genotype and may highlight the potential for either epigenetic-unstable behavior or the genetic local-influences that results in a stochastic change rather than a deterministic change. Schmitz et al. (2013) performed a genome-wide association to map the factors that influence DNA methylation for DMRs in Arabidopsis. They found evidence for local-effects on some DMRs and found that a number of these had nearby insertion/deletions. However, they also found a number of cases with complex control by more than one genomic region. Examples of DMRs that are influenced by multiple genomic regions may be reflected by the DMRs that exhibit "complex" patterns in the RILs.

In contrast to both Arabidopsis and rice, the maize genome has a much higher number of transposable elements and many maize genes are located near transposons (Schnable et al., 2009). It has been hypothesized that a significant function of epigenetic regulation is the control of transposons (Hollister et al., 2011). Previous studies in maize have found that some retrotransposons exhibit spreading of heterochromatin to neighboring sequences while other retrotransposon are marked by DNA methylation that does not spread beyond their borders (Eichten et al., 2012). We found that DMRs are enriched for being located near retrotransposons that exhibit spreading of heterochromatin (Figure 5D). This would suggest that a number of the DMRs may be influenced by local variation for the presence of these elements. As these retrotransposons have been annotated specifically in the B73 reference genome, it is possible that many DMRs that do not appear near an annotated retrotransposon, may have unique retrotransposon insertion patterns across the other diverse genotypes that would be unassessed in the current analysis. It is unclear whether the insertion of one of these spreading retrotransposons results in an obligatory change in DNA methylation levels or simply facilitates the potential to acquire higher levels of DNA methylation.

From the perspective of crop improvement a major question is how to fully capture the heritable information in populations in order to improve predictions of phenotypes. It is relatively inexpensive to collect dense SNP genotypes but collecting data on DNA methylation profiles is currently more expensive and complex. Therefore, there is interest in understanding what proportion of DNA methylation variation might be "capturable" by genotype information. Our results suggest that a substantial portion of the DMRs might be significantly associated with local SNPs. However, we might expect even greater ability to predict DNA methylation state if the genotyping information provided the ability to survey the presence-absence of specific transposon insertions throughout the genome. The potential for certain retrotransposons to spread heterochromatin to flanking sequences suggests that being able to identify the genomic positions of these elements might improve our predictions of DNA methylation variation at low-copy sequences among individuals.

**EXPERIMENTAL METHODS & PROCEDURES**

**Experimental Design:** Three replications of the 20 NAM parental genotypes and single replicates of the additional 31 maize and teosinte HapMap2 genotypes (Table 1) were grown in a randomized block design where each replication was a block. For each replication, four seedlings of each line were grown in a single pot that was randomly assigned to a location within the replication block. Seedlings were grown under controlled conditions with 15 hours of light each day in the University of Minnesota Agricultural Research station, Falcon Heights, Minnesota. Seedlings were watered daily as needed. After 18 days of growth, the 3$^{rd}$ leaf (L3) of each plant was harvested and pooled with other plants from the same pot/replication or harvested independently and immediately frozen in liquid nitrogen. DNAs were isolated using the CTAB method (Doyle 1987) from frozen leaf tissue as described in Eichten et al., 2011. RNAs were isolated using TRIZOL (Invitrogen) per manufacturer's protocol.

**meDIP-chip Epigenomic Profiling:** Methods were adapted from Eichten et al., 2011. Briefly, Methylated DNA was immunoprecipitated with an anti-5-methylcytosine monoclonal antibody from 400ng sonicated DNA using the Methylated DNA IP Kit (Zymo Research, Orange, CA; Cat # D5101). To serve as a negative control, water was substituted for the monoclonal antibody on B73 sonicated DNA. For each replication and genotype, whole genome amplification was conducted on 50-100ng IP DNA and also 50-100ng of sonicated DNA (input control) using the Whole Genome Amplification kit (Sigma Aldrich, St. Louis, MO, Cat # WGA2-50RXN). For each amplified IP input sample, 3ug amplified DNA were labeled using the Dual-Color Labeling Kit (Roche NimbleGen, Cat # 05223547001) according to the array manufacturer's protocol (Roche NimbleGen Methylation UserGuide v7.0). Each IP sample was labeled with Cy5 and each input/control sonicated DNA was labeled with Cy3. Samples were hybridized to the custom 2.1M, 1.4M, or 270k probe array depending on sample set for 16-72hrs at 42°. Slides were washed and scanned according to NimbleGen's protocol for the GenePix4000B (2.1M platform) and Nimblegen MS200 (1.4M and 270k platform) scanner. Images were aligned and quantified using NimbleScan software (Roche NimbleGen) which produced raw pair reports containing fluorescent intensity readings for each probe on the array.

**Normalization and Linear Modeling:** Pair files for 2.1M and 1.4M platforms that were exported from NimbleScan were imported into the Bioconductor statistical environment (Gentleman et al., 2004). 2.1M platform pair files were subset to the 1.4M platform probes to facilitate combined analysis. For each genotype, designated as HapMap2 'HM' lines (excluding B73), microarray data channels were assigned the following factors: B73, HM_line, B73 input, or HM_line input depending on sample derivation. For samples in which three replicates were not available, individual replicates were duplicated to maintain normalization methods across all samples. Non-maize probes and vendor-supplied process control probes were configured to have analytical weights of zero. Variance-stabilizing normalization was used to account for array-specific effects. Factor-specific hybridization coefficients were estimated by fitting fixed linear model accounting for dye and sample effects to the data using the limma package (Smyth 2004). To compute biologically relevant information about each HM line's DNA methylation, the following contrasts were then computed: B73 IP vs B73 input (B73 methylation); HM_line input vs B73 input (CGH and differential hybridization efficiency); HM_line IP vs [HM_line input vs B73 input] (HM_line methylation corrected for differential hybridization efficiency); B73 IP vs HM_line IP vs [HM_line input vs B73 input]] (differential DNA methylation corrected for differential hybridization efficiency). Moderated t-statistics and the log-odds score for differential MeDIP enrichment were computed by empirical Bayes shrinkage of the standard errors with the False Discovery Rate controlled to 0.05 (Smyth 2004).

**Analysis of Variable Methylation:** To identify segments showing differential methylation between B73 and other HM lines, the DNAcopy algorithm (Venkatraman et al., 2007) was used on differential methylation values (B73 – HM line) to identify multi-probe regions displaying similar patterns of differential methylation. Prior to segmentation, probesets were filtered to remove probes displaying CGH effects between the specific HM line and B73 ([HM_line input vs B73 input] < -1). Segments defined by DNAcopy were filtered to those showing at least two-fold difference between B73 and the HM line (log2 value of < -1 or > 1). The filtered segments were used for DMR identification.

A non-redundant 'Discovery Set' of 9899 DMR segments was developed by stepwise filtering of segments and segment borders. Each genotype segment was split whenever there was a gap between probes of over 700bp. Any resulting region that was a single probe was discarded. Segments were further summarized across genotypes by obtaining segment coverage for each

position. Each segment contained ranges specifying how many genotypes also had a segment within the genome space. Ranges with a ratio of genotype count to segment's maximum genotype below 0.6 were removed. Consecutive ranges were combined producing 9899 segments. For each segment, genotypes were split into two classes (methylated or unmethylated) based on log2(IP/input) signal values across each segment. For each class, the mean and variance were calculated for methylation level to identify a goodness of fit based on Fisher's linear discriminant (FLD). The best FLD values provided a threshold value per DMR to categorize genotype signal values as methylated or unmethylated. Then, FLD value was adjusted by removing genotypes with log-odds ratio below 3. This classification was performed both on the entire set of 51 genotypes as well as a set of 49 genotypes in which the highest and lowest methylated genotypes were removed from the analysis to eliminate major classification effects due to outliers. Regions were classified as being common in the population (≥3 genotypes display the minority methylation state) or rare (≤2 genotypes) to facilitate downstream analysis. The resulting regions were further filtered to require an adjusted FLD value > 8, have a different in class means (methylated vs. unmethylated) of 1 and at least one class mean falling between -0.8<X<0.8. Additional rare DMRs were selected if they passed one of the following two tests. The first was based on calculating z-score between the outlier and the other genotypes. We used 0.05 significance level for the corresponding p-value corrected for multiple testing using Storey's FDR approach (Storey, 2001). In addition, Grubbs' test for outliers (Grubbs, 1969) was performed on DMRs that passed Anderson-Darling normality test. False discovery rate was also controlled at 0.05 using Storey's method (Storey, 2001). Both tests allowed us to identify genotypes that have a methylation state significantly different from the other genotypes. These additional DMRs were merged with the previously described 'rare' DMRs. This resulted in 1966 high confidence 'common' and 1754 'rare' DMRs for subsequent analysis (Figure 1A; Table 1).

**qPCR Methylation Validation:** qPCR validation of DMRs was adapted from Eichten et al., 2011. Briefly, primers were designed to amplify regions of 10 DMRs (Figure 2). Sample DNA from the discovery set of genotypes was digested with the methyl-sensitive restriction enzyme FspEI (New England Biolabs) along with glycerol mock digests. The difference between digest C(t) and mock C(t) was calculated for each sample tested. As our selected enzymes target methylated cytosines, higher methylation leads to increased digestion and subsequently longer C(t) times.

**Bisulfite Sequencing DMR validation:** Whole-genome bisulfite sequencing was performed adapting methods from Schmitz et al., 2011. Briefly, 14-day whole-seedling gDNA was isolated from B73 and Mo17 inbred lines planted independently from samples used in meDIP analysis. Samples were fragmented, and TruSeq methylated adapters were ligated to DNA fragments. 500ng of adapter-ligated DNA underwent bisulfite conversion using the MethylCode bisulfite conversion kit (Life technologies) per manufacturers protocol. Converted DNA was split into 4 reactions and amplified using Pfu Turbo Cx DNA Polymerase (Agilent) for four cycles and subsequently pooled. Libraries were sequenced on the HiSeq 2000 (Illumina) for 100 cycles, paired-end. Sequencing reads were processed to identify and filter poor 3' quality and incomplete conversion. Sequenced were aligned using the Bismark aligner (v0.7.2, Krueger & Andrews, 2011) against the B73 RefGen v2 genome under the following parameters (-n 2, -l 50). Methylated cytosines were extracted from aligned reads using the Bismark methylation extractor under standard parameters. The proportion of CG, CHG, and CHH methylation was determined for 100bp windows across the genome. The methylation levels across DMRs were created by averaging methylation levels from intersecting 100bp windows using BEDTools (Quinlan & Hall, 2010).

**DMR / SNP Association Analysis:** In order to identify DMRs that display significant association to local genetic variation, a custom-designed pipeline was developed. SNPs from the maize HapMapv2 project (Chia et al., 2012) were extracted for all DMRs and flanking 1kb of genomic sequence. The 36 DMRs with no SNPs among the 51 genotypes were not tested. For each DMR and regional SNP pairing, genotypes with ambiguous SNP calls were omitted and a two-tailed t-test was performed for each SNP call and methylation values (log2(IP/input)) for the DMR region. Given the number of tests performed, it is important to define a threshold of significance and control for false associations. A set of 100 regions of 1000 SNPs were randomly selected throughout the maize genome and were used to test for random associations between these 'control' SNPs and the methylation values for each DMR. For each test of control SNPs and genotype methylation states, a significance value was determined. The 1% quantile of all control t-test p-values was used as a significance threshold in which experimental p-values must be below in order to be defined as significant. From this, a DMR was identified as putatively associated with local genetic state, if at least 3 SNPs in the DMR region displayed significance based on the 1% quantile cutoff. In order to further define a false discovery rate for the classification of DMRs as a whole, the proportion of SNPs passing the 1% threshold for all 100

control regions was calculated. The proportions for each DMR-region calculation (101 in total, 1 experimental and 100 control) were ranked based on proportion value. Experimental DMRs with a proportion in the 95$^{th}$ percentile were classified as DMRs associated with local genetic state.

**Analysis of retrotransposable element enrichment near DMRs:** All high confidence common and rare DMRs were annotated with respect to the closest annotated maize repeats (ZmB73_5a_MTEC_repeats; Schnable et al., 2009; Baucom et al., 2009) using BEDtools (Quinlan and Hall., 2010). The nearest elements were classified as either spreading (5mc and H3K9) or non-spreading according to Eichten et al (2012). Control regions were developed by randomly selecting 6613 4-probe regions <4000bp in size to simulate the sizes of experimental DMRs. The number of control regions was scaled according to maize chromosome size to better mimic the distribution of DMRs across maize chromosomes. Control regions were mapped to MTEC repeats as described for high confidence DMRs. For each DMR, B73 CG methylation levels > 80% across the region or < 20% were classified as 'high methylation' and 'low methylation' respectively. Methylation levels 20-80% were deemed 'unclassified'.

**RIL Array experimental design and data processing:** 17 RILs were grown and profiled in a similar fashion to those HapMap2 genotypes described above. Methylation profiles across 789 DMRs were created on a custom 270k feature microarray platform (see Data Access). Raw hybridization signals of IP, input, and CGH contrasts between each RIL and inbred line B73 was made for the IP and the Input channels, respectively. RIL genotypes were validated with a subset of array probes selected to show differential hybridization between B73 and Mo17 (Springer et al., 2009) and would thus be able to verify the genotypes of the RILs derived from those two inbred lines. Prior genotyping of this RIL panel (Li et al., 2013) was used as to validate array genotype calls. Raw signals for those lines were normalized using a variance-stabilizing method implicated in RINGO package (vsn; Toedling et al. 2007). Methylation level for each DMR was calculated as the log2(IP / input) for all probes defined within DMR boundaries.

**Classification of DMR in recombinant inbred line (RIL) population:** RILs were divided into two groups based on their genotypes at each DMR. Methylation levels for both groups were compared using student's *t* test followed by Benjamini-Hochberg correction of p-values. DMRs with a corrected p-value < 0.05 were considered to be regulated by '*cis*-' local control, whereas all the other DMRs were further classified into 3 other categories: *trans*-regulated, paramutaion-like or complex DMR using the following criteria. The Fisher linear discriminant (FLD) statistic

was computed for each DMR and maximized by numerating all the possible groupings of the RILs at each DMR. The maximized FLD value was used to divide the RILs into two groups. DMRs were classified as 'trans' if both groups have ≥ 5 RILs, and the two parental genotypes were within different groups. To identify putative paramutation-like DMRs, the probability of the parental lines B73 and Mo17 grouping with RIL methylation values was computed. If there is a high probability that one parent displays a similar methylation level as the RIL population ($P \geq 0.2$), while the other parent does not ($P < 0.001$), the DMR was classified as 'paramutation-like'. DMRs that were not classified by these methods were classified as "complex DMR".

*Cis*-regulated DMRs were subdivided by their stability across generations. If the smallest difference in methylation value between the highly methylated and lowly methylated group was > 0, the DMR was classified as completely regulated in *cis*. For the remaining *cis* DMRs, the genotype of each RIL was predicted based on its methylation level using the linear discriminant analysis (LDA) function in R. The predicted genotype was then compared with the published genotype (Li et al. 2013). If there was a discrepancy between the published genotype and predicted genotype and the posterior probability for this difference was ≥ 0.7, the methylation state for this RIL was classified as "swapped" relative to the genotype of this RIL.

**Validation of local associations between SNP and methylation level:** The associations between SNP and methylation level were tested in a panel of 12 additional HapMap2 genotypes (Table 1). Those samples were separated into 2 groups based on the genotype of the SNP. Average methylation levels between those two groups were contrasted. If same allele showed association with high-methylation state as in the panel used to identify the association, the association was called "confirmed" in the new panel, otherwise, "not confirmed" was assigned.

**Annotation of DMRs for other maize features:** The distance of maize filtered genes (ZmB73_5b_FGS) on both sides of defined DMRs was calculated using the BEDtools utility suite (Quinlan and Hall, 2010). Genes were annotated for their syntenic position, maize sub-genome status, and known orthology to rice and sorghum gene models using data from (Schnable et al., 2012).

**RNA-seq and Expression Analysis:** RNA isolated from 50 genotype seedling L3 leaf samples used for meDIP profiling were prepared for sequencing at the University of Minnesota Genomics Center in accordance with the TruSeq library creation protocol (Illumina). Samples were

sequenced on the HiSeq 2000 developing 8-24 million reads per replicate. Raw reads were filtered to eliminate poor quality reads using CASAVA (Illumina). Transcript abundance was calculated by mapping reads to the maize reference genome (AGPv2) using TopHat (Trapnell et al., 2009) under standard parameters. RPKM values were developed using 'BAM to Counts' across the exon space of the maize genome reference working gene set (ZmB73_5a) within the iPlant Discovery Environment (www.iplantcollaborative.org).

Common and rare DMRs were profiled separately for significant association to their nearest gene expression state. The significant DMR-gene pairs from the common data set were chosen based on the union between Rank Sum ($q < 0.05$) and Kendall's tau ($q < 0.05$) significance tests corrected for multiple testing using Storey's FDR approach (Storey, 2001). The pairs from the rare data set was selected based on the union between z-score ($q < 0.05$) and Kendall's tau ($q < 0.05$) after FDR control.

**REFERENCES**


Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genetics 5: e1000732.

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245–249.

Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83: 725–734.

Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398.

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Current Opinion in Plant Biology 9: 172–176.

Chan SW-L, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Reviews Genetics 6: 351–360.

Chandler VL, Eggleston WB, Dorweiler JE (2000) Paramutation in maize. Plant Molecular Biology 43: 121–145.

Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, et al. (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nature Genetics.

Chodavarapu RK, Feng S, Ding B, Simon S a, Lopez D, et al. (2012) Transcriptome and methylome interactions in rice hybrids. Proceedings of the National Academy of Sciences of the United States of America 109: 12040–12045.

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157–161.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19.1: 11–15.

Durand S, Bouché N, Perez Strand E, Loudet O, Camilleri C (2012) Rapid establishment of genetic incompatibility through natural epigenetic variation. Current biology?: CB 22: 326–331.

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, et al. (2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics 7: e1002372.

Eichten SR, Ellis N a, Makarevitch I, Yeh C-T, Gent JI, et al. (2012) Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genetics 8: e1003127.

Erhard KF, Hollick JB (2011) Paramutation: a process for acquiring trans-generational regulatory states. Current Opinion in Plant Biology 14: 210–216.

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447: 433–440.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5: R80.

Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, et al. (2012) Trans chromosomal methylation in Arabidopsis hybrids. Proceedings of the National Academy of Sciences of the United States of America 109: 3570–3575.

Grubbs F (1969) Procedures for detecting outlying observations in samples. Technometrics 11: 1–21.

Gutzat R, Mittelsten Scheid O (2012) Epigenetic responses to stress: triple defense? Current Opinion in Plant Biology 15: 568–573.

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, et al. (2011) Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proceedings of the National Academy of Sciences of the United States of America 108: 2322–2327.

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. The Quarterly Review of Biology 84: 131–176.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. Science 277: 1100–1103.

Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. Nature Reviews Genetics 8: 253–262.

Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics 9: 883–890.

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27: 1571–1572.

Law J a, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11: 204–220.

Lee M, Sharopova N, Beavis WD, Grant D, Katt M, et al. (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Molecular Biology 48: 453–461.

Li L, Petsch K, Shimizu R, Liu S, Xu WW, et al. (2013) Mendelian and non-mendelian regulation of gene expression in maize. PLoS Genetics 9: e1003202.

Lippman Z, Gendrel A V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.

Luff B, Pawlowski L, Bender J (1999) An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. Molecular Cell 3: 505–511.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948–952.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135–1138.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Melquist S, Luff B, Bender J (1999) Arabidopsis PAI gene arrangements, cytosine methylation and expression. Genetics 153: 401–413.

Messing J, Dooner HK (2006) Organization and variability of the maize genome. Current Opinion in Plant Biology 9: 157–163.

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, et al. (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. The EMBO Journal 28: 1078–1086.

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

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

Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology 9: 149–156.

Rasmusson DC, Phillips RL (1997) Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Science 37: 303–310.

Reinders J, Wulff BBH, Mirouze M, Marí-Ordóñez A, Dapp M, et al. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939–950.

Richards EJ (2008) Population epigenetics. Current Opinion in Genetics & Development 18: 221–226.

Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews Genetics 7: 395–401.

Schmitz RJ, Schultz MD, Urich M a., Nery JR, Pelizzola M, et al. (2013) Patterns of population epigenomic diversity. Nature 495: 193-198.

Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich M a, et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. Science 334: 369–373.

Schnable JC, Freeling M, Lyons E (2012) Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biology and evolution 4: 265–277.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, et al. (2011) Genome-wide atlas of transcription during maize development. The Plant Journal 66: 553–563.

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272–285.

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3: 3.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, et al. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics 5: e1000734.

Storey JD (2002) A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64: 479–498.

Stroud H, Greenberg MVC, Feng S, Bernatavichute Y V, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152: 352–364.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research 20: 1689–1699.

Toedling J, Skylar O, Sklyar O, Krueger T, Fischer JJ, et al. (2007) Ringo--an R/Bioconductor package for analyzing ChIP-chip readouts. BMC Bioinformatics 8: 221.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology 5: e174.

Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23: 657–663.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39: 61–69.

Table 1: Summary of Lines and data collected for analysis

| Genotype | Plant Type | Source | Methylation Set | Methylation Platform | meDIP Replicates | RNA-seq replicates | RNA-seq reads (million bp) |
|---|---|---|---|---|---|---|---|
| B73 | Inbred | PI 550473 | Discovery | 2.1M | 3 | 3 | 46,822,299 |
| B97 | Inbred | PI 564682 | Discovery | 2.1M | 3 | 3 | 64,184,793 |
| CML103 | Inbred | Ames 27081 | Discovery | 2.1M | 1 | 3 | 59,555,282 |
| CML133 | Inbred | CML133 | Profiling | 1.4M | 1 | 1 | 10,108,029 |
| CML192 | Inbred | CML192 | Profiling | 1.4M | 1 | 1 | 10,236,500 |
| CML202 | Inbred | CML202 | Profiling | 1.4M | 1 | 1 | 11,140,488 |
| CML206 | Inbred | CML206 | Profiling | 1.4M | 1 | 1 | 8,787,540 |
| CML228 | Inbred | Ames 27088 | Discovery | 2.1M | 3 | 1 | 13,453,936 |
| CML277 | Inbred | PI 595550 | Discovery | 2.1M | 3 | 1 | 16,421,326 |
| CML322 | Inbred | Ames 27096 | Discovery | 2.1M | 3 | 3 | 55,299,611 |
| CML330 | Inbred | CML330 | Profiling | 1.4M | 1 | 1 | 10,488,563 |
| CML333 | Inbred | Ames 27101 | Discovery | 2.1M | 2 | 3 | 65,190,728 |
| CML504 | Inbred | CML504 | Profiling | 1.4M | 1 | 1 | 21,361,420 |
| CML505 | Inbred | CML505 | Profiling | 1.4M | 1 | 1 | 9,919,395 |
| CML52 | Inbred | PI 595561 | Profiling | 2.1M | 3 | 1 | 16,817,165 |
| Hp301 | Inbred | PI 587131 | Discovery | 2.1M | 3 | 3 | 65,354,901 |
| Il14H | Inbred | Ames 27118 | Discovery | 2.1M | 3 | 1 | 12,653,042 |
| KI11 | Inbred | Ames 27124 | Discovery | 2.1M | 3 | 3 | 38,564,095 |
| KI3 | Inbred | Ames 27123 | Discovery | 2.1M | 2 | 1 | 14,579,133 |
| Ky21 | Inbred | Ames 27130 | Profiling | 2.1M | 3 | 1 | 26,260,623 |
| M162W | Inbred | Ames 27134 | Discovery | 2.1M | 3 | 1 | 16,162,782 |
| M37W | Inbred | Ames 27133 | Profiling | 2.1M | 3 | 3 | 59,915,642 |
| Mo17 | Inbred | PI 558532 | Discovery | 2.1M | 3 | 3 | 47,883,583 |
| Mo18W | Inbred | PI 550441 | Discovery | 2.1M | 3 | 1 | 10,285,877 |
| MR01 | Landrace | Ames 30522 | Profiling | 1.4M | 1 | 1 | 33,293,971 |
| MR02 | Landrace | Ames 30523 | Profiling | 1.4M | 1 | 1 | 22,701,337 |

| Genotype | Plant Type | Source | Methylation Set | Methylation Platform | meDIP Replicates | RNA-seq replicates | RNA-seq reads (million bp) |
|---|---|---|---|---|---|---|---|
| MR09 | Landrace | Ames 30527 | Profiling | 1.4M | 1 | 1 | 40,668,557 |
| MR10 | Landrace | Ames 30528 | Profiling | 1.4M | 1 | 1 | 30,073,450 |
| MR11 | Landrace | Ames 30529 | Profiling | 1.4M | 1 | 0 | NA |
| MR14 | Landrace | Ames 30531 | Profiling | 1.4M | 1 | 1 | 21,719,977 |
| MR18 | Landrace | Ames 30535 | Profiling | 1.4M | 1 | 1 | 31,043,442 |
| MR19 | Landrace | Ames 30532 | Profiling | 1.4M | 1 | 1 | 29,427,624 |
| MR20 | Landrace | Ames 30533 | Profiling | 1.4M | 1 | 1 | 32,141,650 |
| MR22 | Landrace | Ames 30534 | Profiling | 1.4M | 1 | 1 | 30,555,562 |
| MR25 | Landrace | Ames 30537 | Profiling | 1.4M | 1 | 1 | 31,394,449 |
| MS71 | Inbred | PI 587137 | Discovery | 2.1M | 3 | 3 | 50,995,326 |
| NC358 | Inbred | Ames 27175 | Profiling | 2.1M | 3 | 1 | 13,260,113 |
| Oh43 | Inbred | Ames 19288 | Discovery | 2.1M | 3 | 3 | 44,889,405 |
| Oh7b | Inbred | Ames 19323 | Discovery | 2.1M | 3 | 3 | 76,253,237 |
| P39 | Inbred | Ames 28186 | Discovery | 2.1M | 3 | 1 | 21,748,449 |
| TIL01 | Teosinte | Ames 28399 | Profiling | 1.4M | 1 | 1 | 32,746,000 |
| TIL06 | Teosinte | Ames 28401 | Profiling | 1.4M | 1 | 1 | 33,196,714 |
| TIL07 | Teosinte | Ames 28402 | Profiling | 1.4M | 1 | 1 | 30,048,309 |
| TIL09 | Teosinte | Ames 28403 | Profiling | 1.4M | 1 | 1 | 34,970,995 |
| TIL11 | Teosinte | Ames 28405 | Profiling | 1.4M | 1 | 1 | 22,214,335 |
| TIL14 | Teosinte | Ames 28406 | Profiling | 1.4M | 1 | 1 | 21,011,623 |
| TIL15 | Teosinte | Ames 28407 | Profiling | 1.4M | 1 | 1 | 21,884,742 |
| TIL16 | Teosinte | Ames 28408 | Profiling | 1.4M | 1 | 1 | 24,542,055 |
| Tx303 | Inbred | Ames 19327 | Discovery | 2.1M | 3 | 3 | 69,017,698 |
| W22 | Inbred | NSL 30053 | Discovery | 2.1M | 2 | 1 | 11,254,201 |
| W64a | Inbred | PI 587152 | Profiling | 1.4M | 1 | 1 | 22,014,734 |
| CML411 | Inbred | CML411 | Prediction | 270k | 1 | 0 | NA |
| CML418 | Inbred | CML418 | Prediction | 270k | 1 | 0 | NA |

| Genotype | Plant Type | Source | Methylation Set | Methylation Platform | meDIP Replicates | RNA-seq replicates | RNA-seq reads (million bp) |
|---|---|---|---|---|---|---|---|
| CML479 | Inbred | CML479 | Prediction | 270k | 1 | 0 | NA |
| CML511 | Inbred | CML511 | Prediction | 270k | 1 | 0 | NA |
| CML69 | Inbred | Ames 28184 | Prediction | 270k | 1 | 0 | NA |
| CML85 | Inbred | CML85 | Prediction | 270k | 1 | 0 | NA |
| CML96 | Inbred | CML96 | Prediction | 270k | 1 | 0 | NA |
| CML99 | Inbred | CML99 | Prediction | 270k | 1 | 0 | NA |
| NC350 | Inbred | Ames 27171 | Prediction | 270k | 1 | 0 | NA |
| TIL07 | Teosinte | Ames 28402 | Prediction | 270k | 1 | 0 | NA |
| TZI8 | Inbred | PI 506246 | Prediction | 270k | 1 | 0 | NA |
| TIL17 | Teosinte | Ames 28409 | Prediction | 270k | 1 | 0 | NA |
| IBM0008 | RIL | M0008 | RIL stability | 270k | 1 | 0 | NA |
| IBM0012 | RIL | M0012 | RIL stability | 270k | 1 | 0 | NA |
| IBM0032 | RIL | M0032 | RIL stability | 270k | 1 | 0 | NA |
| IBM0048 | RIL | M0048 | RIL stability | 270k | 1 | 0 | NA |
| IBM0054 | RIL | M0054 | RIL stability | 270k | 1 | 0 | NA |
| IBM0061 | RIL | M0061 | RIL stability | 270k | 1 | 0 | NA |
| IBM0062 | RIL | M0062 | RIL stability | 270k | 1 | 0 | NA |
| IBM0099 | RIL | M0099 | RIL stability | 270k | 1 | 0 | NA |
| IBM0298 | RIL | M0298 | RIL stability | 270k | 1 | 0 | NA |
| IBM0311 | RIL | M0311 | RIL stability | 270k | 1 | 0 | NA |
| IBM0315 | RIL | M0315 | RIL stability | 270k | 1 | 0 | NA |
| IBM0325 | RIL | M0325 | RIL stability | 270k | 1 | 0 | NA |
| IBM0352 | RIL | M0352 | RIL stability | 270k | 1 | 0 | NA |
| IBM0357 | RIL | M0357 | RIL stability | 270k | 1 | 0 | NA |
| IBM0360 | RIL | M0360 | RIL stability | 270k | 1 | 0 | NA |
| IBM0368 | RIL | M0368 | RIL stability | 270k | 1 | 0 | NA |
| IBM0379 | RIL | M0379 | RIL stability | 270k | 1 | 0 | NA |

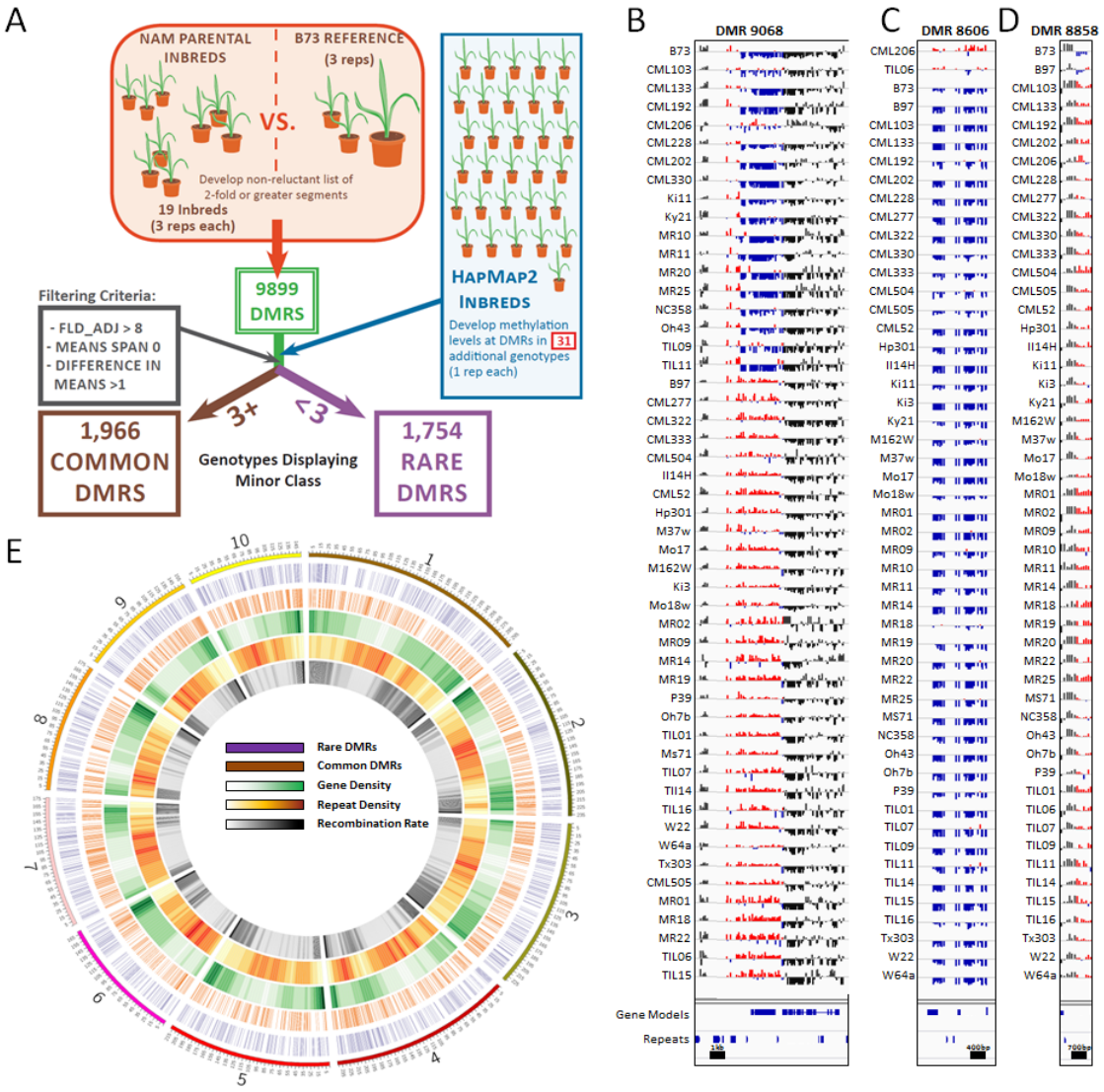



Figure 1: (A) An outline of the methods to identify DMRs. DMR discovery was performed by contrasting replicated measurements of 19 NAM parental lines (ref) with B73. The methylation levels for these 9,899 regions was also determined for a single replicate of 31 additional genotypes that were present in the HapMap2 panel (Chia et al., 2012). A series of filtering criteria were applied to identify robust common (3 or more genotypes with both high and low methylation) or rare (only one or two genotypes with alternate methylation state) DMRs. (B-D) Examples of common (B), rare methylated (C), and rare unmethylated (D) DMRs are visualized in all 51 genotypes. In each case the genotypes are ordered according to DNA methylation state. The red (high) and blue (low) indicate variable methylation with the DMR while probes with black coloring are outside of the DMR. The genes and repetitive elements annotated in each region are shown at the bottom. (E) A genome-wide view of DMRs in relationship to other genomic features. Circos (Krzywinski, et al. 2009) was used to show (outside to inside tracks): rare DMRs (purple ticks), common DMRs (brown ticks), gene density in 1Mb windows based on annotation from Schnable et al., 2009 (green is high), repeat density in 1Mb windows based on (orange is high), recombination rate as cM/Mb from Liu et al., 2009 (black is high).

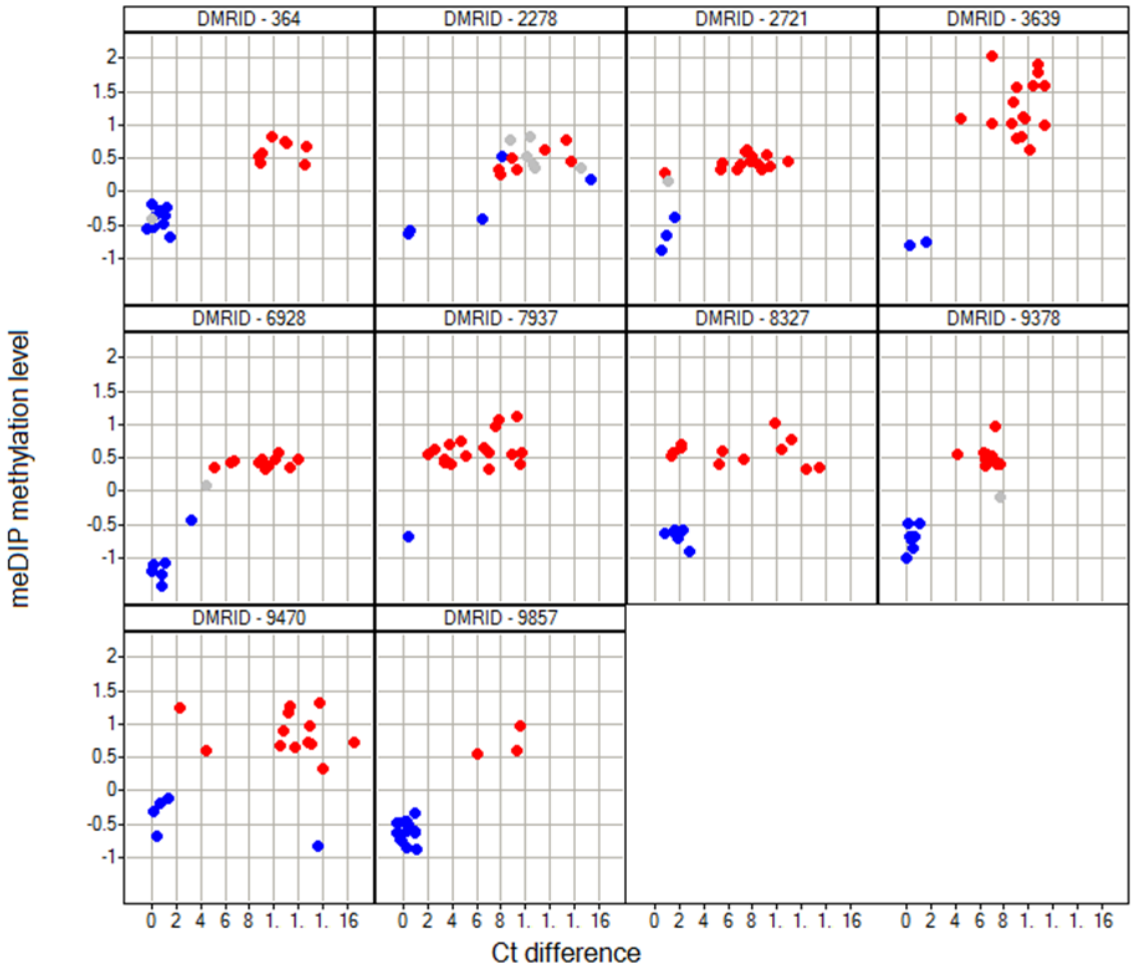



Figure 2: Validation of DMRs by methylation-sensitive qPCR. Values are colored according to the methylation classification (red = methylated, blue = unmethylated, grey = unclassified) from the meDIP profiling. Ten of the DMRs were selected for locus-specific validation assays. The level of DNA methylation was assessed in 18 genotypes by qPCR on genomic DNA following digestion with the methyl-dependent enzyme *FspEI* or with a mock-treatment. The level of DNA methylation detected using this assay was highly correlated with the level of DNA methylation estimated by the meDIP-chip analysis. For six of the ten DMRs assessed, the locus-specific assay completely supported the genome-wide profiling calls. For the other four DMRs the locus-specific and genome-wide profiles matched for at least 15 of the 18 genotypes surveyed within each DMR

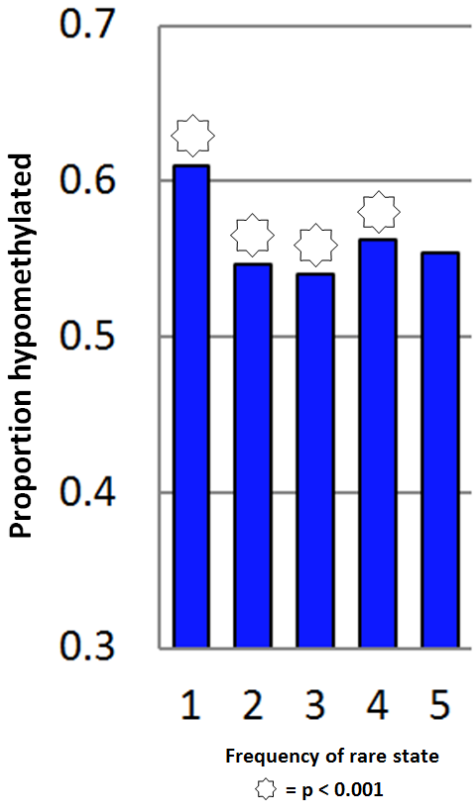



Figure 3: Significant enrichment for hypomethylation in rare DMRs

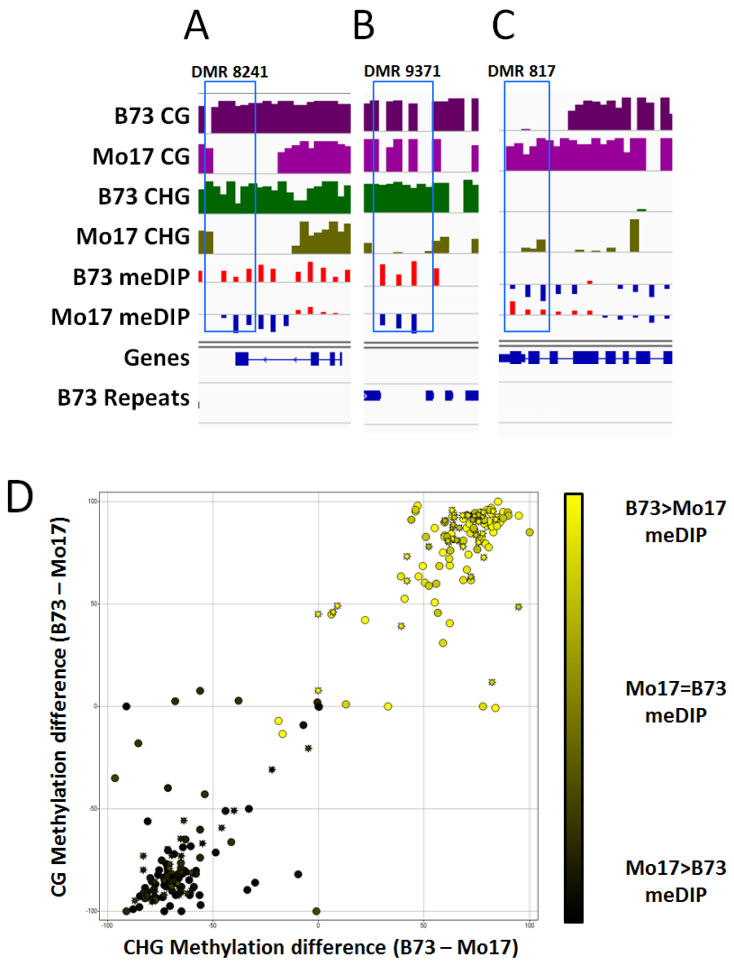



Figure 4: Validation of differential methylation levels by MethylC-Seq data on an independent sample of B73 and Mo17. (A-C) provide examples of DMRs that exhibit differences in CG and CHG (A), CHG only (B) or CG only (C). The CG methylation levels are shown in different shades of red and CHG methylation are shown in different shades of green (for B73 and Mo17. The meDIP-array methylation estimates are shown for both B73 and Mo17 with high methylation indicated by red and low methylation indicated by blue. The annotation of genes and repetitive sequences for each region is shown at the bottom. (D) For the 244 DMRs that exhibit significant meDIP-array variation and have full coverage of MethylC-Seq reads in both B73 and Mo17 the relative levels of CG (y-axis) and CHG (x-axis) methylation are shown. The color coding of each DMR indicates the meDIP-array difference in DNA methylation levels between B73 and Mo17 (yellow – higher in B73; black – higher in Mo17). The majority of DMRs show substantial differences in both CG and CHG methylation in the direction predicted by the meDIP-array data. A small number of the DMRs only exhibit differences in CG or CHG methylation or do not show any difference in methylation levels in this independent sample of B73 and Mo17.

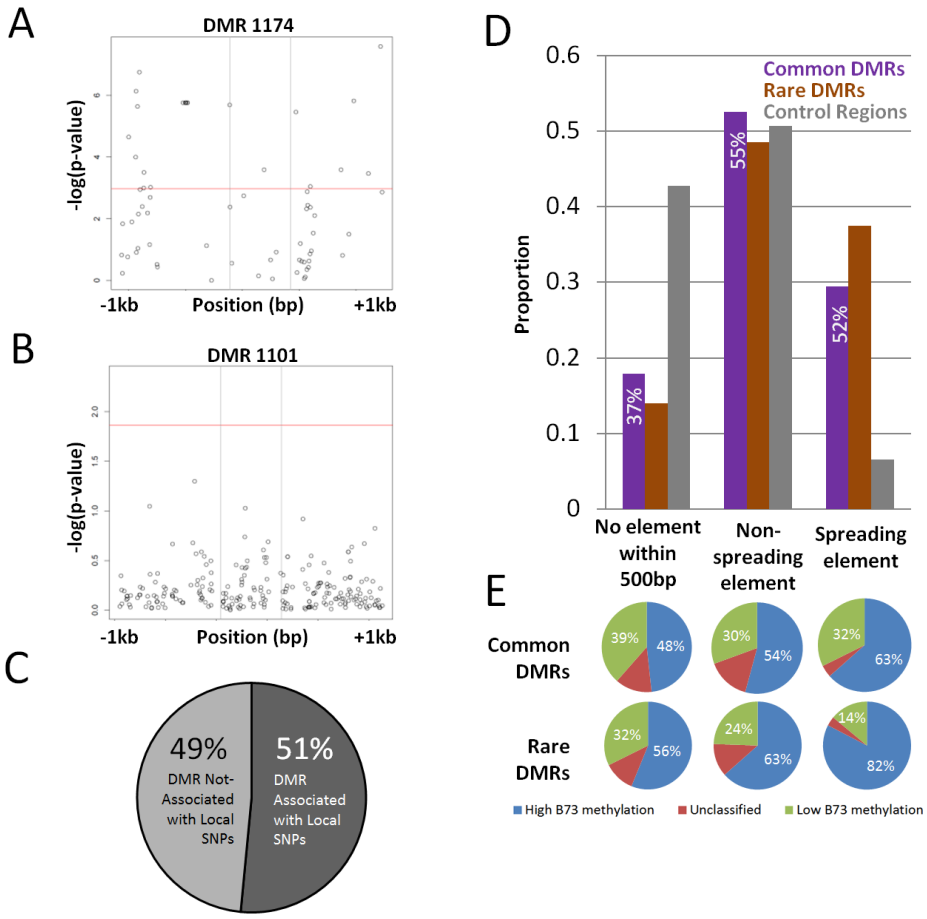



Figure 5: Many common DMRs are associated with local genetic variation (A) An example DMR displaying significant association to local SNPs. Vertical lines indicate DMR boundaries. Horizontal line (red) indicates 1% quantile p-value cutoff based on permutation analysis of other SNPs with the methylation variation for this region. SNPs above this line display significant association to DMR methylation state. (B) Example DMR displaying no significance for local SNPs with methylation state. (C) The proportion of DMRs with and without significant association with local SNPs. (D) Enrichment for heterochromatin spreading transposable elements near both common and rare DMRs. Common and rare DMRs were mapped to nearby repetitive elements within 500bp and classified as having no annotated repeat, non-spreading elements, or heterochromatin spreading elements (based on genome-wide B73 reference genome annotations from Schnable et al., 2009 and spreading assignments from Eichten et al., 2012). The DMRs were compared to a set of 10,000 control regions selected to reproduce features of our experimental DMRs. The common and rare DMRs are enriched for having spreading transposable elements within 500bp compared to the control regions. For each of the three group, the proportion of common DMRs associated to local SNPs is presented (white text in purple bars). (E) For each of the three repeat classes in Figure 3D, the proportion of high (> 80%) and low (< 20%) CG methylation in B73 is presented for both common and rare DMRs. An increase in methylation level in B73 is observed for DMRs near spreading elements.

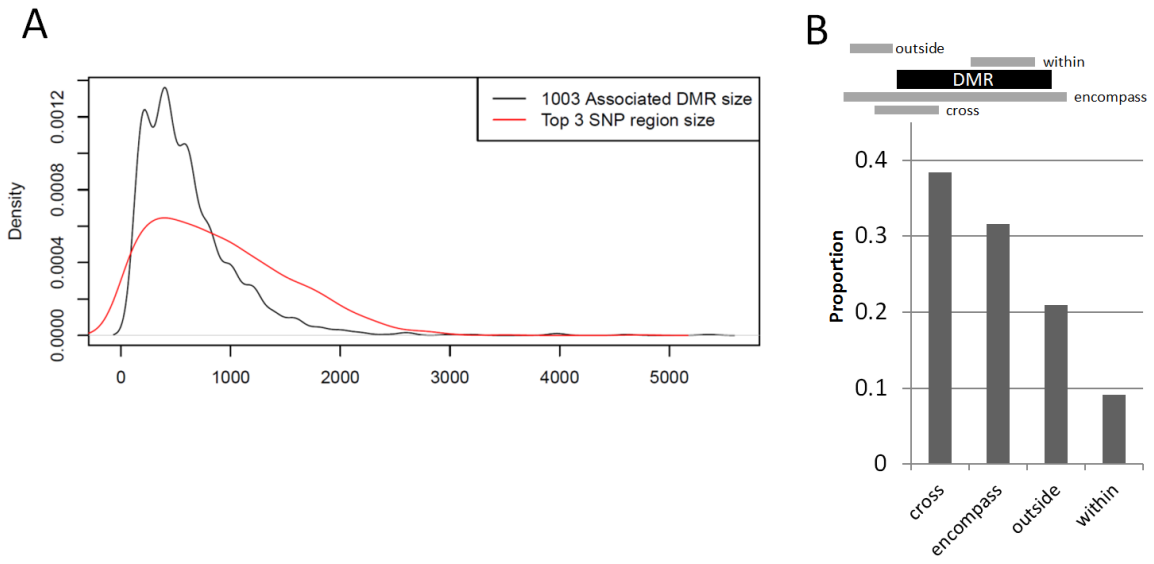



Figure 6: (A) Density plot of size (bp) of DMRs significantly associated with local genetic variation (n = 1003) compared to the region defined by the 3 SNPs displaying the highest significance in association per DMR (red). (B) Significant SNPs are found in larger regions that cross or encompass DMR boundaries. Four possible regions containing SNPs are shown relative the DMR position. The proportion of each of these classes is presented below.

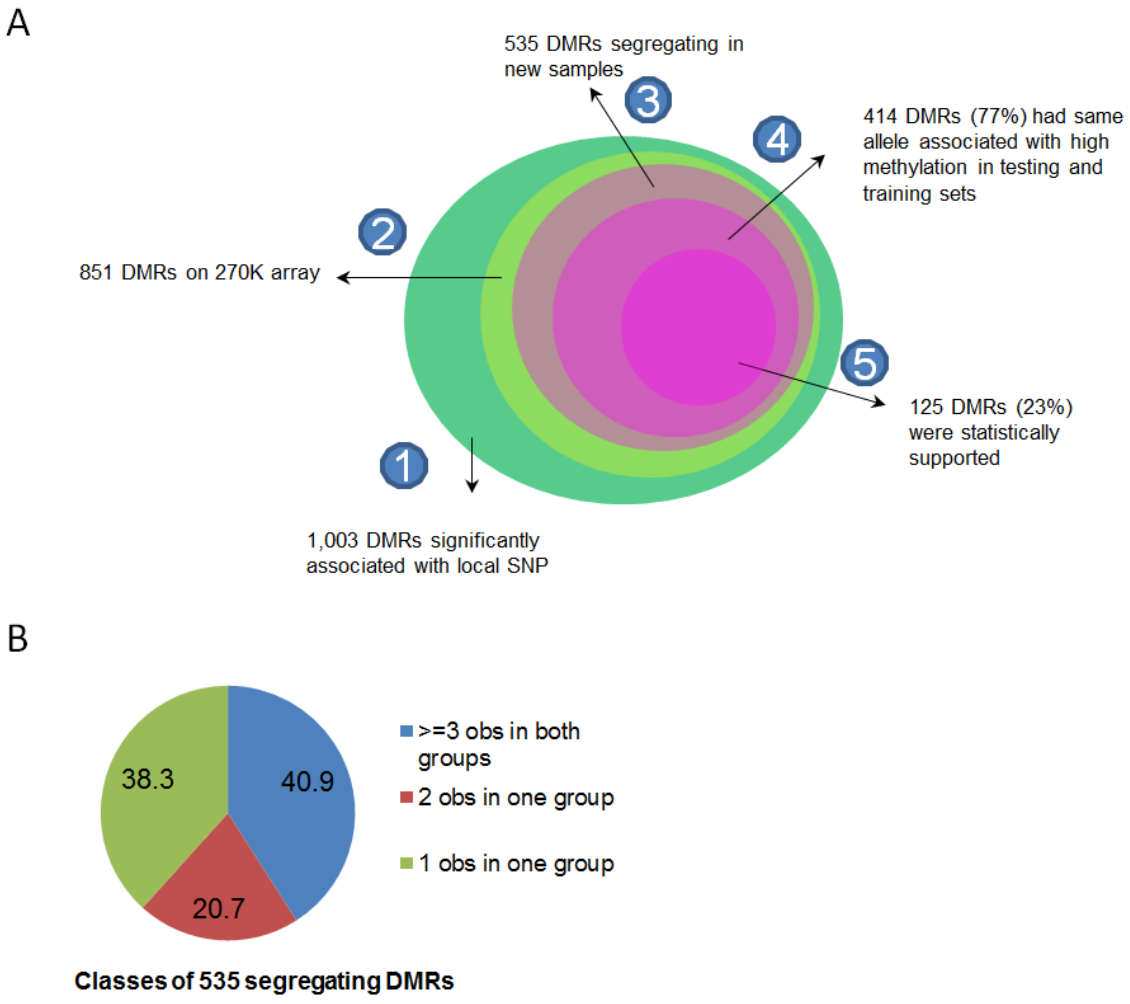



Figure 7: Prediction of methylation state based on genotype. (A) A diagram of how DMRs were filtered to allow for predictions to occur. (B) Proportion of 535 segregating DMRs that display 1 (green), 2 (red), or 3+ (blue) genotypes with the minor allele.

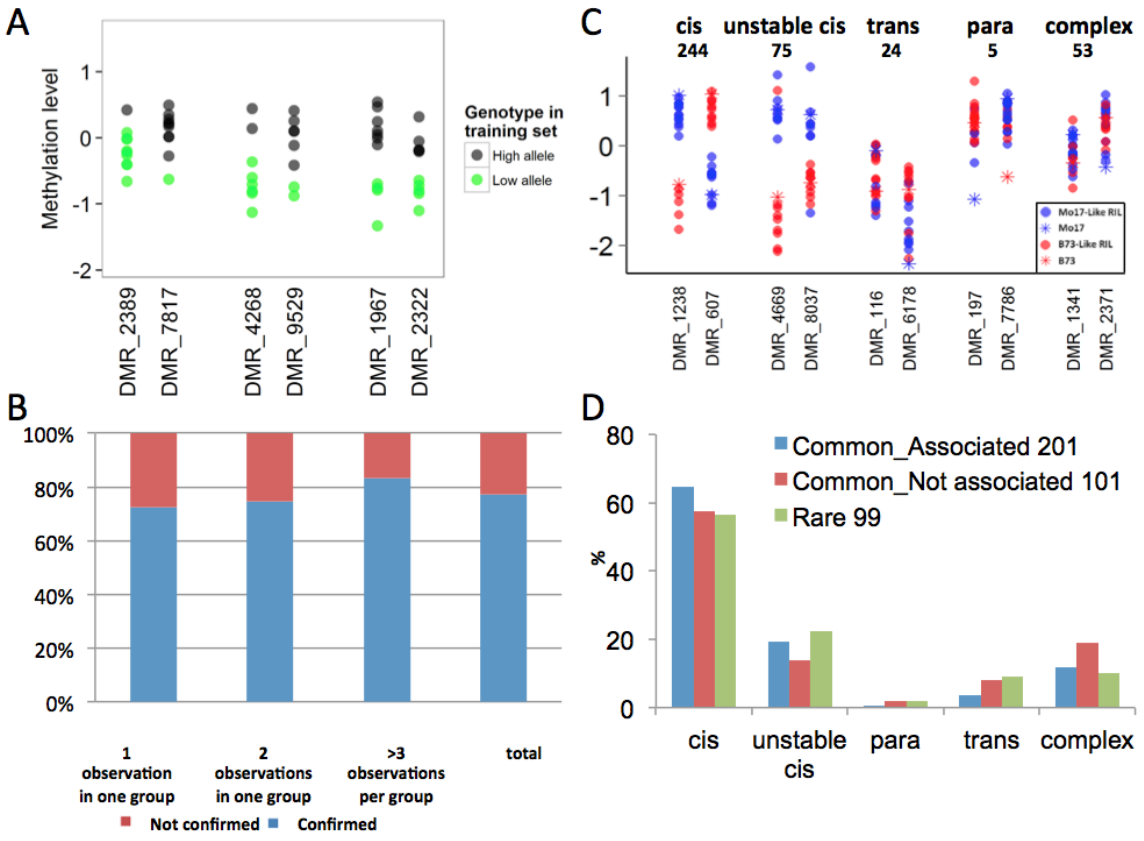



Figure 8: DMRs appear heritable and predictable across diverse and recombinant inbred lines. (A) Example methylation values for additional genotypes. Color indicates the predicted 'high methylation' and 'low methylation' allele based on their genotype calls. (B) Prediction results were divided into those with a single, two, and 3 or more genotypes displaying each allele. The proportion of DMRs where the prediction was confirmed (blue), confirmed and significant (red), and not confirmed (green) are presented. (C) Examples of five classes of DMR stability across 16 B73-Mo17 recombinant inbred lines (RILs) are shown. These are divided into locally inherited 'cis' patterns, locally inherited 'cis' patterns with occasionally methylation state shift, remote inheritance of methylation state by non-local region of the genome, DMRs displaying paramutation like states in which all lines regardless of local genetic content appear like one parent, and complex DMRs that display methylation state instability or multi-region control. The total number of DMRs for each category is displayed below. (D) The proportion of each RIL inheritance class are presented. Few changes in inheritance states observed across rare DMRs as well as genetically-associated / non-associated DMRs.

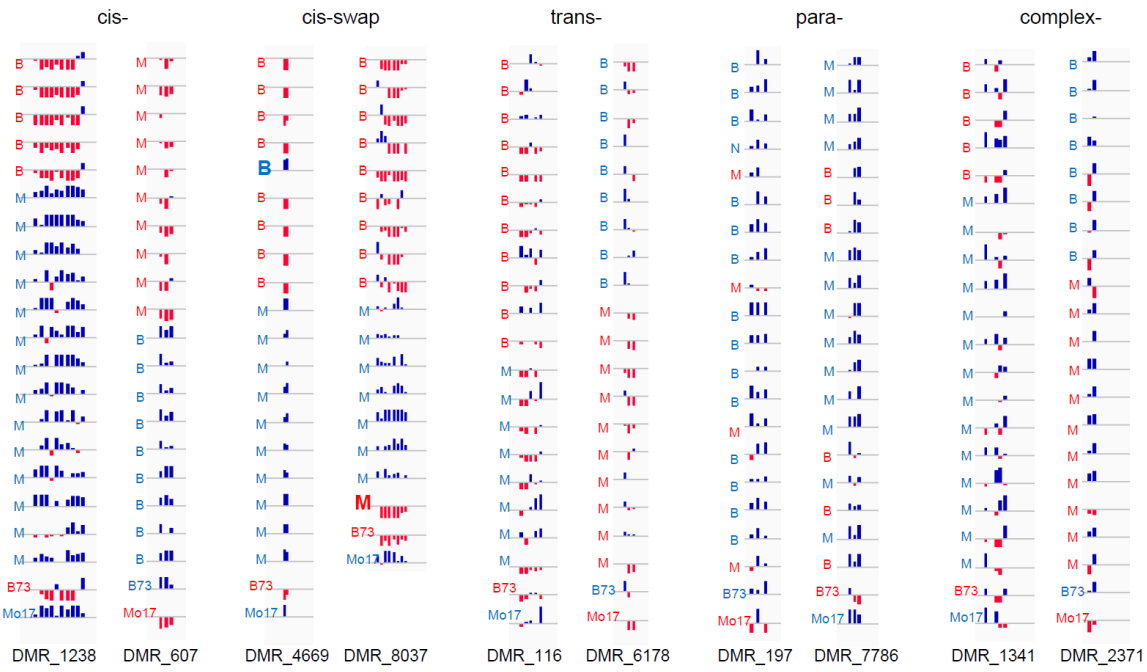



Figure 9: IGV view of RIL lines displayed in Figure 8C

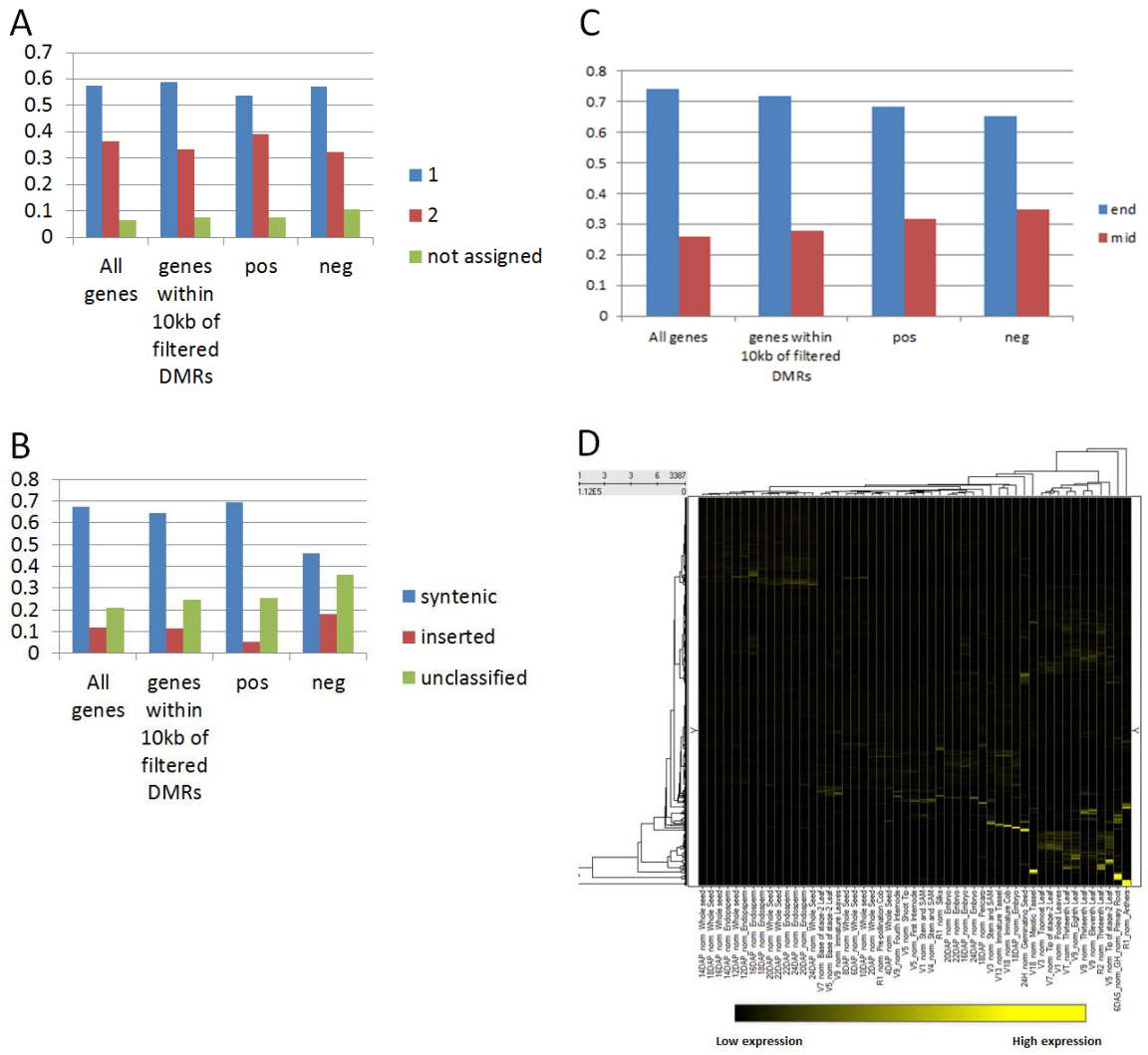



Figure 10: Analysis of genes located near DMRs. (A) The proportion of all maize filtered genes, filtered genes within 10kb of all 1966 common and 1754 rare DMRs, and those genes displaying significant positive or negative association to local DMR state that display subgenome classification of maize genes (Schnable et al., 2011) was plotted. No enrichment for either subgenome 1 or 2 was present in the data. (B) Proportion of genes displaying syntenic or inserted positioning within their genomic context (Schnable et al., 2012) was calculated. An enrichment of inserted genes negatively correlated to methylation state was observed. (C) All maize chromosomes were subdivided into high-recombination chromosome arms or pericentromeric low-recombination regions. The proportion of genes in each category was calculated. (D) Hierarchical clustering of genes within 10kb of DMRs across the maize expression atlas was performed. A per-gene normalization was performed by dividing the expression value by the average expression for each gene. Clusters of tissue-specific expression are observed. A clear separation of anther and root tissues displaying tissue-specificity and high expression was observed.

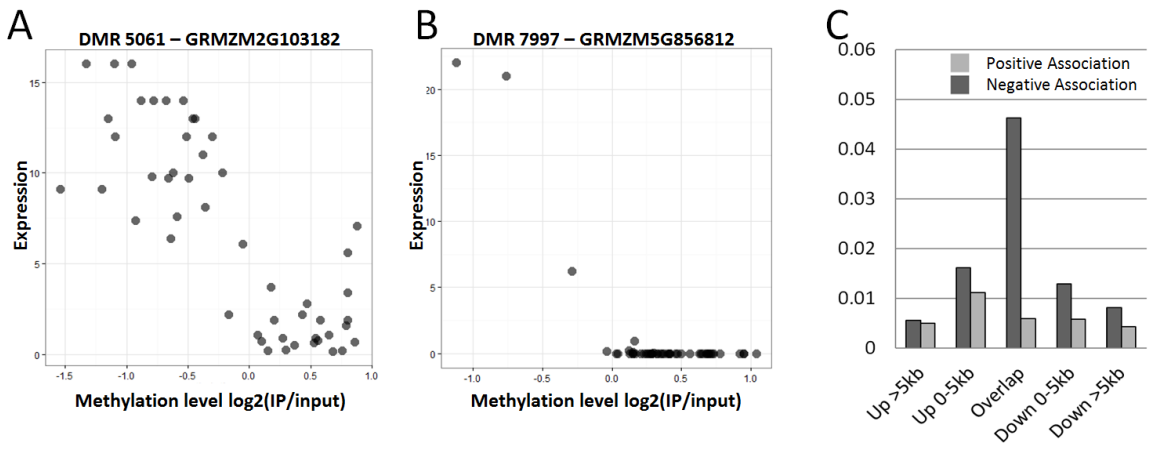



Figure 11: DMRs associated with gene expression state. Examples of Common (A) and rare (B) DMRs showing correlation to nearby gene expression state. Y axis displays log(FPKM) values for genes compared to array methylation value across 50 genotypes. B73 and Mo17 are highlighted for comparison. (C) The position of DMRs in relationship to their associated gene are displayed for both common (left) and rare (right) DMRs. An enrichment for negative methylation-expression correlations for DMRs overlapping genes is present in both classes.

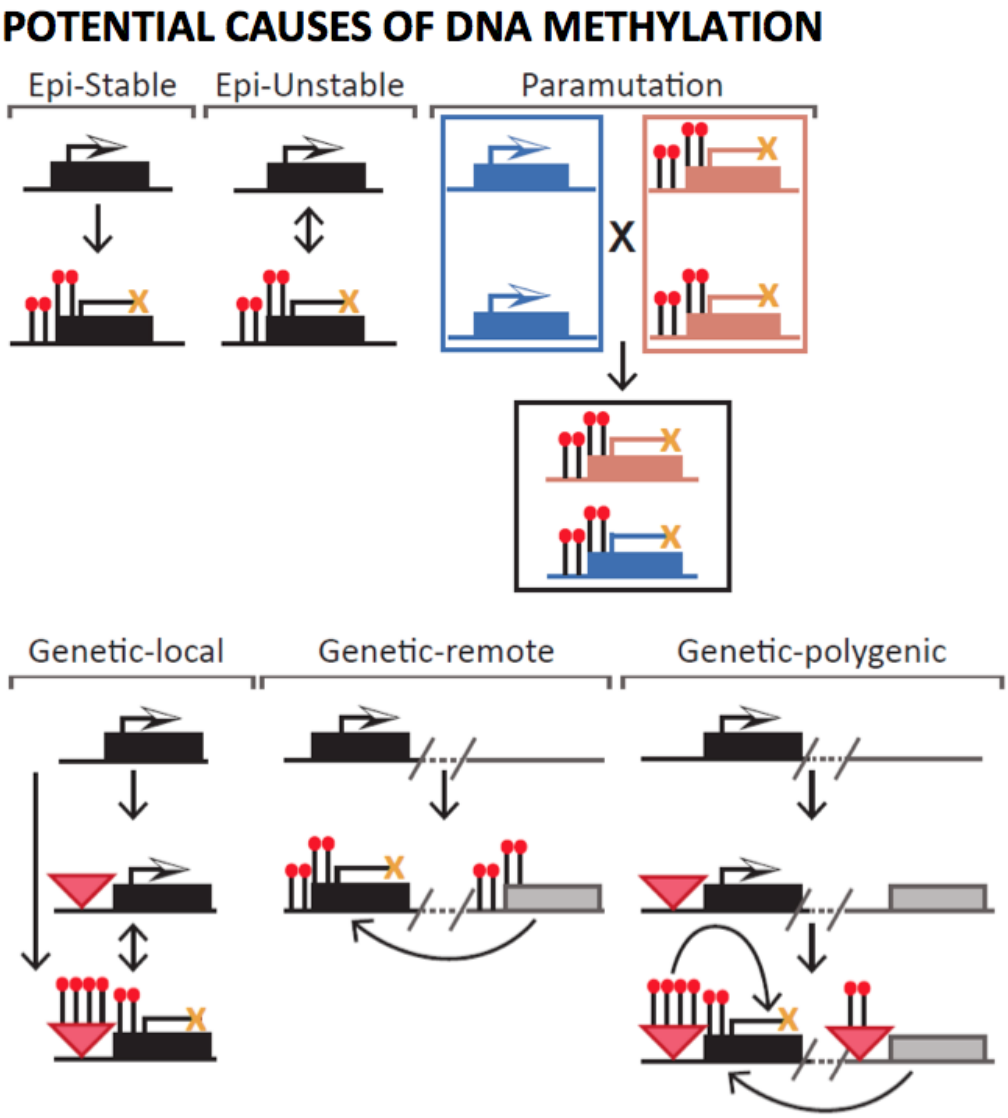



Figure 12: Diagram of potential causes of DNA methylation. DNA methylation variation may act as either an epigenetic mark independent from genetic variation (top), or as a heterochromatin mark linked to genetic variation (bottom). Epi-Stable variation acts independently of genetic context and is stable through generations. Epi-Unstable acts in a similar fashion, however the instability of methylation states allows for reversion of methylation state at some frequency. Paramutation acts through a unique mechanism due to the activity of differently methylated alleles in the same nucleus of a hybrid. DNA methylation linked to genetic variation can be separated into three distinct categories. Genetic-local, where DNA methylation may be either controlled or facilitated by a local genetic variation such as a transposon insertion. Genetic-remote, in which a genetic variant in *trans* acts to change the methylation state at the observed loci. The final class, Genetic-polygenic, involves the actions of both local and *trans*-acting factors to initiate a DNA methylation change at the observed loci. This mechanism is expected to be complex in nature requiring the knowledge of all controlling site genotypes in order to predict DNA methylation state.

**CLOSING REMARKS**

The methylation of genomic DNA within plant systems has been viewed as both a heterochromatin mark as well as a possible source of epigenetic variation across both mitotic and meiotic cellular divisions. The work presented here provided detailed analysis of DNA methylation variation within maize at both the mitotic and meiotic levels, and an attempt to disentangle the relationship between DNA methylation and genetic variation.

It is clear that there are thousands of regions within the maize genome that display DNA methylation variation in the form of differentially methylated regions (DMRs). These regions are found across the genome. The number of regions identified is somewhat less important as a variety of filtering and technological limitations impact the final number, however the high-confidence regions described have a high validation rate using alternative assessment techniques. Examples of DMRs that fall within identical-by-decent regions between the two genotypes provide support for possible pure epialleles within maize. Beyond this, the stability of DNA methylation states was profiled using a panel of near isogenic lines providing support for local regulation of DMR state, even for those DMRs within IBD regions. The ability to identify and validate DMRs identified with our microarray-based platform provides the support to conduct the more complex studies presented in later chapters.

As a heterochromatin mark that can be actively added, or removed, from genomic DNA, there is reason to suspect that the transient regulation of genes by DNA methylation could play a role in proper plant development. Surprisingly an analysis of DNA methylation across different developmental tissues of maize resulted in little evidence for DNA methylation variation across developmental time. Although this study did not identify a major role of DNA methylation in plant development, it did show that DNA methylation patterns are extremely similar across tissues and that the majority of variation is seen across lines rather than across tissues.

With the identification and rate of differential methylation across maize profiled, a need to separate out epigenetic signals from those controlled by genetic factors was conducted. As transposons are known the be repressed by stable DNA methylation, and that maize has a large proportion of its genome devoted to these elements, profiling the effects of transposon heterochromatin on low-copy intergenic space was performed. DNA methylation of adjacent low-copy sequence was compared across different types of repetitive elements. Specific families

of retrotransposable elements in the maize genome exhibited spreading of DNA methylation into the nearby low-copy sequence. In cases where these spreading elements displayed presence-absence variation between maize lines, DMRs were identified. Without knowing the insertion pattern of these spreading elements, it is possible that many of the DMRs identified in these studies are linked to the presence of these retrotransposons. This study produced examples of a possible phenomenon leading to variable DNA methylation patterns between different lines that are directly linked to the genetic content of the individual.

The link between genetic variation and DNA methylation variation was further profiled by assessing the methylation state of a variety of diverse maize inbred lines. DMRs exist between many different lines as both rare and common variants. DMRs assessed appear predominantly across all sequence contexts and roughly half of all DMRs show significant association to local genetic variation within their background. This connection to genetic state allowed methylation states to be predicted across other diverse inbreds based on their known genetic variants. DMRs were also shown to impact the expression level of nearby genes, providing a possible pathway for phenotypic variation to be due to DNA methylation variation.

The results presented here outline a complex relationship between DNA methylation, genetic variation, and phenotype. Future research will definitely involve more detailed methylation profiling methods, such as bisulfite sequencing, along with acute phenotyping of individuals to determine how genes and pathways are influenced by DNA methylation regulation. As DNA methylation may allow for a more plastic regulation of genes compared to genetic mutation, understanding the role of DNA methylation variation due to environmental stress or other genomic perturbations should be conducted in the future.

With maize as a crop system, the application of this work for plant breeding leads to unique conclusions compared to a purely scientific endeavor. DNA methylation does not appear to contain a large proportion of the 'missing heritability' in crop traits that is not being captured for breeding purposes. Many DMRs do not appear to affect nearby genes, and individual DMRs may not have any major downstream phenotypic effect. Beyond this, with roughly half of the DMRs identified showing associations to local genetic variation, it is possible that many methylation states could be predicted based on current genotypic assays eliminating the need for direct epigenetic profiling of germplasm in a breeding program. Indeed there are examples of DMRs

that appear to act independently of genetic variation, are trans-generationally stable, and can impact local gene expression. However, as these are the minority, plant breeders would not gain much more applicable information by directly profiling the DNA methylation patterns of their germplasm.

Instead, as DNA methylation is often associated with transposable elements, and it may be likely that many of the genetically-associated DMRs could be caused by the presence or absence of transposable element insertions, breeders could obtain more useful information by identifying transposon insertion sites across diverse maize lines. As transposons can impact the local methylation state, while also impacting the local genetic state by insertion, having a detailed view of transposon locations could provide breeders with higher-value information to integrate into their breeding programs.

# Bibliography

Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H (2011) Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Research 39: 6919–6931.

Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, et al. (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442: 295–298.

Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, et al. (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proceedings of the National Academy of Sciences of the United States of America 107: 12986–12991.

Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, et al. (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proceedings of the National Academy of Sciences of the United States of America 109: 10444-10449

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, et al. (2009) Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genetics 5: e1000732.

Becker C, Hagmann J, Müller J, Koenig D, Stegle O, et al. (2011) Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature 480: 245–249.

Belo A, Beatty MK, Hondred D, Fengler KA, Li B, et al. (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theoretical and Applied Genetics 120: 355–367.

Bender J, Fink GR (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell 83: 725–734.

Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome / National Research Council Canada 37: 565–576.

Bernatavichute Y V, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana. PloS ONE 3: e3156.

Biémont C (2010) A brief history of the status of transposable elements: from junk DNA to major players in evolution. Genetics 186: 1085–1093.

Biémont C, Vieira C (2006) Junk DNA as an evolutionary force. Nature 443: 521–524.

Bird A (2007) Perceptions of epigenetics. Nature 447: 396–398.

Bird A (1997) Does DNA methylation control transposition of selfish elements in the germline? Trends in Genetics13: 469–472.

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. Genome Biology 12: 232.

Brunner S, Fengler K, Morgante M, Tingey S, Rafalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. The Plant Cell 17: 343–360.

Buckler ES, Gaut BS, McMullen MD (2006) Molecular and functional diversity of maize. Current Opinion in Plant Biology 9: 172–176.

Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nature Reviews Genetics 6: 351–360.

Chandler VL, Eggleston WB, Dorweiler JE (2000) Paramutation in maize. Plant Molecular Biology 43: 121–145.

Chandler VL, Stam M (2004) Chromatin conversations: mechanisms and implications of paramutation. Nature Reviews Genetics 5: 532–544.

Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11: 203.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452: 215–219.

Comfort NC (2001) From controlling elements to transposons: Barbara McClintock and the Nobel Prize. Trends in Genetics17: 475–478.

Cubas P, Vincent C, Coen E (1999) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157–161.

Dempster AP, Laird NM, Rubin DB (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical SocietySeries B (Methodological) 39: 1–38.

Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, et al. (2000) Mediator of Paramutation1 is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. The Plant Cell 12: 2101–2118.

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19.1: 11–15.

Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. Proceedings of the National Academy of Sciences of the United States of America 106: 19916–19921.

Eichten SR, Foerster JM, de Leon N, Kai Y, Yeh CT, et al. (2011) B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiology 156: 1679–1690.

Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, et al. (2011) Heritable epigenetic variation among maize inbreds. PLoS Genetics 7: e1002372.

Emberton J, Ma J, Yuan Y, SanMiguel P, Bennetzen JL (2005) Gene enrichment in maize with hypomethylated partial restriction (HMPR) libraries. Genome Research 15: 1441–1446.

Fedoroff N, Schlappi M, Raina R (1995) Epigenetic regulation of the maize Spm transposon. BioEssays 17: 291–297.

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. Nature 447: 433–440.

Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, et al. (2010) Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America 107: 8689–8694.

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics 9: 397–405.

Finnegan EJ (2002) Epialleles - a source of random variation in times of stress. Current Opinion in Plant Biology 5: 101–106.

Finnegan EJ, Peacock WJ, Dennis ES (1996) Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proceedings of the National Academy of Sciences of the United States of America 93: 8449–8454.

Franklin GC, Adam GI, Ohlsson R (1996) Genomic imprinting and mammalian development. Placenta 17: 3–14.

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proceedings of the National Academy of Sciences of the United States of America 99: 9573–9578.

Gdula DA, Gerasimova TI, Corces VG (1996) Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. Proceedings of the National Academy of Sciences of the United States of America 93: 9378–9383.

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

Gent JI, 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 Research 40: 1550–1560.

Girard L, Freeling M (1999) Regulatory changes as a consequence of transposon insertion. Developmental Genetics 25: 291–296.

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, et al. (2009) A first-generation haplotype map of maize. Science 326: 1115–1117.

Haag JR, Pikaard CS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature Reviews Molecular Cell Biology 12: 483–492.

Hansey CN, Johnson JM, Sekhon RS, Kaeppler SM, Leon N de (2011) Genetic Diversity of a Maize Association Population with Restricted Phenology. Crop Science 51: 704–715.

Haun WJ, Laoueille-Duprat S, O'connell MJ, Spillane C, Grossniklaus U, et al. (2007) Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. The Plant Journal 49: 325–337.

Haun WJ, Springer NM (2008) Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. The Plant Journal 56: 903–912.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491.

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Research 19: 1419–1428.

Hollister JD, Smith LM, Guo Y-L, Ott F, Weigel D, et al. (2011) Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. Proceedings of the National Academy of Sciences of the United States of America 108: 2322–2327.

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451–1454.

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN Epigenetic Alleles in Arabidopsis. Science 277: 1100–1103.

Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, 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 Genetics 5: e1000737.

Jirtle RL, Skinner MK (2007) Environmental epigenomics and disease susceptibility. Nature Reviews Genetics 8: 253–262.

Johannes F, Colome-Tatche M (2011) Quantitative epigenetics through epigenomic perturbation of isogenic lines. Genetics 188: 215–227.

Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: a quantitative genetics perspective. Nature Reviews Genetics 9: 883–890.

Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, et al. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. PLoS Genetics 5: e1000530.

Kalisz S, Purugganan MD (2004) Epialleles via DNA methylation: consequences for plant evolution. Trends in Ecology & Evolution 19: 309–314.

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nature Genetics 33: 102–106.

Kent WJ (2002) BLAT---The BLAST-Like Alignment Tool. Genome Research 12: 656–664.

Kohler C, Weinhofer-Molisch I (2010) Mechanisms and evolution of genomic imprinting in plants. Heredity 105: 57–63.

Lauria M, Rupe M, Guo M, Kranz E, Pirona R, et al. (2004) Extensive maternal DNA hypomethylation in the endosperm of Zea mays. The Plant Cell 16: 510–522.

Levin HL, Moran J V (2011) Dynamic interactions between transposable elements and their hosts. Nature Reviews Genetics 12: 615–627.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, et al. (2010) The developmental dynamics of the maize leaf transcriptome. Nature Genetics 42: 1060–1067.

Li W, Liu H, Cheng ZJ, Su YH, Han HN, et al. (2011) DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. PLoS Genetics 7: e1002243.

Lippman Z, Gendrel A V, Black M, Vaughn MW, Dedhia N, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.

Lisch D (2002) Mutator transposons. Trends in Plant Science 7: 498–504.

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annual Review of Plant Biology 60: 43–66.

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Current Opinion in Plant Biology 14: 156–161.

Lisch D, Carey CC, Dorweiler JE, Chandler VL (2002) A mutation that prevents paramutation in maize also reverses Mutator transposon methylation and silencing. Proceedings of the National Academy of Sciences of the United States of America 99: 6130–6135.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.

Liu J, He Y, Amasino R, Chen X (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. Genes & Development 18: 2873–2878.

Liu S, Yeh CT, Ji T, Ying K, Wu H, et al. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genetics 5: e1000733.

Lukens LN, Zhan S (2007) The plant genome's methylation status and response to stress: implications for plant improvement. Current Opinion in Plant Biology 10: 317–322.

Makarevitch I, Stupar RM, Iniguez AL, Haun WJ, Barbazuk WB, et al. (2007) Natural variation for alleles under epigenetic control by the maize chromomethylase zmet2. Genetics 177: 749–760.

Manning K, Tor M, Poole M, Hong Y, Thompson AJ, et al. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature Genetics 38: 948–952.

Martienssen R, Barkan A, Taylor WC, Freeling M (1990) Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. Genes & Development 4: 331–343.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, et al. (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature 461: 1135–1138.

McClintock B (1983) The significance of responses of the genome to challenge. Science 226: 792–801.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. Plant Molecular Biology 16: 753–770.

Messing J, Dooner HK (2006) Organization and variability of the maize genome. Current Opinion in Plant Biology 9: 157–163.

Meyers B, Tingey S, Morgante M (2001) Abundance, Distribution, and Transcriptional Activity of Repetitive Elements in the Maize Genome. Genome Research 11: 1660–1676.

Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, et al. (1994) Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes & Development 8: 1463–1472.

Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, et al. (2009) An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. The EMBO Journal 28: 1078–1086.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. Nature Genetics 23: 314–318.

Ngernprasirtsiri J, Chollet R, Kobayashi H, Sugiyama T, Akazawa T (1989) DNA methylation and the differential expression of C4 photosynthesis genes in mesophyll and bundle sheath cells of greening maize leaves. The Journal of Biological Chemistry 264: 8241–8248.

Ohtani K, Dimmeler S (2011) Epigenetic regulation of cardiovascular differentiation. Cardiovascular Research 90: 404–412.

Ohtsu K, Smith MB, Emrich SJ, Borsuk LA, Zhou R, et al. (2007) Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). The Plant Journal 52: 391–404.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, et al. (2010) The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. Science 327: 92–94.

Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, et al. (2003) Maize genome sequencing by methylation filtration. Science 302: 2115–2117.

Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM (2001) Maize chromomethylase Zea methyltransferase2 is required for CpNpG methylation. The Plant Cell 13: 1919–1928.

Pereira V, Enard D, Eyre-Walker A (2009) The effect of transposable element insertions on gene expression evolution in rodents. PloS ONE 4: e4321.

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

Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Current Opinion in Plant Biology 9: 149–156.

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, et al. (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genetics 23: 305–308.

Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, et al. (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proceedings of the National Academy of Sciences of the United States of America 100: 2538–2543.

Rakyan V, Whitelaw E. (2003) Transgenerational epigenetic inheritance. Current Biology 13: 6.

Rasmusson DC, Phillips RL (1997) Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. Crop Science 37: 303–310.

Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, et al. (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genetics 7: e1002301.

Reik W (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. Nature 447: 425–432.

Reinders J, Wulff BB, Mirouze M, Mari-Ordonez A, Dapp M, et al. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. Genes & Development 23: 939–950.

Richards EJ (1997) DNA methylation and plant development. Trends in Genetics13: 319–323.

Richards EJ (2008) Population epigenetics. Current Opinion in Genetics & Development 18: 221–226.

Richards EJ (2006) Inherited epigenetic variation--revisiting soft inheritance. Nature Reviews Genetics 7: 395–401.

Richards EJ (2011) Natural epigenetic variation in plant species: a view from the field. Current Opinion in Plant Biology 14: 204–209.

Riddle NC, Richards EJ (2002) The control of natural variation in cytosine methylation in Arabidopsis. Genetics 162: 355–363.

Riddle NC, Richards EJ (2005) Genetic variation in epigenetic inheritance of ribosomal RNA gene methylation in Arabidopsis. The Plant Journal 41: 524–532.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nature Biotechnology 29: 24–26.

Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL (1996) Demethylation-induced developmental pleiotropy in Arabidopsis. Science 273: 654–657.

Roux F, Colome-Tatche M, Edelist C, Wardenaar R, Guerche P, et al. (2011) Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. Genetics 188: 1015–1017.

SanMiguel P, C. V (2009) The LTR-retrotransposons of maize. Handbook of maize: Genetics and genomcs.

Saze H, Kakutani T (2007) Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. The EMBO Journal 26: 3641–3652.

Saze H, Shiraishi A, Miura A, Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. Science 319: 462–465.

Schmitz RJ, Ecker JR (2012) Epigenetic and epigenomic variation in Arabidopsis thaliana. Trends in Plant Science 17: 149–154.

Schmitz RJ, Schultz MD, Urich M a., Nery JR, Pelizzola M, et al. (2013) Patterns of population epigenomic diversity. Nature 495: 193-198.

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PloS ONE 6: e17855.

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America 108: 4069–4074.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Skinner MK (2011) Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Research93: 51–55.

Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. Genetics 182: 845–850.

Slotkin RK, Freeling M, Lisch D (2005) Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. Nature Genetics 37: 641–644.

Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics 8: 272–285.

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, et al. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

Smyth GK (2005) Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor . New York: Springer. 397–420.

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3: 3.

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, et al. (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Molecular Cell 6: 791–802.

Springer NM, Gutierrez-Marcos F. J (2009) Imprinting in Maize. Handbook of Maize. University of Warwick Warwick HRI: Springer New York. 429–440.

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, et al. (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genetics 5: e1000734.

Stokes TL, Kunkel BN, Richards EJ (2002) Epigenetic variation in Arabidopsis disease resistance. Genes & Development 16: 171–182.

Suter CM, Martin DI, Ward RL (2004) Germline epimutation of MLH1 in individuals with multiple cancers. Nature Genetics 36: 497–501.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, et al. (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Research 20: 1689–1699.

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Trapnell C, Williams B a, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology 28: 511–515.

Vaughn MW, Tanurdzic M, Lippman Z, Jiang H, Carrasquillo R, et al. (2007) Epigenetic natural variation in Arabidopsis thaliana. PLoS Biology 5: e174.

Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23: 657–663.

Vicient CM (2010) Transcriptional activity of transposable elements in maize. BMC genomics 11: 601.

Wang Q, Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. Proceedings of the National Academy of Sciences of the United States of America 103: 17644–17649.

Wang X, Elling AA, Li X, Li N, Peng Z, et al. (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. The Plant Cell 21: 1053–1069.

Weil C, Martienssen R (2008) Epigenetic interactions between transposons and genes: lessons from plants. Current Opinion in Genetics & Development 18: 188–192.

Whitelaw C a, Barbazuk WB, Pertea G, Chan a P, Cheung F, et al. (2003) Enrichment of gene-coding sequences in maize by genome filtration. Science 302: 2118–2120.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. (2007) A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8: 973–982.

Woo HR, Pontes O, Pikaard CS, Richards EJ (2007) VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes & Development 21: 267–277.

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178: 539–551.

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, et al. (2010) Local DNA hypomethylation activates genes in rice endosperm. Proceedings of the National Academy of Sciences of the United States of America 107: 18729–18734.

Zhang M, Xu C, von Wettstein D, Liu B (2011) Tissue-specific differences in cytosine methylation and their association with differential gene expression in sorghum. Plant Physiology 156: 1955–1966.

Zhang X, Shiu SH, Cal A, Borevitz JO (2008) Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. PLoS Genetics 4: e1000032.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126: 1189–1201.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genetics 39: 61–69.