UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Active DNA Demethylation in the Arabidopsis Endosperm

Permalink
https://escholarship.org/uc/item/1d62f8g0

Author
Ibarra, Christian Anthony

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Active DNA Demethylation in the *Arabidopsis* Endosperm

By

Christian Anthony Ibarra

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

In

Plant Biology

In the

GRADUATE DIVISION

Of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Robert L. Fischer, Chair
Professor Daniel Zilberman
Professor Jennifer Fletcher
Professor Jasper Rine

Active DNA Demethylation in the *Arabidopsis* Endosperm

Copyright 2012

By

Christian Anthony Ibarra

Abstract

Active DNA Demethylation in the *Arabidopsis* Endosperm

By

Christian Anthony Ibarra

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Robert L. Fischer, Chair

DNA methylation is one of the most well studied repressive epigenetic marks in eukaryotes. In plants, DNA methylation silences genes and transposons, and establishes genomic imprinting. Genomic imprinting is the mono-allelic expression of a gene occurring in a parent-of-origin specific manner. Imprinted genes tend to be expressed in nutritive tissues and structures that function to support the growing embryo. In mammals, this is the placenta, and in plants this is the endosperm.

In *Arabidopsis*, the female gametophyte contains the egg and central cell, while the male gametophyte contains the vegetative cell and two sperm cells. Both the central cell and vegetative cells are companion cells to the egg and sperm cells respectively. When pollen lands on specialized cells at the tip of the gynoecium, the vegetative cell forms a pollen tube that transports the two sperm cells to the ovule. Within the ovule, one sperm fertilizes the egg cell to form the embryo, whereas a second sperm fertilizes the central cell to form the endosperm, which supports embryo development by providing nutrients.

In *Arabidopsis*, the DME DNA glycosylase carries out active DNA demethylation in the central cell of the female gametophyte, which has been shown to establish genomic imprinting in *Arabidopsis*. Recently, DME-mediated DNA demethylation has also been shown to occur in the vegetative cell of the male gametophyte, as well.

The aim of the work carried out in my thesis has been to investigate the role of DNA methylation and DME-mediated DNA demethylation in the central cell and vegetative cell. In particular, I was interested in identifying and understanding the regions in the genome where DME-mediated DNA demethylation occurs. My overall approach was to perform high-throughput bisulfite sequencing using the illumina™ platform on the *Arabidopsis* endosperm, embryo, and *dme*-mutant endosperm, and to analyze their DNA methylation profiles (methylomes). In my initial study (Chapter II), I measured non-allele-specific methylomes in which endosperm and embryo dissected from selfed-

plants were used. In my subsequent study (Chapter III), I determined allele-specific (maternal *versus* paternal) methylomes by analyzing endosperm and embryo derived from seeds generated by crossing different *Arabidopsis* ecotypes. Because there are no current methods for determining the *Arabidopsis* central cell methylome, I sought to use the maternal endosperm genome as a proxy for the maternal central cell genome. Through collaboration, I also had the opportunity to investigate and explore the methylomes of wild-type and *dme*-mutant vegetative cell and sperm cells, which enabled me to compare the maternal and paternal reproductive methylomes of the plant.

To the best of my knowledge, this is the first time a high-throughput DNA methylation study was done on the endosperm of a flowering plant. With my collaborators, we discovered that the *Arabidopsis* endosperm exhibits a genome-wide reduction in DNA methylation compared to the embryo. We also found that the DME DNA glycosylase regulates non-CG methylation in the endosperm, particularly siRNA-mediated CHH methylation. Moreover, our analysis of DME-dependent DNA demethylation in companion cells suggested a mechanism for reinforcing silencing of transposons in the plant male and female gamete cells.
We found that DME demethylated many of the same loci in both the maternal endosperm and vegetative cell genomes, suggesting that DME has a much more robust and primitive role in *Arabidopsis* reproduction than previously thought. Interestingly, many of these DME-dependent sites were heavily enriched in small transposons and repetitive elements. Moreover, these small transposons tended to reside adjacent to genes, enabling us to propose a new model for the observed phenomenon of gene-adjacent DNA demethylation in *Arabidopsis*. That is, regulation of gene imprinting may not be the basal function of DME demethylation. These studies suggest that the primary function of demethylation of transposons in companion cells may be to reinforce transposon silencing in plant gametes.

Contents

Table of Contents i
Acknowledgements vii

Chapter I 1

Introduction and Background 1
Megasporogenesis and Gametogenesis 2
Microsporogenesis and Microgametogenesis 4
Double Fertilization Produces the Embryo and Endosperm 4
Embryogenesis 6
Endosperm Development and Function 6
DNA Methylation is a Repressive Epigenetic Mark in Plants and Animals 7
The Establishment of DNA Methylation 9
Mechanisms for Maintaining CG Methylation in Plants and Mammals 11
Mechanisms for Maintaining non-CG Methylation in Plants 12
Active DNA Demethylation Occurs in Plants and Animals 12
Active DNA Demethylation in Plants 13
Active DNA Demethylation in Mammals 14
DME-Mediated DNA Demethylation in the Central Cell Establishes Genomic Imprinting in the *Arabidopsis* Endosperm 15
Epigenetic Events in the Companion Cells Ensure Gamete Genome Integrity 19
What is the Role of DME-Mediated DNA Demethylation in the *Arabidopsis* Endosperm? 20

Chapter II 22

Genome-Wide DNA Demethylation of the *Arabidopsis* Endosperm 22
INTRODUCTION 23
RESULTS 24
Sequencing and DNA methylation analysis 24
Gene body methylation occurs in the tissues of the seed 26
Ends Analysis shows the endosperm to be hypomethylated compared to the embryo 26
Most DMRs tend to be hypomethylated in the endosperm genome 29
Heat maps of methylation in genic and transposable elements further show the endosperm to be hypomethylated compared to the embryo 30
Using kernel density plots to graph DMRs is a powerful tool for studying genome-wide DNA methylation 30
Kernel density plots also reveal the effects of the dme-mutant allele on the endosperm genome 33
The methylation profiles of several known imprinted-related DMRs coincide with the genome-wide DNA demethylation event 33
The endosperm contains DME-independent hypermethylated non-CG DMRs 34
The embryo methylome is comprised of extensive CHH hypermethylation suggesting RdDM is elevated during early plant development 34
DISCUSSION 37
Genome-wide DNA demethylation in the endosperm 37
Genome-wide DNA demethylation occurs in the central cell and is an indirect effect of DME-mediated DNA demethylation 37

Is there a link between genome-wide DNA demethylation and genomic imprinting? 38
What role does DEMETER play in the endosperm genome-wide demethylation? . 39
A model for the DME-dependent up-regulation of non-CG methylation in the endosperm 39
Can genome-wide DNA demethylation in the endosperm reinforce transposon silencing in the plant germline? 41
METHODS 41
Bisulfite sequencing 42
Sequence analysis 42
Methylation analysis 43
siRNA analysis 44
Chapter III 45
DEMETER-mediated DNA demethylation targets similar loci in gamete companion cells 45
INTRODUCTION 47
RESULTS 47
Allele-specific sequencing and methylation analysis 47
Kernel Density plot analysis of the allele-specific endosperm shows a subpopulation of maternal loci potentially targeted by DME-mediated DNA demethylation 48
Ends Analysis shows the maternal genome to be hypomethylated in the CG context compared to the paternal genome on the endosperm 49
DME-mediated DNA demethylation occurs in all methylation contexts 49
DEMETER activates PRC2 to promote non-CG methylation in the endosperm 53
PRC2 has no effect on CG methylation in the endosperm 54
Analysis of the vegetative and sperm cell methylomes 54
Does DEMETER target the same loci in the central cell and vegetative cell for active DNA demethylation? 58
Where in the genome are the DME-dependent hypomethylated loci located? 58
DISCUSSION 59
DME targets loci on the maternal endosperm genome for active DNA demethylation 59
DME targets the same loci in the companion cells for active DNA demethylation .. 60
A mechanism for DME-mediated gene adjacent demethylation 60
METHODS 61
Isolation of *Arabidopsis thaliana* endosperm 61
Bisulfite sequencing library construction 62
Allele-specific determination of DNA methylation 62
Kernel density plots 62
Box-whisker plots 63
Chapter IV 64
Conclusion 64
References 68
Appendix A 77

List of Figures

FIG. 1. 1. MEGASPOROGENESIS AND MICROSPOROGENESIS IN *ARABIDOPSIS*.....3
FIG. 1. 2. DOUBLE FERTILIZATION IN ARABIDOPSIS.. 5
FIG. 1. 3. RNA DIRECTED DNA METHYLATION 10
FIG. 1. 4. THE CLASSIC MODEL OF ARABIDOPSIS GENOMIC IMPRINTING.....17

FIG. 2. 1. AN *ARABIDOPSIS* 'LATE-WALKING STICK' STAGE SEED. 25
FIG. 2. 2. ENDS ANALYSIS COMPARING OUR AERIAL TISSUE METHYLATION DATA, TO THAT OF COKUS *ET AL. (2008)*.. 27
FIG. 2. 3. PROFILES OF DNA METHYLATION IN EMBRYO, WILD-TYPE ENDOSPERM, AND *DME*-ENDOSPERM. 28
FIG. 2. 4. DNA METHYLATION HEAT MAPS OF GENES IN THE EMBRYO AND WILD-TYPE ENDOSPERM. 31
FIG. 2. 5. DNA METHYLATION HEAT MAPS OF TRANSPOSONS AND REPEATS IN THE EMBRYO AND WILD-TYPE ENDOSPERM. 32
FIG. 2. 6. KERNEL DENSITY PLOTS DEMONSTRATING GENOME-WIDE DEMETHYLATION OF THE *ARABIDOPSIS* ENDOSPERM. 36
FIG. 2. 7. ASSOCIATION BETWEEN NON-CG METHYLATED DMRS IN THE ENDOSPERM AND SIRNAS.. 36
FIG. 2. 8 A MODEL SUGGESTING HOW DME-MEDIATED DNA DEMETHYLATION MAY REGULATE CHH METHYLATION IN THE CENTRAL CELL, AS WELL AS REINFORCE THE SILENCING OF TRANSPOSONS IN THE GERM CELL 40

FIG. 3. 1 ALLELE-SPECIFIC PROFILES OF *A. THALIANA* ENDOSPERM AND EMBRYO DNA METHYLATION. ..50
FIG. 3. 2. DNA METHYLATION IN *A. THALIANA* EMBRYO. 51
FIG. 3. 3. DNA METHYLATION IN *A. THALIANA* GENES. 52
FIG. 3. 4. PROFILES OF *A. THALIANA* SPERM AND VEGETATIVE CELL 55
FIG. 3. 5. DEMETHYLATION OF SMALL TES IN VEGETATIVE CELLS REINFORCES METHYLATION IN SPERM.56
FIG. 3. 6. CG METHYLATION NEAR *A. THALIANA* IMPRINTED GENES. 57

Appendix

Fig. A2.1. A schematic of library construction for Illumina™ sequencing of bisulfite converted DNA.....77
Fig. A2.2. Illustration of kernel density plots from two fictitious tissues.....78
Fig. A2.3. Kernel density plots tracing the frequency distribution of the differences between embryo and aerial tissue methylation.....78

List of Tables

Table. 2.1. Bulk methylation amounts in embryo, endosperm, *dme*-endosperm, and aerial tissue.....25

Table. 2.2. Differentially methylated regions (DMRs) between embryo and endosperm……25

Appendix
Table. 2.1. Methylation statistics for embryo, WT endosperm, *dme* endosperm and aerial tissues……...78

Acknowledgements

In the Name of God, The Most Gracious, The Most Merciful.

First and foremost, I would like to thank my advisor Robert Fischer, for taking me as a graduate student, over six years ago. He has been an incredible mentor to me, not only as scientist but as a human being. It has truly been an honor being a graduate student in his lab.

Next, I have to thank Dr. Tzung-Fu Hsieh, who I have sat next to for the last six years of my life. I am forever indebted to him for teaching me how to actually do science, and I am not sure I'd be finishing graduate school if it were not for his help.

I am grateful to Professor Daniel Zilberman for exposing me to cutting-edge science, and intellectually stimulating lab meetings.

I must also thank my wife and true sole mate Sabina Khan, (now Khan-Ibarra), who I met and married during my 3rd year of grad school. Her understanding and support for me throughout graduate school has been essential to my health and well-being.

Special shout-outs to:

- All the Fischer lab members (current and past) - Juhyun Shin, Rie Uzawa, Denisse Rojas, Jenny Frost, Matt Bauer, Wenyan, Jon Penterman, Mary Gehring, Karandeep,
- The entire PMB Class of 2006
- The Zilberman lab
- Barb and all the folks at the green house

….And, above all…..I must thank my mother, Sarah, who always told me, "people can take everything away from you, but they can never take away your education!"…This is for you mom!

Chapter I

Introduction and Background

Megasporogenesis and Gametogenesis

At the heart of plant and animal reproduction is the formation of gametes which can partake in fertilization and develop into the next generation. For animals, the formation of gametes starts early in embryogenesis, as specific cells are selected and set aside, eventually to become the germ cells capable of undergoing meiosis and forming into the gametes (Feng et al., 2010b). Undifferentiated and diploid, these primordial germ cells (PGCs) will undergo gametogenesis to produce the male and female gametes. Plant gametogenesis is strikingly different than mammalian gametogenesis, as there is no setting aside of germ cells during development. Instead, plants make their gametes by selecting somatic cells at the appropriate time and place, and then subject them to two reductive processes: sporogenesis and gametogenesis**.** The products of sporogenesis are haploid spores, of which there is a female megaspore and a male microspore. The megaspore undergoes megagametogenesis producing the female gametophyte which houses the female gamete (egg cell), while the microspore undergoes microgametogenesis resulting in the male gametophyte which holds the male gametes (two sperm cells). The plant life cycle is described as an alternation of generations, moving between the multicellular gametophyte generation and the multicellular sporophyte generation (Leyser and Day, 2003).

The megasporogenesis in plants occurs at structural outgrowths called ovule primorida (Drews, 2011). The cells of the primordia are somatic and genetically diploid, but as these finger-like projects begin to arise, one of the primordia cells undergoes a differentiation process and forms the megaspore mother cell. Although genetically identical to the surrounding somatic cells from which it originated, the megaspore mother cell is considerably larger, having a denser cytoplasm and larger nucleus (Drews, 2011). Eventually, the megaspore mother cell undergoes meiosis producing four daughter haploid megaspores, of which three undergo a process of programmed cell death and degenerate. The surviving cell is called the megaspore and marks the beginning of megagametogenesis (Fig. 1.1, A).

Embedded in the young ovule and protected by the integuments, the haploid megaspore develops into the multicellular female gametophyte. Megagametogenesis begins with the megaspore undergoing two mitotic events without cytokinesis. The outcome of this is a coenocyte comprised of four nuclei, two of each located at the polar ends of the cell. A third mitosis occurs making a total of eight nuclei, of which four are now located at each pole of the cell (now called the embryo sac). During this last mitosis a cellularization process happens in which phragmoplasts and cell walls form around the nuclei, demarcating their locations within the developing female gametophyte. Coinciding with cellularization is the migration of two nuclei from opposite poles to the center of the female gametophyte, where they meet and then fuse to form the diploid central cell (Huh et al., 2008). At the end of cellularization the seven-celled female gametophyte is comprised of two synergid cells, three antipodal cells, the

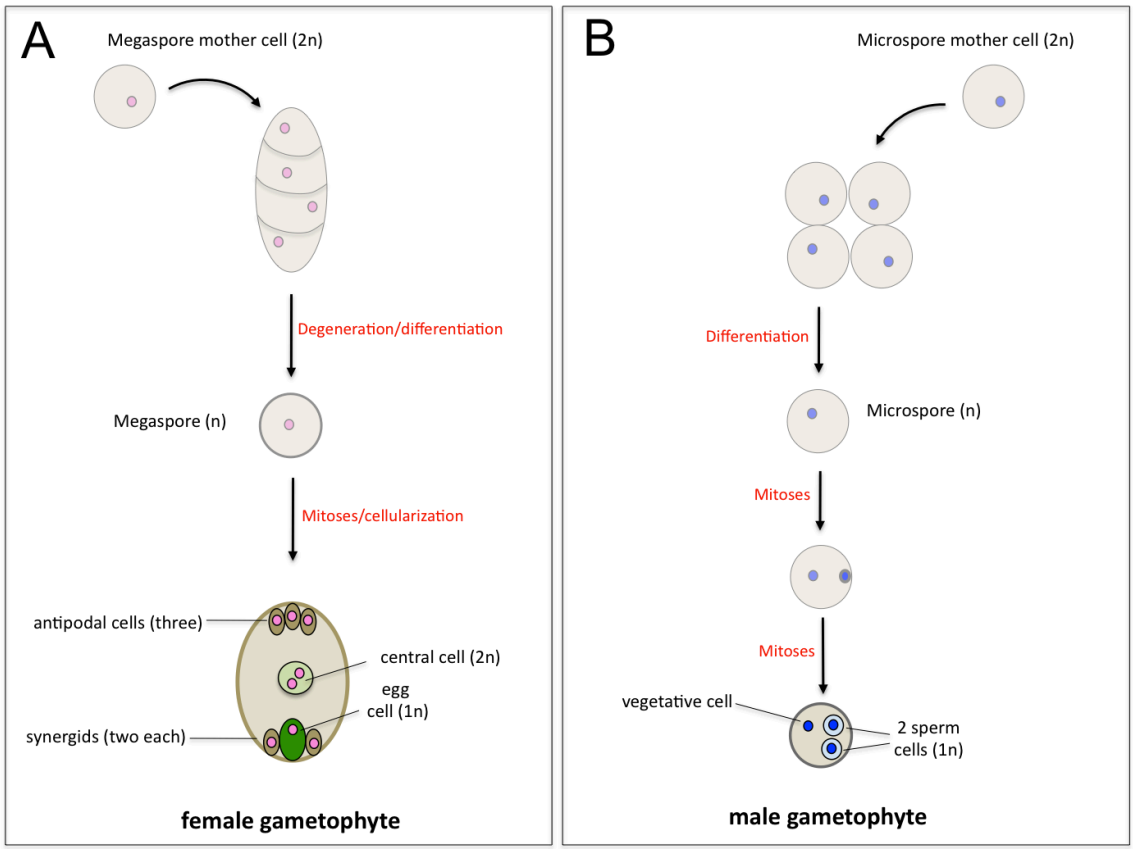


Fig. 1. 1 Megasporogenesis and microsporogenesis in *Arabidopsis.* Small pink and blue circles represent maternal and paternal nuclei respectively.

central cell, and the egg cell. All of the cells are genetically haploid, except the central cell which is diploid (Fig. 1.1, A).

Microsporogenesis and Microgametogenesis

The male counterpart to megasporogenesis, is microsporogenesis (Fig. 1.1, B) which occurs in the anthers, the male reproductive structures of the plant. Early in development, a somatic cell in the developing anther is selected to become the diploid microspore mother cell, which undergoes meiosis resulting in four haploid daughter cells. Unlike megasporogenesis, where three of the haploid daughter cells degenerate via programmed cell death, during microsprogenesis all four of the resulting daughter cells survive and form a tetrad of haploid microspores. Microgametogenesis is initiated when a single haploid microspore is released from the tetrad and commences to grow by cell expansion and the formation of a large vacuole (Twell, 2011). At this time the nucleus migrates to the microspore wall where the mitotic spindle begins to appear. The spindle is asymmetric in terms of its location being on one side of the cell, rather than being in the middle. This asymmetric placement of the spindle has a significant consequence, as its corresponding mitosis gives way to a highly asymmetric nuclear division, resulting in one of the daughter cells being significantly smaller than the other (Twell, 2011). The smaller cell is the germ cell and the larger cell the vegetative cell. Because no phragmoplast or cell walls are formed after mitosis, the germ cell is essentially engulfed by the large vegetative cell. A final mitosis occurs in the germ cell resulting in two daughter germ cells which continue to develop into the sperm cells. Thus, the mature tricellular pollen granule, the male gametophyte, consists of two sperm cells engulfed by a surrounding vegetative cell (Fig. 1.1, B).

Double Fertilization Produces the Embryo and Endosperm

Both plants and mammals carry out sexual reproduction in which their haploid gametes undergo a fertilization event resulting in a diploid embryo. In plants however, a modification of this process has evolved where two fertilization events occur simultaneously within the female gametophyte (Fig. 1.2). This type of fertilization is called, double fertilization, and it appears to be unique to flowering plants (Leyser and Day, 2003). Double fertilization is initiated when the male microspore (pollen granule) lands on specialized cells called stigmatic papillar cells which reside on the tip of the flower carpel. The microspore is in a desiccated state, but upon adherence to the papillar cell, a reaction takes place causing the microspore to rehydrate. The rehydration event puts the microspore into a physiologically active state causing the vegetative cell to germinate and generate a pollen tube (Palanivelu and Tsukamoto, 2012). The growing pollen tube penetrates the papillar cell, and continues to traverse down the inside style

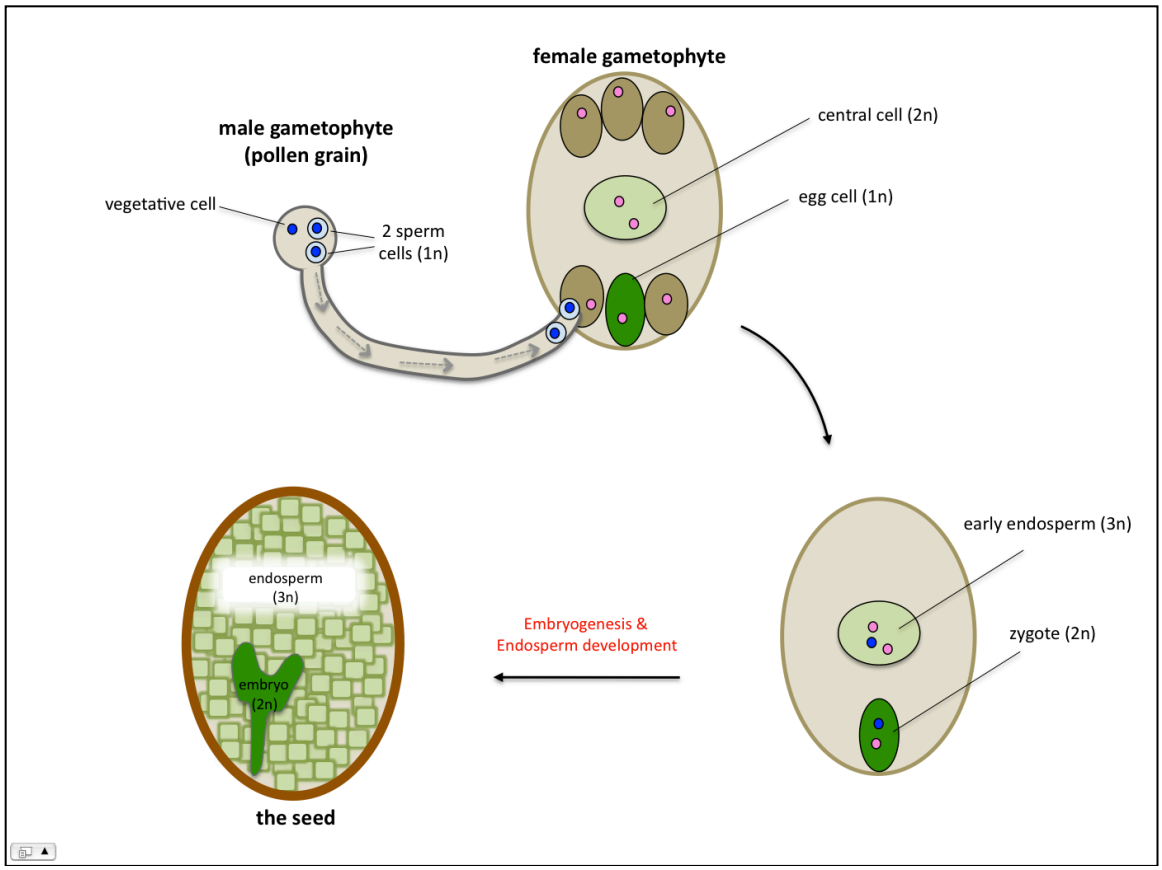


Fig. 1. 2. Double fertilization in *Arabidopsis.* Small pink and blue circles represent maternal and paternal nuclei respectively.

of the flower until it reaches the ovule, where it then moves through the funicle, into the micropylar, and eventually penetrating one of the synergids of the female gametophyte (Iwano et al., 2004). Upon entry into the synergid, the pollen tube ceases to grow and then bursts open, releasing its two sperms into the synergid cytoplasm (Palanivelu and Tsukamoto, 2012). The two sperm cells traverse the synergid cytoplasm and eventually migrate to the haploid egg cell and diploid central cell respectively, thus fertilizing both. Hence, the product of double fertilization is a diploid embryo and triploid endosperm (Fig. 1.2).

Embryogenesis

The initial stages of *Arabidopsis* embryogenesis involve setting up the future body plan of the multicellular adult plant, which is composed of a polar apical-basal axis surrounded by a radial pattern of concentric tissues (De Smet et al., 2010). Soon after fertilization, the zygote elongates along the future apical-basal axis, until an asymmetric division occurs resulting in a small apical daughter cell and a much larger basal daughter cell (De Smet et al., 2010). The basal daughter cell proliferates develops into the suspensor, a structure that connects the embryo to the surrounding maternal tissue and endosperm, functioning as a nutrient conduit for the embryo (Kawashima and Goldberg, 2010). The apical daughter cell undergoes a series of highly coordinated cell divisions and eventually develops into the embryo proper, which contains the meristimatic tissues essential for continual plant development (Fig. 1.2).

Endosperm Development and Function

In parallel to *Arabidopsis* embryogenesis is the development of the endosperm. Upon fertilization of the diploid central cell by the haploid sperm, the nucleus of the nascent triploid endosperm begins a process of rapid nuclear division without cell wall formation, which results in a large coenocyte containing ~200 nuclei (Berger et al., 2007; Olsen, 2004). Soon after coenocyte formation a celluarization process begins with each nuclei emitting a radial microtubule system (RMS) from its nuclear envelope. The outer most ends of the RMS forms the boundary between each nuclei, providing a place for phragmoplast vesicles to deposit cell wall components for eventual cell wall formation (Olsen, 2004). The developing endosperm is comprised of several different types of tissues (Berger, 1999; Lopes and Larkins, 1993). Cytological studies reveal three distinct domains, each exhibiting the expression of specific genes (Berger et al., 2007). Surrounding the embryo is the (1) micropylar endosperm (MCE) consisting of small extremely metabolically active cells (Berger, 1999). The cellular activity and juxtaposition of the MCE to the embryo suggests MCE may direct nutrient transfer from endosperm to embryo (Berger, 1999). Opposite of the embryo and MCE is the (2) chalazal endosperm (CZE) consisting of a cluster

of syncytially dividing cells at the chalazal cyst (CZC), above the maternal chalazal proliferation tissue (CPT). The basal end of the CZC forms a haustoria with tentacle-like processes that penetrate into the CPT eventually reaching the vascular tissues of the maternal plant (Nguyen et al., 2000). The complex structure of the CZE and its proximity to the maternal tissue, suggests the CZE is critical in regulating maternal nutrient flow to the seed. The remainder of the seed's volume is filled with small thin cells making up the (3) central/peripheral endosperm (PEN) (Olsen, 2004).

The endosperm is a storehouse of proteins, starch, and lipids which function to supply the embryo with essential nutrients for development. Because of its location, the endosperm also serves to mediate the flow of information transfer between the maternal sporophytic plant and the embryo. In dicots, such as *Arabidopsis*, the endosperm is consumed during seed development by the embryo, while in monocots it persists in to the mature seed and is only consumed by the embryo during germination (Brown and Lemmon, 2007; Olsen, 2004). Persistent endosperm was a selective advantage for the monocots (i.e. cereals) allowing for better seed dispersal, long-term dormancy, and effective seedling establishment. Moreover, the cereal endosperm has been advantageous to humans, being the primary source of nutrition worldwide, providing at least 60% or more of our daily caloric intake.

DNA Methylation is a Repressive Epigenetic Mark in Plants and Animals

DNA methylation is a process naturally occurring in many eukaryotic genomes, involving the addition of a methyl functional group onto the 5-carbon base of a cytosine within the DNA molecule. The process is carried out by enzymes known as DNA methyltransferases, capable of transferring the methyl group from the well known methyl donor compound S-Adenosyl methionine (SAM). The cytosines selected for methylation may occur in different 'contexts,' which varies between organisms and cell types. In plants, DNA methylation can occur in three contexts, which are CG, CHG, and CHH (where H=A,C, or T) (Cokus et al., 2008). However, in mammals, DNA methylation has been shown to occur almost exclusively in the CG context, although a small amount of CHG and CHH methylation have recently been found in embryonic stem cells (Lister et al., 2009; Ramsahoye et al., 2000). DNA methylation appears to be a critical aspect to both plant and mammalian development, since null-mutations in some of the DNA methyltransferases lead to physical abnormalities and infertility in plants, and embryonic abortion in mammals (Okano et al., 1999).

A methylated cytosine is an epigenetic 'mark' on the genome alerting the cell to silence and transcriptionally inactivate that denoted area of the genome. A few models exist attempting to explain the mechanism of repression due to DNA methyation. One model suggests transcriptional silencing can be caused by the methylated cytosine physically blocking DNA binding proteins, such as transcription factors, from properly associating with the DNA (Watt and Molloy, 1988). Another model suggests the methylated cytosine behaves like a

recruiting agent, recognized by proteins possessing methyl-binding domains (MBDs) which bind to the methylated site (Klose and Bird, 2006). These MBD proteins recruit additional proteins, including chromatin remodeling proteins, which change the local chromatin architecture, forcing it into a tight repressive heterochromatic state (Cedar and Bergman, 2012).

On an evolutionary level, both plants and mammals have found several ways to take advantage of the repressive quality of DNA methylation. It seems that one common use of DNA methylation is that of a defense system for the genome, working to silence transposons, repetitive elements, and any type of deleterious mobile element (Feng et al., 2010a; Zemach et al., 2010b). This is especially the case in *Arabidopsis*, where a significant amount of the methylation in the genome appears to be specifically delegated to transposable elements (Zemach and Zilberman, 2010). DNA methylation leads to silencing, which prevents transposable element movement. A transposon with aberrant expression, if next to an endogenous gene in the genome, can affect the normal expression of that gene, causing the gene to abnormally expressed as well (Kato et al., 2003). Hence, DNA methylation's function as a defense mechanism ensures the overall integrity of the genome.

Genomic imprinting is another function of DNA methylation in which its ability to silence gene transcription is utilized. While most genes in the genome are expressed bialleleically, a small amount (less than 1% in both plants and mammals) exhibit monoallelic expression. The silencing of the non-expressed allele is due to DNA methylation and additional epigenetic mechanisms such as histone modifications and chromatin remodeling (Barlow, 2011).

A firm explanation of the evolutionary reason for genomic imprinting has been a subject of debate for sometime, but the most widely accepted model suggests imprinting arose out of a conflict between parents regarding how much of the mother's nutrients should go into supporting the offspring (Haig, 2000; Haig and Westoby, 1989). For organisms whose reproductive strategy allows the mother to be impregnated multiple times by different males over the course of her life, a natural conflict will arise between the mother and the father in regards to how much of the mother's resources should be given to the offspring. Under these conditions of multiple paternity, the father would favor abundant maternal resource allocation to the offspring of the specific brood he fathered, to ensure the most optimal conditions for their survival. There would be no evolutionary drive for him to support the growth of the offspring from other broods, the ones he did not father. In the case of the mother, however, she would favor a more equal distribution of her resources to all of her offspring, since they all are indeed her offspring. As a consequence of this theory, it is suggested that paternally expressed imprinted genes (PEGs) tend to promote resource allocation, while maternally expressed imprinted genes (MEGs) are more likely to suppress it (Haig, 2000). Moreover, imprinted genes tend to be expressed in tissues or structures which function to mediate the supply of nutrients from the mother to the offspring. In mammals the structure responsible for this is the placenta and its analogue in plants is the endosperm.

The Establishment of DNA Methylation

For a particular cytosine in the genome targeted for DNA methylation, the methylation at that site must initially be established *de novo*, and then subsequently maintained as the cell mitotically divides. During mammalian development, two major genomic reprogramming events take place where epigenetic marks including DNA methylation and histone modifications, are stripped away and subsequently reestablished (Feng et al., 2010b; Hayashi and Surani, 2009). In both reprogramming events, the *de novo* methylation is preceded by genome-wide active DNA demethylation . The first wave of *de novo* methylation occurs in the somatic cells of the nascent embryo, as it implants itself into the uterine wall, and is carried out by the DNA methyltransferases DNMT3A and DNMT3B (Hayashi and Surani, 2009). The demethylation followed by methylation essentially functions to set the embryonic cells to a more pluripotent state, consequently preparing them for upcoming development and growth (Hayashi and Surani, 2009). The second wave of *de novo* methylation happens during gametogenesis in PGCs and involves the DNA methyltransferaes, DNMT3A and DNMT3-like (DNMT3L) establishing the proper methylation patterns reflective of the gender of the gonads they reside in. It is also at this time that the DNA methylation marks responsible for genomic imprinting, are established as well (Hemberger et al., 2009). Although DNMT3A and DNMT3L are methyltransferases, only DNMT3A can catalytically transfer the methyl group to the cytosine (DNMT3L lacks this ability) (Bourc'his et al., 2001). However, DNMT3L has demonstrated preferential binding to unmethylated H3K4 tails (Ooi et al., 2007). Thus, a current model describing the targeting of specific loci for *de novo* methylation suggests DNMT3L initially binds to unmethylated H3K4 tails following recruitment of DNMT3A for active DNA methylation (Jia et al., 2007; Ooi et al., 2007).

In plants *de novo* methylation occurs via an elaborate process known as the RNA-directed DNA Methylation (RdDM) pathway (Henderson and Jacobsen, 2007; Matzke et al., 2009; Wassenegger et al., 1994). As shown in Figure 1.3, RdDM is initiated by a plant-specific DNA-dependent RNA polymerase, RNA-POLYMERASE IV (POLIV), which is believed to target repetitive and transposable elements in the genome for transcription (Matzke et al., 2007). The single-stranded RNA-transcripts generated by POLIV activity become the template for RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), which generates double-stranded RNA (dsRNA). The formation of dsRNA activates the endoribonuclease DICER-LIKE-3 (DCL3), which processes the dsRNA into small-interfering RNAs (siRNAs) of 24 nucleotides (nt) in length. These siRNAs are further processed by the SAM-dependent RNA methyltransferase HUA ENHANCER 1 (HEN1) which brings the siRNA into maturity by methylating the 2'OH group on the 3'-terminal nucleotides of siRNAs (Yang et al.; Yu et al., 2005).

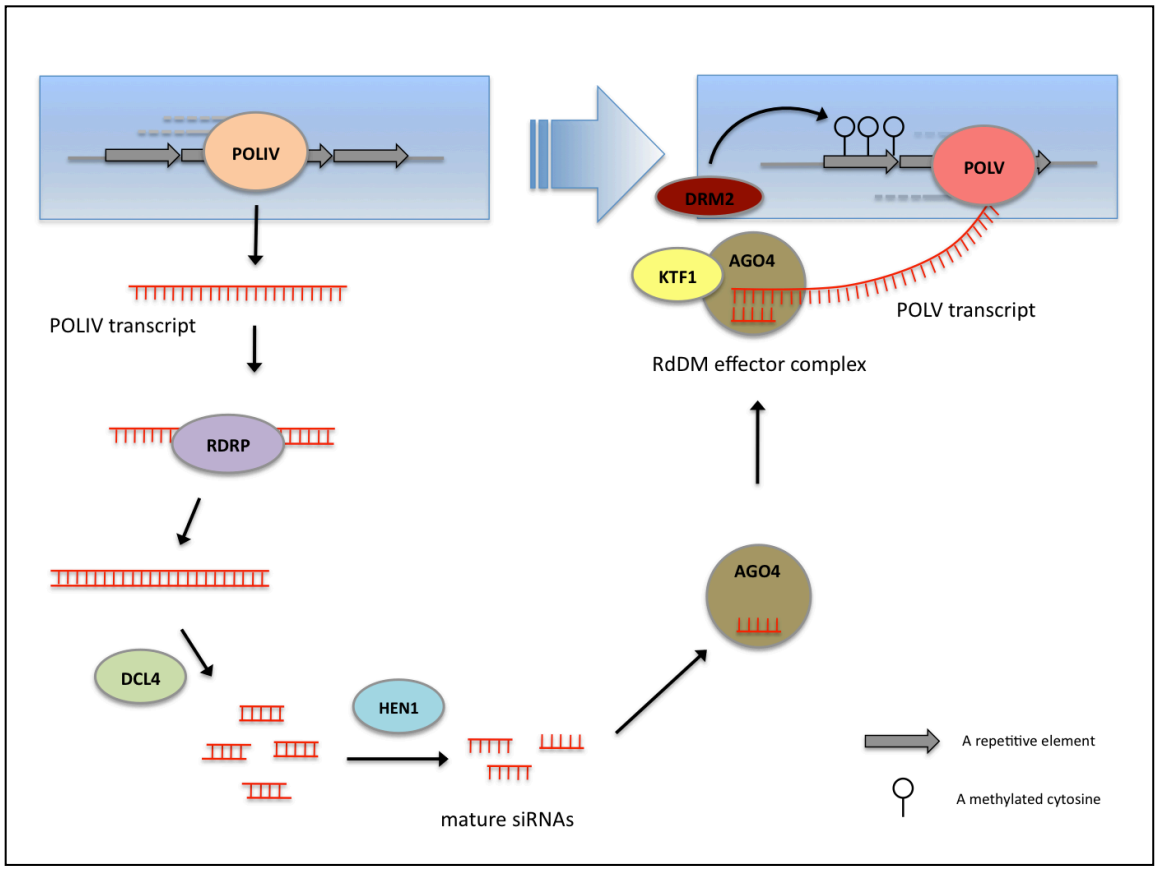


Fig. 1. 3. RNA directed DNA methylation. The large grey arrows represent transposons and/or repetitive elements.

The mature siRNAs are then loaded onto ARGONAUTE 4 (AGO4) - a member of the Argonaute family of small RNA (sRNA) - binding proteins. Argonaute (AGO) proteins are highly conserved across eukaryotes and associate with different classes of sRNAs, such as micro RNAs (miRNAs), siRNAs, and the animal related piwi-interacting RNAs (piRNAs) (He et al., 2011b). Most AGOs assemble into the RNA-induced Silencing Complex (RISC), which functions to carryout posttranscriptional gene silencing (Czech and Hannon, 2011). However in the case of RdDM, the siRNA-loaded AGO4 (siRNA/AGO) will further recruit an additional component, RNA POLYMERASE V (POLV). Like POLIV, POLV is also a plant-specific DNA-dependent RNA polymerase, but its suggested function is to generate single-stranded (ss) RNA transcripts from transposon or repetitive element.intergenic non-coding (IGN) regions (Wierzbicki et al., 2008). These ssRNA transcripts act as linkers bringing the siRNA/AGO complex to the targeted loci. The nascent POLV transcript is fastened to the siRNA/AGO complex through base-pair complementary between the POLV transcript and siRNA (He et al., 2009; Wierzbicki et al., 2009). The RNA binding protein KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1) reinforces the base-pair binding between the POLV transcript and siRNA by simultaneously binding to the POLV transcript and AGO4, resulting in the functional RdDM effector complex. Formation of the mature RdDM effector complex signifies the recruitment of the signature RdDM *de novo* methyltransferase, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), an ortholog of DNMT3A, which then carries out *de novo* DNA methylation at the targeted site (He et al., 2009; Zheng et al., 2007; Zilberman et al., 2003).

Mechanisms for Maintaining CG Methylation in Plants and Mammals

While *de novo* methylation is used to establish methylation, maintaining the methylation mark at a selected site from one cell division to the next is carried out by different processes. A methylated CG will always have its complementary methylated CG on the reverse strand. An apparent outcome of this methylation symmetry can be seen after mitosis, where semiconservative replication results in two nascent hemimethylated daughter DNA stands. The hemimethylated DNA must be restored to a fully methylated state in the newly made cell, since ensuing replications of hemimethylated DNA would eventually remove the methylation mark. This type of DNA demethylation is known as *passive* DNA demethylation and occurs when a methylated cytosine loses its DNA methylation mark due to a failure in maintenance methylation.

In mammals, symmetrical CG methylation is maintained from one generation to the next by the maintenance DNA methylatransferase DNMT1, which is brought to the replication fork during the time of DNA synthesis by the chromatin-associated-protein Ubiquitin-like Plant Homeodomain and RING Finger Domain-1 (UHRF1), which contains a SET or RING ASSOCIATED domain (SRA) (Bostick et al., 2007; Sharif et al., 2007). SRA domains are known to bind hemimethylated DNA and it is believed that UHRF1 recruits DNMT1 to

the replication fork as it binds to the nascent hemimethylated CG sites formed immediately after replication (Bostick et al., 2007). The UHRF1 recruitment of DNMT1 to the replication fork restores the two hemimethylated daughter DNA strands to fully methylated states. The same pathway may occur in plants where the homologue to UHRF1, VIM1, binds to hemimehtylated CG sites via its SRA domain. MET1, the plant DNMT1 homologue, is then recruited by VIM1 in order to restore DNA methylation (Woo et al., 2008; Woo et al., 2007).

Mechanisms for Maintaining non-CG Methylation in Plants

In addition to CG methylation, plants also have CHG and CHH methylation, which need to be maintained throughout mitotic replication as well. The maintenance of CHG methylation occurs by interactions between the CHROMOMETHYLASE 3 (CMT3) methyltransferase, and a histone 3 lysine 9 dimethylation (H3K9me2) mark (Bartee et al., 2001; Lindroth et al., 2001). A methyltranseferase unique to the plant kingdom, CMT3 has a chromodomain motif which binds to H3K9me2 modifications functioning to guide CMT3 to the chromatin. The chromodomain is also believed to be involved in remodeling the targeted heterochromatin (Eissenberg, 2001). The H3K9me2 modifications are mainly carried out by the SET and SRA domain containing histone methyltransferase, SUPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUV4). The model suggests CHG methylation in plants is maintained through a reinforcing loop (Law and Jacobsen, 2010) where SUV4 binds to methylated CHG sites via its SRA domain and then carries out H3K9me2 on the histone. The H3K9me2 mark in turn recruits CMT3 to the loci via its chromodomain, enabling CMT3 to methylate the CHG site.

The maintaining of CHH methylation is done through persistent *de novo* methylation via the RdDM pathway of which DRM2 is the main DNA methyltransferase (Chan et al., 2005). The requirement for constant *de novo* methylation to maintain CHH methylation is a direct result of the asymmetrical nature of CHH methylation, and the immediate loss of methylation in the newly synthesized DNAs.

Active DNA Demethylation Occurs in Plants and Animals

In addition to harboring mechanisms that establish and maintain DNA methylation, eukaryotes have also evolved methods for removing the repressive methylation mark. As previously mentioned, *passive* DNA demethylation occurs when maintenance methylation fails to target a cytosine for methylation while DNA replication and cell division continue. These conditions effectively titer-out the methylation mark at that particular cytosine, resulting in the entire population of cells no longer harboring the methylation mark. The process of *active* DNA methylation has also been demonstrated in organisms as well, and it involves the

enzymatic removal of DNA methylation, specifically 5mC (He et al., 2011a). A profound consequence to the active removal of the repressive methylation mark is the derepression of transcription. Indeed, active DNA demethylation is antagonistic to DNA methylation resulting in gene expression (Ooi and Bestor, 2008). Over the last decade, a considerable amount of work has gone into elucidating the mechanisms of active DNA demethylation in both plants and animals.

Active DNA Demethylation in Plants

In *Arabidopsis*, active DNA demethylation is carried out by the DEMETER (DME) family of DNA glycosylases and components of the base excision repair (BER) pathway (He et al., 2011a). DNA glycosylases function to remove damaged or mis-paired bases from genomic DNA, and are the first enzymes in the BER pathway (Jacobs and Schar, 2012). A damaged base is initially recognized by a DNA glycosylase, which acts to flip the base out from the double helix and hydrolyze its N-glycosidic bond (ribose-base), resulting in a complete removal of the base from the double helix. This produces an apurinic/apyrimidinic site (A/P site), which is further processed by an AP lyase which nicks the sugar-phosphate backbone. The single strand break in the sugar-phosphate backbone activates an AP endonuclease which creates a 3' –hydroxyl at the region, allowing a DNA polymerase to add the correct base. A DNA ligase is then used to seal the nick, thusly completing the BER pathway.

Unlike other DNA glycosylases, the DME family of DNA glycosylases function to remove 5mC, instead of a damaged or misspaired base (Choi et al., 2002; Gehring et al., 2006). In essence, the methylated cytosine is flipped out of the double helix and its N-glycosidic bond hydrolyzed by the DME DNA glycosylase enzyme, resulting in the expected A/P site and a free methylated base (5mC). DME itself further nicks the sugar-phosphate backbone at the A/P site, as a consequence of being a bifunctional helix-hairpin-helix DNA glycosylase, which have both DNA glycosylase and apurinic/apyrimidinic (AP) lyase activity (Gehring et al., 2006). At this point, downstream components of the BER pathway are activated (i.e. AP endonuclease, DNA polymerase, and DNA ligase), resulting in replacement of 5mC with unmethylated cytosine.

The DME family of DNA glycosylases consisting of DEMETER (DME), REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2), and DEMETER-LIKE 3 (DML3) have different functions in *Arabidopsis*. ROS1, DML2, and DML3 are expressed broadly throughout the plant, mainly in adult vegetative tissue. A genome-wide methylation analysis of whole plant tissue possessing a triple mutation in *ros1 dme2 dml3*, found 179 hypermethylated loci compared to the wild type tissue, suggesting the hypermethylated loci to be targeted by these three DNA glycosylases in the wild type (Penterman et al., 2007). Most of these demethylated loci appeared at the 5' and 3' ends of genes, a pattern opposite of DNA methylation. A model was proposed suggesting the three DNA glycosylases worked to 'prune' away DNA methylation from the ends

of genes. In essence, this 5' and 3' gene adjacent demethylation was acting to remove any misdirected DNA methylation that may have made its way into the transcriptionally sensitive areas of genes (Penterman et al., 2007; Zhu et al., 2007).

Unlike ROS1, DML2, and DML3, which are broadly expressed throughout the plant, DME has been shown to be active in reproductive tissues, specifically the companion cells, of the developing plant. DME-mediated DNA demethylation occurs in the central cell of the female gametophyte prior to fertilization, resulting in the maternal expression of several genes (Huh et al., 2008). Indeed, DME's activity in the central cell is essential for setting up the allele-specific methylation patterns that ultimately establish genomic imprinting in the *Arabidopsis* endosperm (Choi et al., 2002; Gehring et al., 2006). Moreover, central cell DME-mediated DNA demethylation is required for subsequent seed development, since those female gametophytes inheriting the maternal *dme*-mutant allele give a mutant phenotype of an over-proliferated endosperm. Thus, *dme*-mutant heterozygous plants will generated seeds in which half of the progeny will abort, while the other half will be normal (Choi et al., 2002).

DME is also expressed in the vegetative cell of the male gametophyte, since whole pollen shows DME expression, while isolated sperm cells shows none (Schoft et al., 2011). Like the *dme*-mutant female gametophyte, male pollen inheriting a *dme*-mutant allele also displays a mutant phenotype in the form of an inability to germinate properly. Locus-specific methylation analysis found the vegetative cell genome to have several loci targeted by DME-mediated DNA demethylation. Interestingly, many of these loci were the same as those in the central cell, suggesting DME may be targeting similar loci in both the central cell and vegetative cell for active DNA demethylation (Schoft et al., 2011). Chapter 3 of this thesis attempts to address this observation of DME targeting similar sites in central cell and vegetative cell.

Active DNA Demethylation in Mammals

In the last decade, multiple mechanistic pathways have been proposed, attempting to explain mammalian DNA demethylation (Bhutani et al., 2011). In mammals, no orthologs of the DME family of DNA glycosylases have been found (Gehring et al., 2009b; Ooi and Bestor, 2008). However, DNA glycosylases from the mammalian BER pathway appear to be used for active base removal, demonstrating a commonality between plant and mammalian active DNA demethylation.

Earlier work in Zebra fish and mouse PGCs, suggest a model where the methylated cytosine in the 5mC/G base pair, is first enzymatically converted to thymine (causing a T/G mismatch), which is then restored to a C/G base pair by thymine DNA glycosylases (Morgan et al., 2004; Rai et al., 2008). The enzymes that generate the T/G mismatch, activation induced deaminase (AID) and apolipoprotein B mRNA-editing catalytic polypeptides 1 (APOBEC1), belong to a family of 5mC-deaminases, which function to deaminate the amine (NH_3) group

of the methylated cytosine, causing a C to T base conversion. The potentially deleterious effects of the T/G mismatch activate the BER glycosylases, thymine DNA glycosylase 1 (TDG1) and methyl-binding domain 4 (MBD4) (an enzyme with DNA glycosylase activity), which remove the thymine, creating an abasic site. The additional downstream components of the BER pathway are then triggered by the abasic site, ultimately resulting in a C (unmethylated) being incorporated into the abasic site. An additional component of this process is Growth Arrest and DNA-damage-inducible protein 45α (Gadd45α), a multifunctional protein which acts as an adaptor between AID and MBD4 as well as a recruiting agent bringing BER factors to the demethylation site (Barreto et al., 2007; Niehrs and Schafer, 2012).

More recently, intense study has been given to the ten-eleven-translocation (TET) family of proteins, responsible for converting 5mC to 5'-hydroxymethylcytosine (5hmC) (Iyer et al., 2009; Tahiliani et al., 2009). In mouse embryonic stem cells (ESCs), TET1/2 and 3 carryout an Fe(II) and alpha-ketoglutarate (α-KG)–dependent oxidation reaction, resulting in the methyl-group on the 5-carbon getting hydroxylated and the formation of 5hmC (Ito et al., 2011). The 5hmC can then be further modified in two ways. In one case, the 5hmC undergoes deamination by AID/AB1,3, resulting in the formation of 5'-hydromethyluricil (5hmU) and a corresponding U/G mismatch. A uracil DNA glycosylase, single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1), then removes the incorrect base (the uracil), creating the abasic site, which is further corrected via BER (Bhutani et al., 2011). However, it has also been suggested that the TET family may further oxidize 5hmC to 5-formylcytosine (5fC) and then to 5-carboxylcytosine (5caC). Both 5fC and 5CaC are believed to then be targeted by the thymine DNA glycosylases for active removal of the base (Ito et al., 2011).

DME-Mediated DNA Demethylation in the Central Cell Establishes Genomic Imprinting in the *Arabidopsis* Endosperm

The molecular processes responsible for genomic imprinting are different between plants and mammals. One possible reason for this difference is that mammals have a *bona fide* germ cell line, while plants generate germ cells in flowers produced late in development. Early in mammalian embryogenesis, diploid pluripotent primitive ectoderm (PEct) cells in the inner cell mass (ICM), initiate a process of differentiation towards a somatic linage, as these cells will eventually spawn all the differentiated cells constituting the adult organism (McLaren and Lawson, 2005). During this time however, a small subpopulation of these cells are blocked from undergoing this somatic fate process. These are the PCGs, and their germ cell-specific fate is initiated by the expression of *Blimp1*, a master regulator which functions to suppress the somatic cell fate program (Ohinata et al., 2005) and activate the more pluripotent state, characteristic of germ cells. The function of PGCs is to undergo gametogenesis and make haploid gametes, that consequently have the proper sex-specific

imprinting patterns reflective of the gonads they reside in (i.e. ovaries or testes). Nascent PGCs must have the maternal and paternal DNA methylation marks they inherited initially removed, before they can reestablish new DNA methylation marks on their gametes. In mammals, genomic imprinting patterns are setup exclusively by targeted *de novo* DNA methylation (Hayashi and Surani, 2009; Surani, 2001).

Plant genomic imprinting appears to have evolved independently from animal imprinting, as it is the active removal of the repressive methylation mark that establishes the imprinting mark instead of targeted *de novo* methylation (Huh et al., 2008; Kohler et al., 2012). In *Arabidopsis* this demethylation activity is specifically carried out by the DME DNA glycosylase, which is expressed in the central cell of the female gametophyte. In plants, no germ cell line exists from which the gametes arise, and therefore no resetting of DNA methylation takes place involving the removal and reestablishment of epigenetic marks (Bourc,Äôhis and Voinnet, 2010; Feng et al., 2010b). Instead, the DNA methylation patterns are stably maintained, as the megaspore mother cell (which is the functional equivalent to the mammalian germ cell) undergoes gametogenesis to produce the central cell and egg cell gametes. DME is then expressed in the central cell, where it demethylates its targeted methylated loci (at the 5' and 3' ends of genes) on the maternal genome, resulting in transcriptional derepression of the maternal allele (Gehring et al., 2006). Because DME is not expressed in the sperm cell, the paternal alleles of the sperm cell remain methylated. Hence, the occurrence of fertilization results in an endosperm with a hypomethylated maternal allele compared to a hypermethylated paternal allele (Fig. 1.4).

An additional difference between mammalian and plant genomic imprinting is the diversity of tissue types in which imprinted genes are found. Imprinting in mammals seems to occur in many different tissue types although a large amount of imprinting is found in the placenta (Wang et al., 2011). In the last decade, mammalian imprinted genes have been found in the brain (Gregg et al., 2010), embryo (Weaver et al., 2009), embryonic stem cells (ESCs) (Frost et al., 2011; Kim et al., 2007), and blood (Frost et al., 2010). This magnitude of tissue variety is impressive when compared to plant imprinting, which is almost completely restricted to the endosperm (Jahnke and Scholten, 2009). It is interesting that plant imprinting evolved to occur in a terminal tissue, incapable of transmitting its genetic material to the next generation. Indeed, one possible reason why plant imprinting lacks the large-scale epigenetic removal and reestablishment events found in mammals is that there is simply no need for such events, since the very tissue which the events take place will terminate (Bauer and Fischer, 2011).

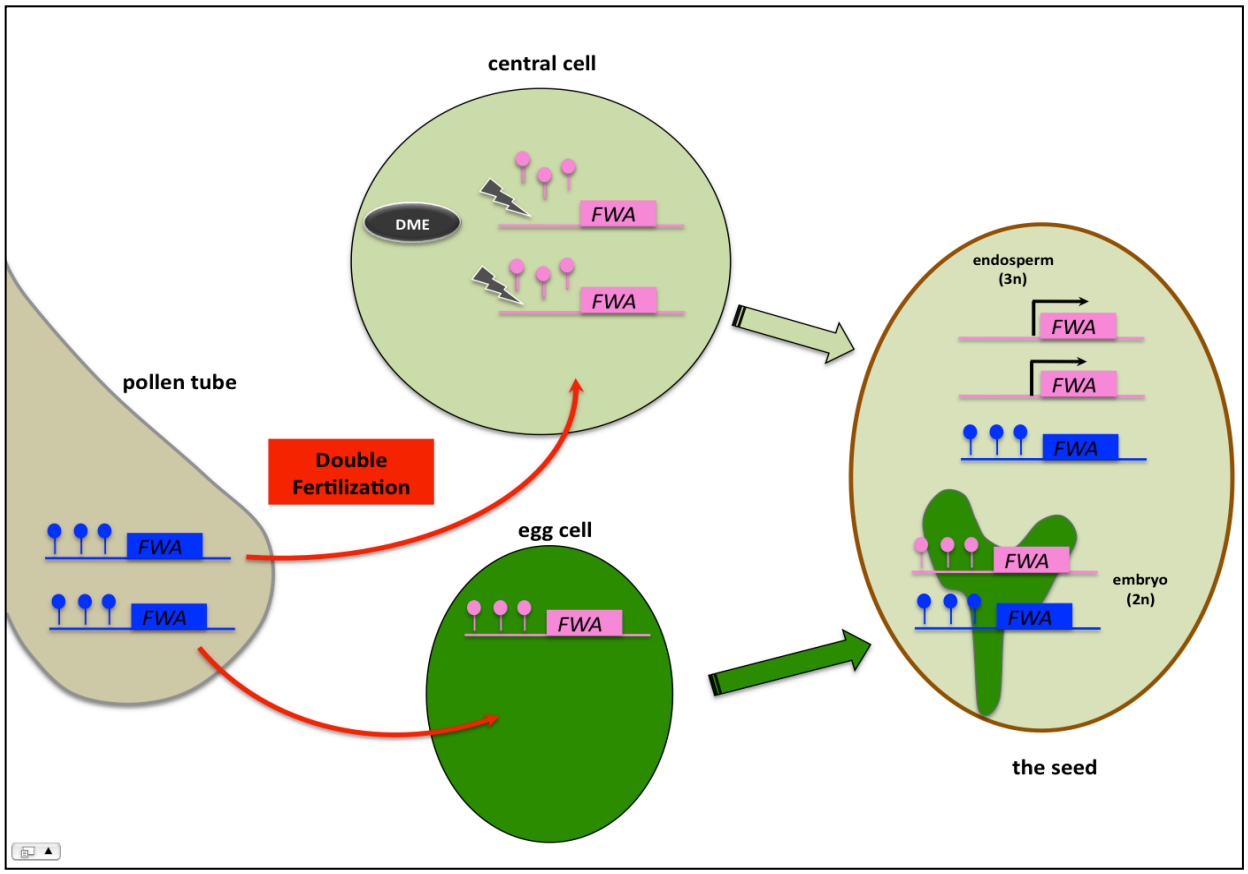


Fig. 1. 4. The classic model of *Arabidopsis* genomic imprinting.

DME-mediated DNA demethylation is not single-handedly responsible for plant imprinting, but rather works synergistically with additional epigenetic marks, an essential one being histone modification. The histone tails that emanate from the core nucleosome of chromatin are subject to numerous chemical modifications including methylation, demethylation, acetylation, deacetylation, phosphorylation, ubiquitinylation, sumolyation, and ADP-ribosylation (Allis et al., 2007). These specific modifications cause structural changes to the chromatin architecture and for the genes residing within them, leading to transcriptional activation or repression. Polycomb group (PcG) proteins are well-studied multiprotein complexes known for carrying out specific histone modifications, namely a trimethylation (me3) of the 27th lysine residue on the tail of histone H3 (H3K27) which causes transcriptional repression (Pirrotta, 1998; Schuettengruber et al., 2007; Schwartz and Pirrotta, 2008). Originally discovered in *Drosophila*, PcG proteins were found essential to animal development by functioning as repressors of the homeotic *Hox* genes, which are responsible for laying down the basic body plan and orientation of the developing fruit fly (Franke et al., 1992; Pirrotta, 1998).

In *Arabidopsis*, the Polycomb Group Repressive Complex-2 (PRC2), functions as a major regulator of endosperm development and imprinting. Homologous to *Drosophila* PRC2, *Arabidopsis* PRC2 consists of the SET-domain protein MEDEA (MEA), C2H2 zinc-finger protein FERTILIZATION INDEPENDENT SEED 2 (FIS2), and two WD-40 proteins FERTILIZATION INDEPENDENT ENDOSPERM (FIE) and MULTICOPY SUPRESSOR OF IRA 1 (MSI1). The molecular function of the PcG complex appears well-conserved between plants and mammals as mutations in any of the four *Arabidopsis* PRC2 components cause dramatic aberrancies in endosperm development resulting in abnormal seed formation (Goodrich et al., 1997; Grossniklaus et al., 1998; Ohad et al., 1999). Remarkably, mutant PRC2 plants produce autonomous endosperm without ever undergoing fertilization. However, the PRC2-mutant seed is severely compromised as a consequence of failing to undergo the distinct phases of endosperm development (i.e. nuclear proliferation, celluarization, and maturation) and is marked by excessive nuclear proliferation, a loss of anterior/posterior polarity, and an aborted embryo (Ohad et al., 1999). PRC2-mutant phenotypes suggest the normal function of PRC2 is to repress central cell proliferation until fertilization. PRC2 also functions in plant imprinting by establishing repressive H3K27me3 marks on the maternal and paternal genomes of the endosperm (Huh et al., 2008).

During the course of this work, several large-scale genome wide studies were done resulting in a dramatic increase in our catalog of the number of *Arabidopsis* imprinted genes from ~10 to ~300 (Gehring et al., 2011; Hsieh et al., 2011; McKeown et al., 2011; Tiwari et al., 2010; Wolff et al., 2011). Although many imprinted genes are known, only a few models exist that attempt to describe the molecular mechanisms of plant genomic imprinting. These models were made by extrapolating data from mutant analyses of the most well studied epigenetic regulators of imprinting - MET1, DME, and the PRC2 complex, and suggest the imprinting status of many of the currently discovered MEGs and

PEGs to be determined by at least one or more of these regulators (Hsieh et al., 2011).

For example, in the case of the maternally expressed *MEA* gene, both DME-mediated DNA demethylation and PRC2 repression are required to establish and maintain imprinting (Gehring et al., 2006). DME-mediated DNA demethylation of the maternal *MEA* allele in the central cell results in the expression of the maternal *MEA* allele, and the formation of a maternal PRC2. Fertilization then brings in the paternal methylated *MEA* allele, which is not subject to DME-mediated DNA demethylation, since DME expression ceases soon after endosperm formation (Gehring et al., 2006). Maternal PRC2 binds to the methylated paternal *MEA* allele causing it to be silenced, resulting in stable *MEA* imprinting.

DME-mediated DNA demethylation and PRC2 binding are also required for the imprinting of the MADS-box transcription factor *PHERES1*, but cause the silencing of the maternal *PHE1* allele and expression of the paternal *PHE1* allele instead (Kohler et al., 2003). In this particular case, DME-mediated DNA demethylation targets maternal *PHE1* in the central cell, resulting in maternal *PHE1* hypomethylation. Maternal PRC2 binds to the hypomethylated site repressing maternal *PHE1* expression. Fertilization then brings in the paternal methylated *PHE1* allele, but PRC2 is unable to bind properly potentially because of the DNA methylation marks (Kohler et al., 2003). The binding of PRC2 to the maternal allele of *PHE1*, and not to the paternal allele, results in *PHE1* paternal expression.

Unlike *MEA* and *PHE1*, which require both DME and PRC2 for imprinting, the homeo-box containing gene *FWA* seems to only require the activity of DME-mediated DNA methylation in the central cell for setting up its maternally expressed imprinting status (Kinoshita et al., 2004).

An underlying goal behind a large portion of the work done in this thesis was to identify processes involved in *Arabidopsis* genomic imprinting. The approach used to ascertain this potential information on imprinting was that of a high-throughput genome-wide DNA methylation analysis (MethylSeq) (Lister et al., 2008) of the endosperm, as well as the embryo. The findings in both Chapters 2 and 3 of this thesis, along with those of Gehring et al. (Gehring et al., 2009a) have resulted in a new understanding of plant genomic imprinting. Excitingly, the work done in this thesis has gained considerable attention in the field plant reproductive epigenetics, being a subject of interest in several recent reviews (Bauer and Fischer, 2011; Bourc‚Äôhis and Voinnet, 2010; Feng et al., 2010b; Ikeda, 2012; Kohler et al., 2012; Law and Jacobsen, 2010).

Epigenetic Events in the Companion Cells Ensure Gamete Genome Integrity

The ability of an organism to produce viable offspring is critical to the success of a particular species. It is not surprising then, that mechanisms have evolved which function to protect and support the integrity and overall well-being

of the germ line and gametes. One good example of this is nurse cells in *Drosophila* which provide nutrient support to the growing oocytes by manufacturing and supplying them with macromolecules such as ribosomes, mRNAs, and proteins (Lan et al., 2010). Accessory cells such as these tend to be in proximity to the germ cells making it easy for them to contribute their cytoplasmic contents into the developing gametic cells (Bastock and Strutt, 2007).

In plants, the central cell and the vegetative cell are supportive accessory cells of the egg cell and the sperm cells, respectively (Drews, 2011). The central cell expresses its supportive role to the gametes, soon after fertilization as it develops into the endosperm which functions to provide the growing embryo with nutrients. The vegetative cell supports the two sperm cells by germinating a pollen tube that physically transports the two gametes to the female gametophyte for fertilization.

Interestingly, specific epigenetic mechanisms appear to have also evolved in the companion cells in order to support the neighboring gametes. During the course of this thesis, Slotkin *et al*. (Slotkin et al., 2009) suggested genome-wide reprogramming events occurred in the vegetative cell, resulting in the de-repression of transposable elements and the biogenesis of siRNAs. These small RNAs could then be shuttled into the sperm cell in order to help reinforce the silencing of transposons. The notion that a tissue would intentionally reactivate transposons seems precarious, as it puts the integrity of its own genome in danger. However, in this particular case, the tissue is the vegetative cell which is degenerative, never passing on its genetic information to subsequent generations. Thus, the vegetative cell provides additional support to the gamete, by acting as a small RNA storehouse to ensure potentially deleterious transposable elements are not expressed in the gamete genome. In Chapter 2 of this thesis I explore the possibility of a similar scenario occurring between the companion central cell and the gamete egg cell of the female gametophyte.

What is the Role of DME-Mediated DNA Demethylation in the *Arabidopsis* Endosperm?

The goal of my thesis research was to investigate the role of DNA methylation and DME-mediated DNA demethylation in the reproductive tissues of *Arabidopsis*. In particular, I wanted to obtain, exhaustively, as much information as possible about the behavior of the DME DNA glycosylase, and the places it chooses to actively demethylate. This me (and my collaborators) to initially carry out a high-throughput DNA methylation analysis of the wild type endosperm and embryo and *dme*-mutant endosperm (Chapter 2) (Hsieh et al., 2009). A major finding from this study was that the endosperm genome was considerably reduced in DNA methylation compared to the embryo (Hsieh et al., 2009). Moreover, this reduction in DNA methylation could lead to transposon expression, and potentially the biogenesis of siRNAs for possible movement into the embryo. The data from this study also provided us with a new understanding

of plant imprinting. We then attempted to further examine the methylomes of the central cell and vegetative cell, which was done by using high-throughput DNA methylation analysis to elucidate methylation of maternal and paternal endosperm genomes, as well as vegetative and sperm genomes. A significant finding from Chapter 3 was that DME-mediated DNA demethylation targets the same loci in both the central cell and vegetative cell, and that these targeted loci are enriched in small transposons located next to genes.

Chapter II

Genome-Wide DNA Demethylation of the *Arabidopsis* Endosperm

INTRODUCTION

Genomic imprinting occurs in the *Arabidopsis* endosperm (Huh et al., 2008). Similar to mammalian imprinting, a critical component in *Arabidopsis* imprinting is an associated parent-of-origin specific DNA methylation, which in the case of some imprinted genes is the repressive mark responsible for the monoallelic expression of the imprinted gene. For other imprinted genes, the parent-of-origin specific methylation is still associated with the imprinting status, but other repressive protein complexes are responsible for the monoallelic expression (Arnaud and Feil, 2006; Gehring et al., 2006). In addition to having DNA methylation, *Arabidopsis* imprinting also involves the enzymatic removal of this repressive methylation mark, done by the plant-specific DNA glycosylase DME (Gehring et al., 2006). Moreover, DME is important for *Arabidopsis* reproduction and development, as a null-mutant maternal *dme* allele results in defective endosperm causing the embryo to abort (Choi et al., 2002). Thus, the epigenetic landscape of the endosperm is one in which DNA methylation and DNA demethylation are properly balanced.

To the best of our knowledge, all published studies examining DNA methylation in the *Arabidopsis* endosperm, prior to this work, were carried out using conventional techniques such as locus-specific bisulfite sequencing and various types of restriction enzyme-based methods (Gehring and Henikoff, 2007; Huh et al., 2008). A Genome-wide approach examining DNA methylation by immunoprecipitating methylated DNA (MeDIP) in non-seed tissues has also been applied (Zilberman et al., 2007; Zilberman and Henikoff, 2007). While these approaches have been useful to the field of plant DNA methylation, they nevertheless have inherent limitations. For example locus-specific bisulfite sequencing is a PCR-based method and is therefore limited to analyzing a single gene, while MeDIP exhibits a preferential bias towards precipitating highly methylated cytosines as opposed to those that are less methylated (Krueger et al., 2012). The net outcome of these aforementioned methods for DNA methylation analysis has resulted in only a partial glimpse of the *Arabidopsis* endosperm DNA methylation landscape.

Sodium-bisulfite conversion coupled to next-generation sequencing (Methyl-Seq) is a high-throughput DNA methylation method that has been applied to *Arabidopsis*. However, only the methylation profiles of easily obtainable tissues such as mature whole plants (Cokus et al., 2008; Zilberman et al., 2007) , roots (unpublished), leaves (unpublished), and flowers (Lister et al., 2008), and has been examined. Prior to this work, a genome-wide methylation study of the major tissues of the seed (i.e. embryo, endosperm, and seed coat) using a high-throughput approach had not been done. One practical reason for this has been an inability to obtain a substantial amount of specific seed tissue, particularly endosperm, in order to extract enough genomic DNA required for Methyl-Seq.

The aim of the work done in this chapter is to investigate on a genome-wide level, the role of DNA methylation and DNA demethylation in the *Arabidopsis* endosperm. A high-throughput approach is used, involving the

bisulfite conversion of genomic DNA coupled to next-generation sequencing using the illumina ™ platform. The analysis involves a comparison of the wild type endosperm and embryo methylomes, as well as that of a more developmentally mature tissue (adult aerial tissue). Endosperm lacking the functional DME DNA glycosylase (*dme*-endosperm) is also examined to determine the contribution of DME-mediated DNA demethylation to the endosperm methylation profile. The single-base resolution of this analysis has the advantage of revealing sequence context-specific methylation. The mechanisms responsible for maintaining CG and CHH methylation in the embryo and endosperm are explored.

RESULTS

Sequencing and DNA methylation analysis

To investigate the methylomes of the *Arabidopsis* seed, wild-type endosperm, wild-type embryos, and *dme*-endosperm were hand-dissected from mid-torpedo to early maturation stage seeds (Fig 2.1). Aerial tissue from 4 week-old plants was also obtained. DNA was then extracted from the plant tissues, followed by library construction, bisulfite conversion, and sequencing using the Illumina Genome Analyzer ™ (Hsieh et al., 2009; Lister et al., 2008). A custom work-flow was then used to carry out base-pair alignment and sequence analysis. We aligned approximately 2.5 billion bases for embryo, 2.2 billion bases for wild-type endosperm, 2.0 billion bases for *dme*-endosperm, and 1.5 billion bases for aerial tissues, which corresponds to 21-fold, 18-fold, 16-fold, and 13-fold coverage of the *Arabidopsis* nuclear genome, respectively (Table. A2.1-appendix, and Table. 2.1). To determine the quality of our methylome study of the seed, we compared our aerial tissue results to those of another high-throughput methylome study, Cokus et al. (2008) (Fig. 2.2, A to F). This was done using Ends Analysis, a software program that generates a methylation profile of a typical gene or transposon residing in a tissue (Cokus et al., 2008; Zilberman et al., 2007). The methylation profile is an average and is done by aligning all the annotated genes (and transposons) at their 5' and 3' ends, and then plotting the average methylation levels for each 100-bp interval. The use of Ends Analysis or methods similar to it allows the methylomes of different tissues to be compared and has become a standard means for analyzing high-throughput DNA methylation data (Cokus et al., 2008; Lister et al., 2008; Penterman et al., 2007; Zilberman et al., 2007). As shown in Figure 2.2, our aerial tissue methylation data generated similar results to that of Cokus *et al*. in both genes (Fig. 2.2, A, C, and E) and transposable elements (Fig. 2.2, B, D, and E), and for all DNA methylation contexts (Fig. 2.2, A to F), demonstrating our method and analysis to be accurate, relative to the published work of others.

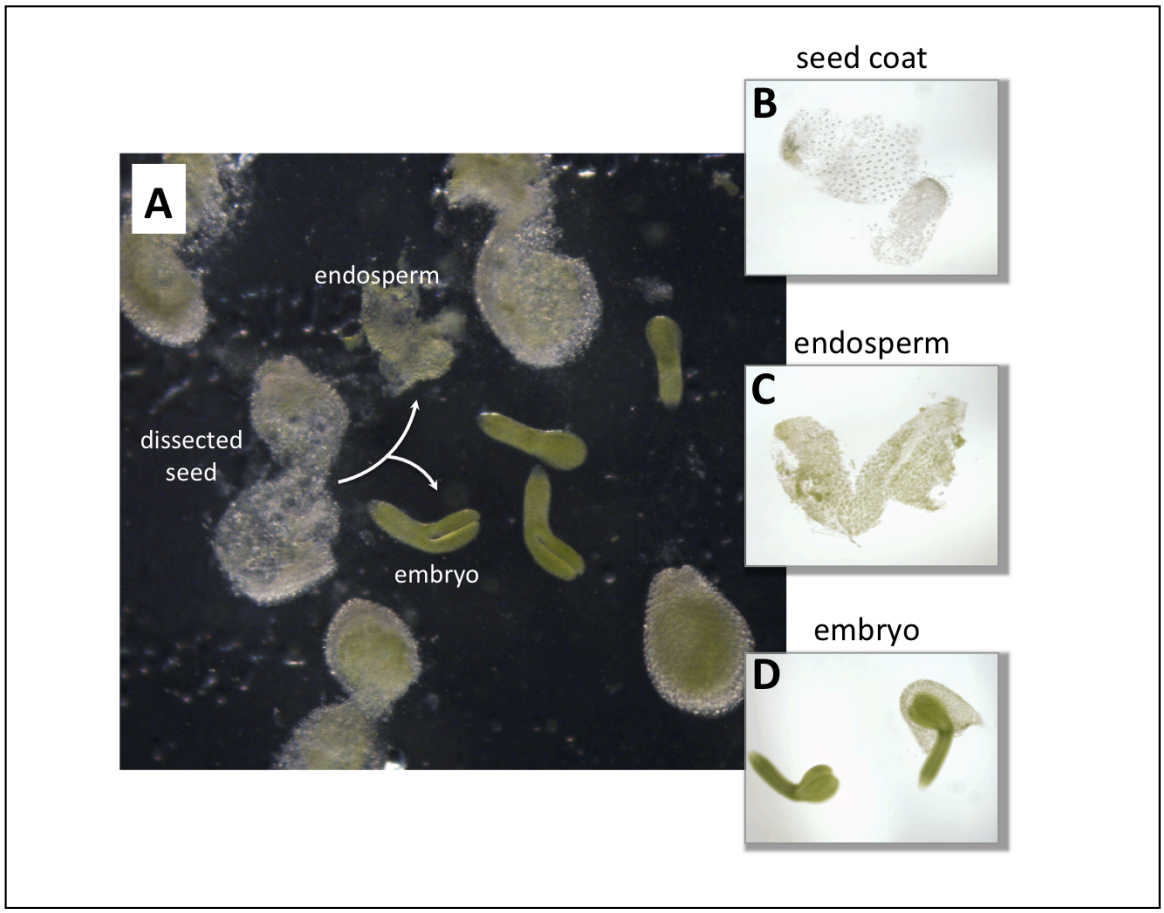


Fig. 2. 1 An *Arabidopsis* 'late-walking stick' stage seed.

Tissue	% CG methylation	% CHG methylation	% CHH methylation
embryo	26.9	10.6	4.4
endosperm	20.9	8.9	2.8
dme-endosperm	23.1	5.8	0.8
aerial	25.7	9.4	2.3

Table. 2.1. Bulk methylation amounts in embryo, endosperm, *dme*-endosperm, and aerial tissue.

DNA Context	Total number of DMRs	More methylated in embryo	More methylated in endosperm
CG	36,749	36,534 (99.4%)	215 (0.6%)
CHG	5,694	5,200 (91.3%)	494 (8.7%)
CHH	9,749	8,760 (89.9%)	989 (10.1%)

Table. 2.2. Differentially methylated regions (DMRs) between embryo and endosperm.

Gene body methylation occurs in the tissues of the seed

Our initial examination of the methylomes of the *Arabidopsis* seed showed the exclusive methylation of gene bodies in the CG context (Fig. 2.3, A, C, and E), while transposon bodies were methylated in all sequence contexts (CG, CHG, and CHH) (Fig. 2.3, B, D, and F). Several genome-wide DNA methylation studies in *Arabidopsis* have revealed 20-30% of genes possessing DNA methylation within the gene body (Cokus et al., 2008; Lister et al., 2008; Teixeira and Colot, 2009; Zhang et al., 2006; Zilberman et al., 2007). The function of plant gene body methylation is unclear, however some studies suggest a relationship exists between gene body methylation and transcription, whereby moderately methylated genes are more likely to be expressed than genes that are either highly methylated or unmethylated (Zilberman et al., 2007). The identification of CG gene body methylation occurring in the nascent tissues of the plant, the endosperm and embryo, suggests the phenomenon to be a fundamental process in plant development.

Ends Analysis shows the endosperm to be hypomethylated compared to the embryo

The egg cell and central cell of the female gametophyte are genetically identical, arising from the same haploid spore. Nonetheless, despite their genetic uniformity, both tissues have very disparate developmental fates, as the embryo continues to mature, eventually becoming the next generation of the plant, while the endosperm terminally differentiates and is consumed by the embryo**.** A comparison of the endosperm and embryo methylomes may elucidate epigenetic processes responsible for the endosperm and embryo's distinct developmental fates. To investigate the epigenetic processes regulating endosperm and embryo development, we examined the DNA methylation profiles of the endosperm and embryo. Our Ends Analysis results shows the endosperm (Fig. 2.3, A to F, blue trace) genome to be reduced in all DNA methylation contexts compared to the embryo (Fig. 2.3, A to F, orange trace). Interestingly, a similar finding was reported in maize, in which the endosperm had a 13% reduction in 5-methylcytosine compared to embryo and leaf tissue (Lauria et al., 2004).

The normal function of DME is to demethylate DNA, and previous locus-specific bisulfite sequencing studies showed loci targeted by DME-mediated DNA demethylation to be hypermethylated in the *dme*-mutant endosperm (Gehring et al., 2006; Kinoshita et al., 2007). However, to what extent is the effect of DME-mediated DNA demethylation throughout the entire endosperm genome? We examined the effects of the DME DNA glycosylase on the endosperm methylation landscape by profiling endosperm lacking a functional DME allele. As shown in Fig 2.3 (A to F, green line), DME affects CG methylation differently than non-CG methylation.

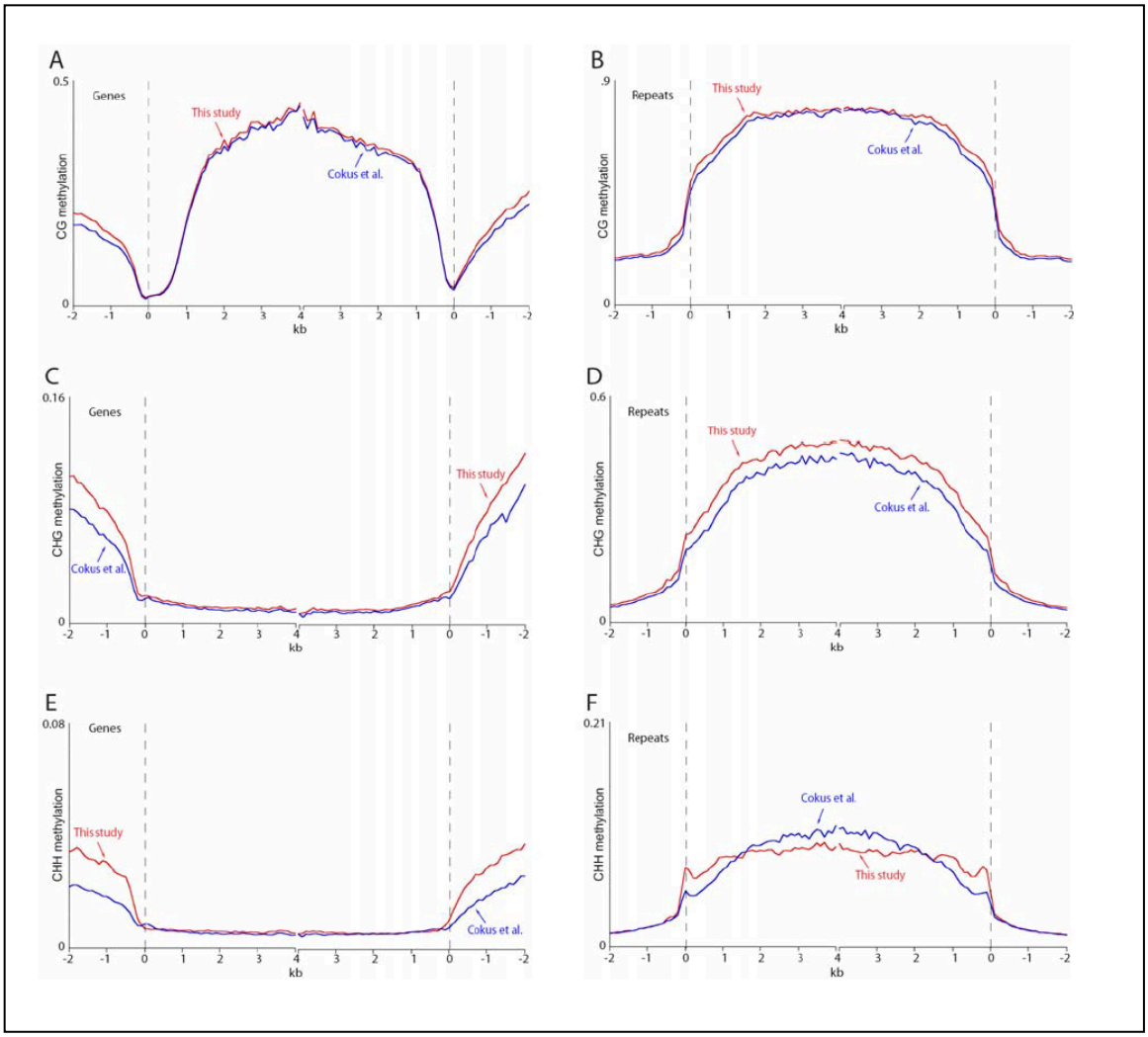


Fig. 2. 2. Ends Analysis comparing our aerial tissue methylation data, to that of Cokus *et al*. (Cokus et al., 2008). TAIR8-annotated genes (A, C, E) or transposons (B, D, F), were aligned at the 5' end (left panel) or the 3' end (right panel) and average methylation levels (y-values) for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 4 kb into the gene (positive numbers). Methylation of wild-type aerial tissues as determined in this study is represented by the red trace, and as determined by Cokus et al. (Cokus et al., 2008) by the blue trace. The dashed line at zero represents the point of alignment. CG methylation is shown in (A-B), CHG in (C D), CHH in (E-F).

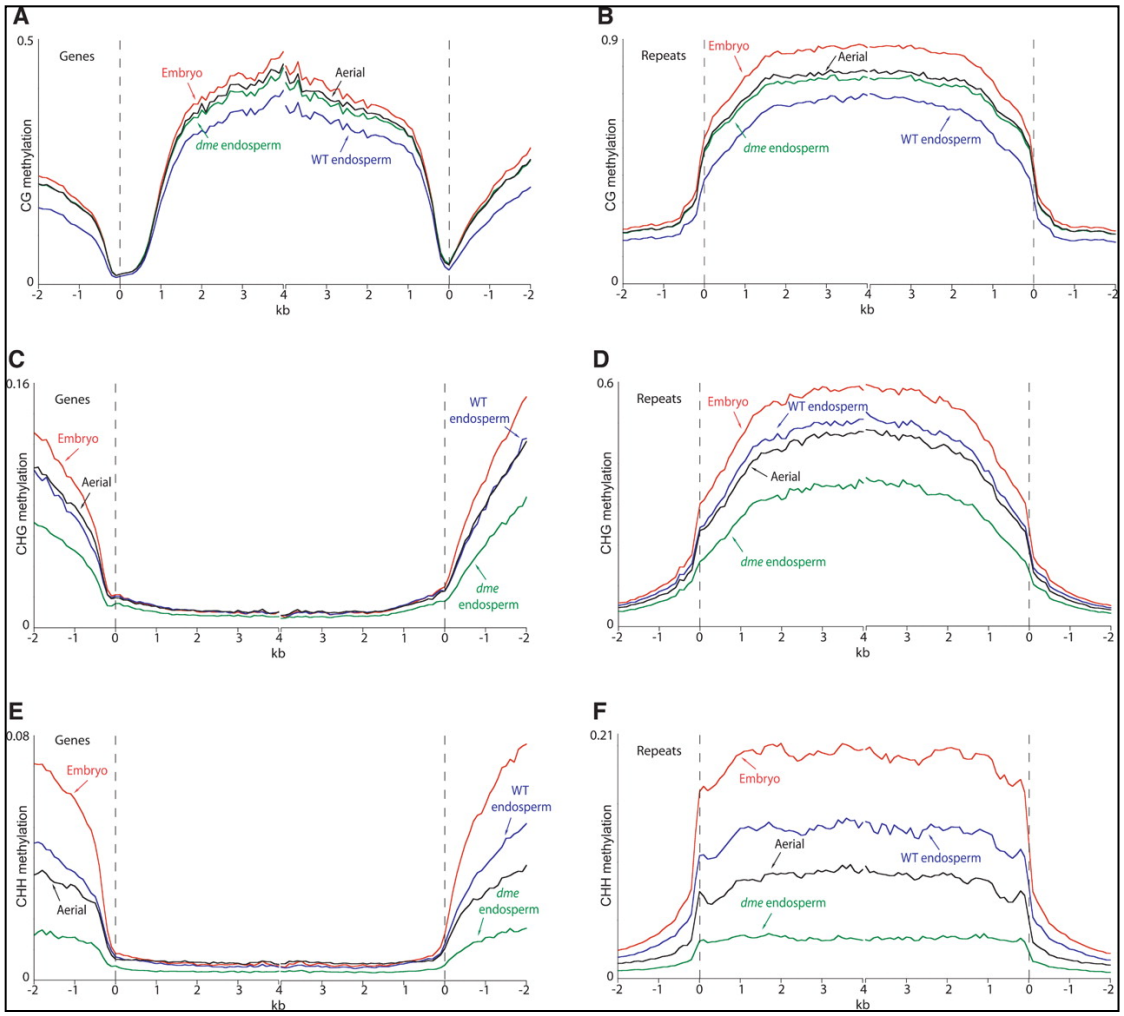


Fig. 2. 3. Profiles of DNA methylation in embryo, wild-type endosperm, and *dme*-endosperm. (A to F) TAIR8-annotated genes [(A), (C), and (E)] or transposons [(B), (D), and (F)] were aligned at the 5′ end (left panel) or the 3′ end (right panel), and average methylation levels (y-values) for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 4 kb into the gene (positive numbers). Embryo methylation is represented by the red trace, wild-type (WT) endosperm by the blue trace, dme endosperm by the green trace, and aerial tissues by the black trace. The dashed line at zero represents the poinalignment. CG methylation is shown in (A) and (B), CHG in (C) and (D), CHH in (E) and (F).

In terms of CG methylation, mutant DME had the effect of globally hypermethylating the endosperm genome, causing an increase in CG methylation for both genes and transposons (Fig. 2.3, A and B, green trace). This is consistent with our prediction that the presence of hypermethylation, is due to a non-functional *dme*-mutant allele. However, in terms of non-CG methylation, the mutant DME allele caused a significant drop in both CHG and CHH methylation (Fig. 2.3, C to F, green trace). This finding was unexpected as it suggested DME was somehow responsible for up-regulating non-CG methylation in the endosperm. In *Arabidopsis*, the maintenance of CHG and CHH methylation is performed by fairly distinct molecular pathways. CHG methylation is done by the plant-specific DNA methyltransferase CMT3, along with the histone methyltransferease SUVH4. Together, CMT3 and SUVH4 work in a reinforcing loop which allows the maintenance of CHG methylation to occur (Law and Jacobsen, 2010). CHH methylation is maintained by the elaborate process of RdDM which uses siRNAs, along with a host of other proteins, to ensure the maintenance of CHH methylation (Fig. 1.3). Hence, the reduction in both CHG and CHH methylation caused by the DME mutation suggests DME is somehow involved in regulating these pathways. Of particular interest is DME's association with CHH methylation, as the *dme*-mutant allele caused almost a complete loss of CHH methylation (Fig. 2.3, E and F, green trace). This dramatic drop in CHH methylation in the *dme*-mutant background suggests DME-mediated DNA demethylation may indirectly up-regulate RdDM in the endosperm.

Most DMRs tend to be hypomethylated in the endosperm genome

To further study the DNA methylation landscapes of the seed, we identified differentially methylated regions (DMRs) in the embryo compared to the endosperm. This was done by calculating the fractional methylation in each context within 50 bp windows and subtracting the endosperm methylation from the embryo methylation. As shown in Table 2.2, we identified 36,749 total DMRs in the CG context, of which 36,534 (99.4%) were more methylated in embryo than the endosperm, and 215 (0.6%) were more methylated in the endosperm. In terms of non-CG methylation, we found 5,694 total DMRs in the CHG context, of which 5,200 (91.3%) were more methylated in embryo, and 494 (8.7%) where more methylated in the endosperm. For CHH methylation we identified 9,749 total DMRs, of which 8,760 (89.9%) were more methylated in embryo than the endosperm, and 989 (10.1%) where more methylated in the endosperm. These values provide a clear indication of the extent of methylation differences between embryo and endosperm, demonstrating that the endosperm is much less methylated, in all contexts, compared to the embryo.

Heat maps of methylation in genic and transposable elements further show the endosperm to be hypomethylated compared to the embryo

We aligned all *Arabidopsis* annotated genes (which included some pseudogenes) and transposable elements, at their 5′ ends, stacked them from the top of chromosome 1 to the bottom of chromosome 5, and displayed the fractional embryo methylation as heat maps (Fig. 2.4, A to C, left panels; Fig. 2.5, A to C, left panels). We also made heat maps by profiling the methylation differences between the embryo and the wild-type endosperm as well (Fig. 2.4, A to C, right panels; Fig. 2.5, A to C, right panels). The use of heat maps to display the methylation profiles of all genes and transposons in a single image provides a useful way to visualize a large amount of data all at once, allowing broad trends in the data set to be quickly identified. According to the heat maps, virtually all sequences methylated in embryo in the CG context were less methylated in the endosperm (Fig. 2.4 A; Fig. 2.5 A). Gene bodies, gene adjacent sequences, and transposable elements were all similarly demethylated in the endosperm, with transposons demethylated to a somewhat greater extent than genes (Fig. 2.4, A to C; Fig. 2.5, A to C). Similar to our non-CG methylation Ends Analysis data (Fig. 2.3, C to F), the heat map data showed both CHG and CHH methylation of most sequences to also be higher in the embryo relative to the endosperm (Fig. 2.4, B and C; Fig. 2.5, B and C).

Using kernel density plots to graph DMRs is a powerful tool for studying genome-wide DNA methylation

Like the heat map, the kernel density plot is also able to display large amounts of DNA methylation data in a single graph and is effective at showing the occurrence of broad tendencies. The kernel density plot is a type of histogram in which the x-axis is a given variable and the y-axis is the frequency or ‘density’ of that particular variable. However, unlike a histogram which is represented as a series of bins, the kernel density estimation is shown as a smoothed-out curve Fig. A2.2 (A and B). Kernel density plots have been used to analyze genome-wide methylation data, which involves plotting the density of DMRs generated from two methylomes. In this case, the x-axis is the numeric value of the ‘fractional methylation difference’ of the DMR (from here on will be called the DMR value), while the y-axis is the density of the DMR. The kernel density plots of DMRs typically give a Gaussian-like distribution, where the center of the main peak shows the average (mean) DMR value (Fig. A2.2), marking the most common occurring DMR value in the entire population. A kernel density plot centered on zero means the average DMR value is zero, suggesting the two methylomes in comparison are, at a genome-wide level, similarly methylated (Fig. A2.2, A). methylome in this case, would be ‘hypomethylated’). An added advantage of using the kernel density plot to study DMRs is its aptness at locating and displaying subpopulations of DMRs with extreme DMR values (i.e. large positive or large negative fractional methylation difference values).

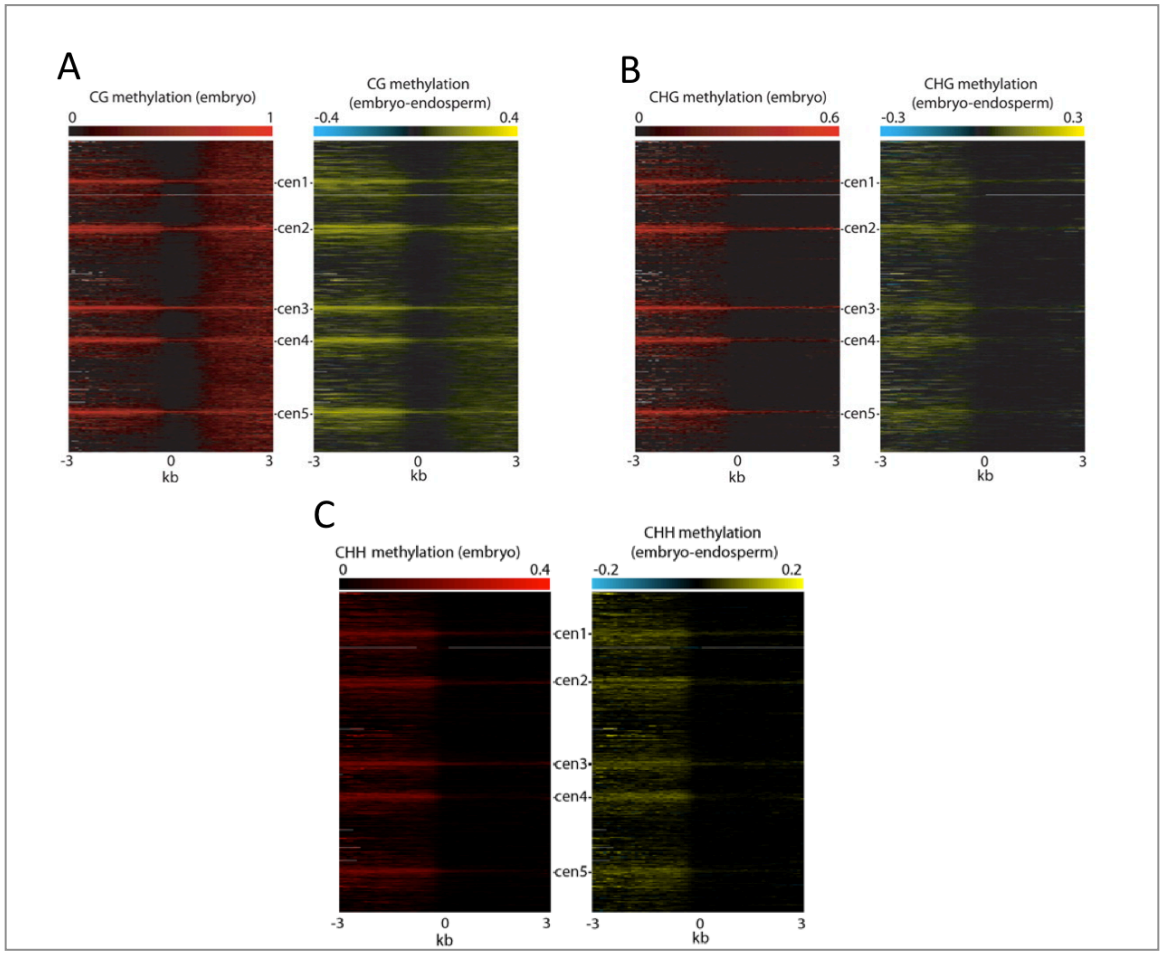


Fig. 2. 4. DNA methylation heat maps of genes in the embryo and wild-type endosperm. All TAIR8-annotated genes (28,244) were aligned at the 5′ end and stacked from the top of chromosome 1 to the bottom of chromosome 5. Embryo methylation is displayed as a heat map in the left panel, while the differences between embryo and wild type endosperm are in the right panel. CG methylation is shown in (A), CHG in (B), and CHH in (C). Numbers at the top of each individual heat map represent fractional methylation values (zero signifies no methylation, while more intense color signifies elevated methylation).

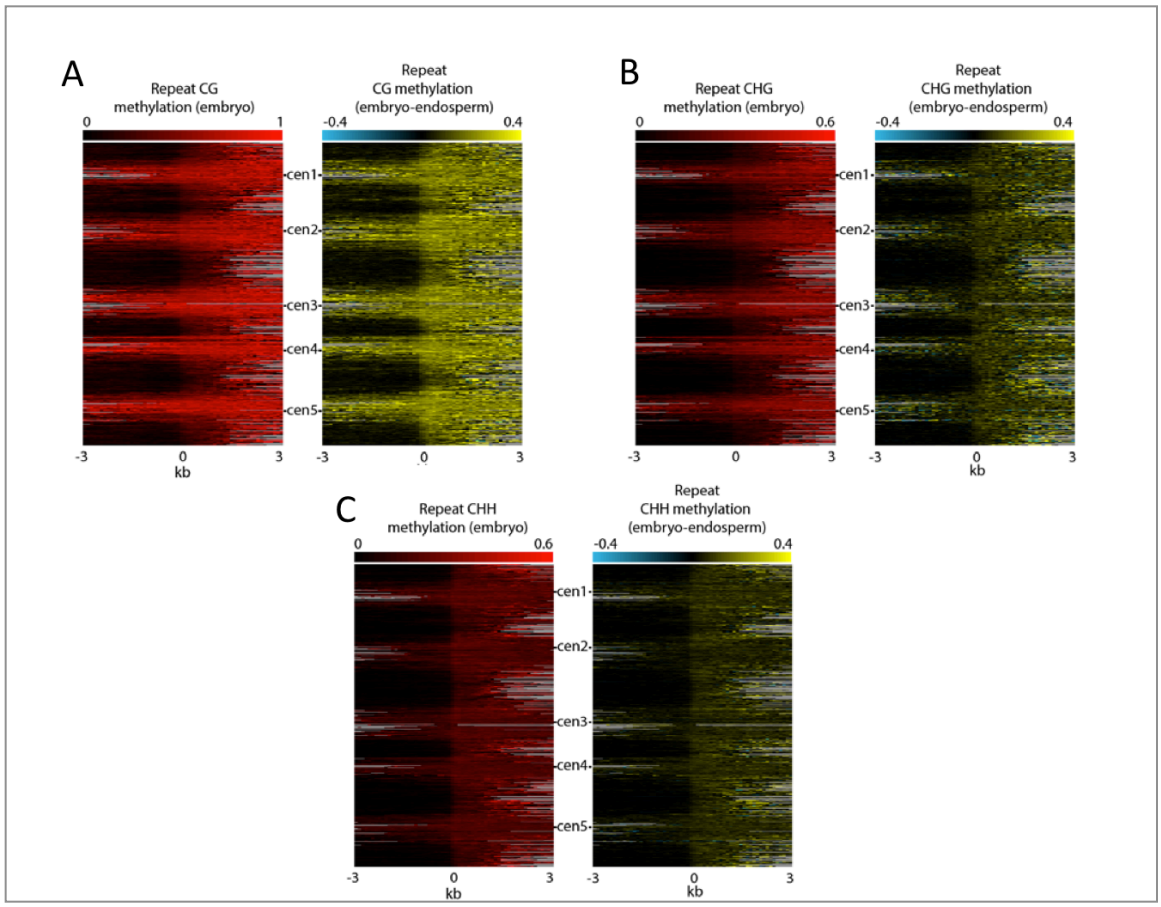


Fig. 2. 5. DNA methylation heat maps of transposons and repeats in the embryo and wild-type endosperm. Heat maps of repeats. All TAIR 8 annotated transposable elements and repeats (31,076) were aligned at the 5' end and stacked from the top of chromosome 1 to the bottom of chromosome 5. Embryo methylation is displayed as a heat map in the left panel, while the differences between embryo and wild type endosperm are in the right panel. CG methylation of transposable elements is shown in (A), CHG in (B), and CHH in (C). Numbers at the top of each individual heat map represent fractional methylation values (zero signifies no methylation, while more intense color signifies elevated methylation).

Such values lie far from the mean DMR value and are not buried in the large data set. DMRs with extreme values can be important as they represent highly methylated (or highly demethylated) sites with in the genome. Figure A2.2 (B) shows a subpopulation of DMRs (red arrow) that are significantly hypermethylated (>0.5) in hypothetical Tissue X relative to Tissue Y, or hypomethylated in Tissue Y compared to Tissue X.

Kernel density plots also reveal the effects of the dme-mutant allele on the endosperm genome

We generated kernel density plots by graphing the distribution density of DMRs from wild-type endosperm subtracted from embryo (Fig. 2.6, A to C, blue trace), *dme*-endosperm subtracted from embryo (Fig. 2.6, A to C, red trace), and aerial tissue subtracted from embryo (Fig. A2.3, A to C). Similar to our Ends Analysis results, the kernel density plot findings also demonstrated the endosperm genome to be hypomethylated compared to the embryo, by being shifted to the right of zero and occurring in all contexts (Fig. 2.6, A to C). The *dme* mutation uniformly restored CG methylation, which was seen as the *dme*-endosperm subtracted from embryo plot (Fig. 2.6 A, orange trace) being shifted to the left of the wild type endosperm subtracted from the embryo plot (Fig. 2.6 A, blue trace). The slight shift to the left in the *dme*-mutant background plots is a consequence of the fractional methylation differences becoming smaller due to the increase in methylation, caused by the lack of DME-mediated DNA demethylation in the endosperm. Also similar to the Ends Analysis results, were the non-CG methylation kernel density plots which demonstrated the unexpected reduction in both CHG and CHH methylation in the *dme*-endosperm background (Fig. 2.6, B and C, red trace). In this case, both of the plots (Fig. 2.6, B and C, red trace) were shifted to the right of their respective wild type plots (Fig. 2.6, B and C, blue trace) due to larger fractional methylation differences caused by the DME-mediated drop in non-CG methylation.

The methylation profiles of several known imprinted-related DMRs coincide with the genome-wide DNA demethylation event

We calculated methylation levels of sequences (i.e. DMRs) either known or strongly inferred to cause imprinted expression of five *Arabidopsis* genes: the *MEA* 3′ repeats, the *FWA* promoter and start of transcription, the *FIS2* promoter, the *PHE1* 3′ repeats, and the *MPC* gene and flanking regions (Fig. 2.6, A to C, symbols). For the DMR associated with *MEA* imprinting, methylation was reduced from 88% CG, 39% CHG, and 42% CHH in embryo to 63% CG, 16% CHG, and 17% CHH in wild-type endosperm. Methylation in the CG context was restored to 87% in *dme*-endosperm, whereas CHG (13%) and CHH (8%) methylation was further reduced. Methylation profiles of the other four aforementioned imprinted-related DMRs behaved in a similar fashion (Fig. 2.6, A

to C, symbols). Hence, our data shows many of the imprinting-related DMRs follow the same methylation patterns occurring on a genome-wide level, throughout the endosperm trends.

The endosperm contains DME-independent hypermethylated non-CG DMRs

As shown in Table 2.2, only a small amount of DMRs occurring in the CG context were shown to be hypermethylated in the wild type endosperm (less than 1.0%) compared to the embryo (99.4%). However, for non-CG DMRs, a larger amount exhibited endosperm hypermethylation (~10% for both CHG and CHH methylation) (Table 2.2). We sought to further investigate the nature of these hypermethylated non-CG sites by seeing if DME played a role in their presence. This was done by identifying the genomic coordinates of the endosperm-specific hypermethylated non-CG sites, and then examining their methylation status in the *dme*-endosperm (subtracted from embryo) background. The resulting loci were then plotted via a kernel density plot (Fig. 2.6, B and C, green trace). If the observed endosperm hypermethylation was associated with DME-mediated DNA demethylation, then we would expect to see a correlation between hypermethylation in the wild type and *dme*-endosperm, which would show as either an increase or decrease in DNA methylation at the observed genomic coordinates. As shown in Figure 2.6 (B and C, green trace), a large portion of the candidate coordinates remained hypermethylated in the *dme*-endosperm background demonstrating their independence from DME-mediated DNA demethylation, despite the overall reduction of non-CG methylation caused by the *dme* mutation. Because the maintenance of non-CG methylation is an siRNA-mediated process, we wanted to determine if these non-CG hypermethylated DMRs, specific to the endosperm, were associated with siRNA sequences. Hence, we used published siRNA data from immature *Arabidopsis* flower buds (Lister et al., 2008) to count the number of siRNAs with 50 bp windows covering the genome. As demonstrated in Figure. 2.7, both CHG and CHH endosperm hypermethylated loci show enrichment for the flower bud-derived siRNAs, suggesting the methylation to be maintained by a highly specific RNAi-targeted process.

The embryo methylome is comprised of extensive CHH hypermethylation suggesting RdDM is elevated during early plant development

Our high-throughput DNA methylation study has shown the genome of the embryo to be more methylated than the genomes of the wild type endosperm, *dme*-mutant endosperm, and adult aerial tissue. In particular was the presence of extensive CHH hypermethylation in the embryo relative to the aerial tissue. As shown in Table 2.1, bulk symmetric methylation was only slightly higher in the embryo (26.9% CG and 10.6% CHG) than in the aerial tissue (25.7% CG and 9.4% CHG), resulting in a 4% and 5% increase in CG and CHG methylation respectively. However, asymmetric methylation in the aerial tissue (2.3% CHH)

appeared to be significantly higher in the embryo (4.4% CHH), showing a ~ 2 fold increase in CHH methylation. The heighten levels of CHH methylation in the embryo suggest the siRNA-mediated DNA methylation pathway to have increased activity in the embryo.

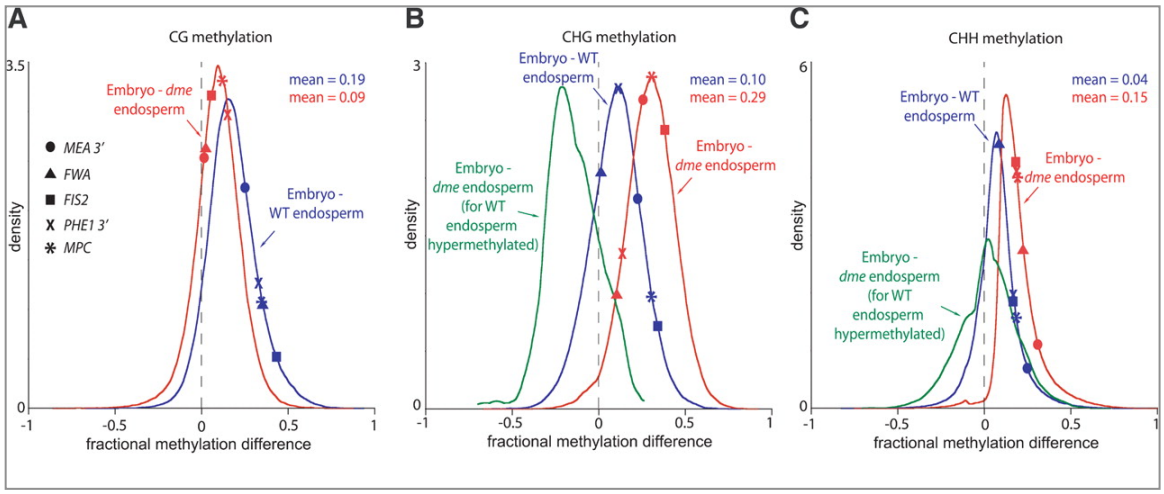


Fig. 2. 6. Kernel density plots demonstrating genome-wide demethylation of the *Arabidopsis* endosperm. (A to C) Kernel density plots of the differences between embryo and WT endosperm methylation (blue trace) and the differences between embryo and dme endosperm methylation (red trace). The green trace in (B) and (C) represents methylation differences between embryo and dme endosperm for windows with absolute fractional methylation increase in WT endosperm compared with embryo of at least 0.4 in the CHG context (B) (n = 135) or at least 0.2 in the CHH context (C) (n = 6168).

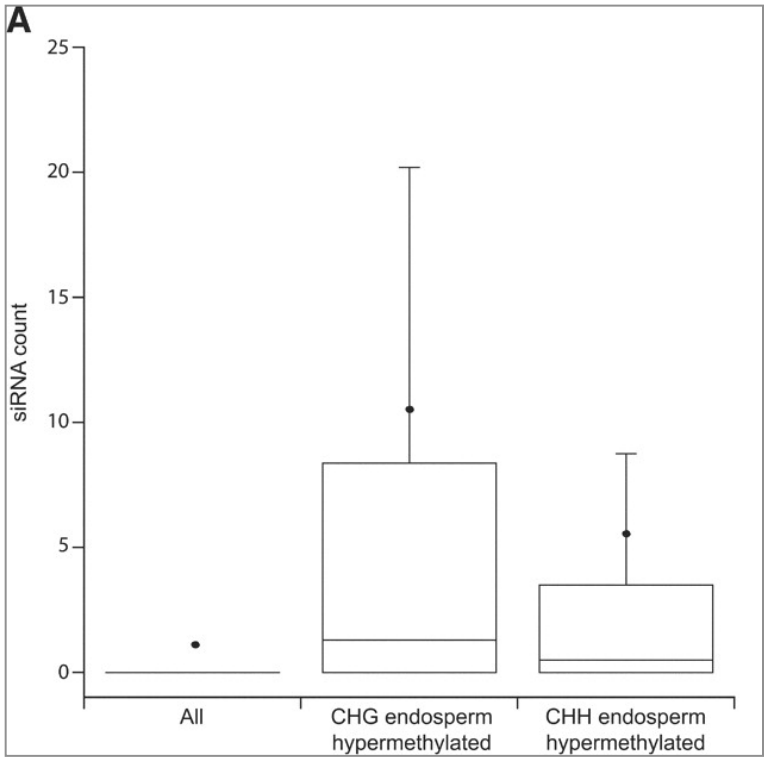


Fig. 2. 7. Association between non-CG methylated DMRs in the endosperm and siRNAs. Box and whisker plots showing siRNA abundance within 50-bp windows in the entire *Arabidopsis* genome (All) and in sequences hypermethylated in *wild-type* endosperm compared with the embryo in the CHG and CHH contexts.

DISCUSSION

Genome-wide DNA demethylation in the endosperm

In this study a high-throughput DNA methylation approach was used in order to gain a detailed image of the *Arabidopsis* endosperm and embryo methylomes. A key finding of our study is the endosperm's marked reduction in DNA methylation compared to the embryo. Confirming this observation has been the work of Gehring *et al.* who also used a high-throughput approach to examine the methylomes of the *Arabidopsis* seed (Gehring et al., 2009a). Using MeDIP coupled to illumina™ sequencing, they too witnessed a genome-wide decrease in DNA methylation in the endosperm compared to embryo (Gehring et al., 2009a).

Microscopy studies show endosperm nuclei to be considerably larger than the nuclei of other tissues (i.e. leaf, meristematic root, seed coat), due to the chromatin being in a relatively decondensed state (Baroux et al., 2007). It has been suggested this particular nuclear architecture of the endosperm nuclei reflects the presence of high metabolic activity involving genome reorganization and reprogramming (Baroux et al., 2007). Our discovery of wide-spread DNA demethylation in the endosperm might, in part, influence endosperm genome architecture.

Genome-wide DNA demethylation occurs in the central cell and is an indirect effect of DME-mediated DNA demethylation

We suspect the observed genome-wide DNA demethylation in the endosperm occurs sometime during central cell development. Indeed, prior to our work, Jullien *et al*. suggest a model where by genome-wide DNA demethylation could occur on the maternal genome in the central cell, before fertilization, due to a down-regulation of the maintenance methylation machinery (Jullien et al., 2008). During the final stages of female gametogenesis in the central cell, a complex formed by RBR1 and MSI1 acts to transcriptionally repress the MET1 methyltransferase required for maintaining CG methylation producing a hemimethylated genome. This hemimethylated genome could then be further demethylated by DME. An attractive aspect of this model suggests DME may prefer targeting hemimethylated and not fully methylated DNA, as it has been demonstrated that active DNA demethylation occurring on symmetrically methylated DNA can lead to double-strand breaks potentially lethal to genome integrity.

Another possible mechanism responsible for the endosperm genome-wide DNA demethylation involves an indirect interaction of DME-mediated DNA demethylation. A recent allele-specific RNA-seq study conducted in our lab (Hsieh et al., 2011), showed both VIM5 and MET1 to be up-regulated in *dme*-

mutant and *fie*-mutant endosperm. Because of the important role VIMs and METs play in maintaining CG-methylation patterns, this suggests the endosperm hypomethylation may also involve coordination between VIM5 and MET1. Thus, genome-wide DNA demethylation in the endosperm could be due to both the RBR1-mediated down regulation of MET1 (Jullien et al., 2008), as well as the DME-dependent activation of PRC2 which works to repress VIM5, and the CG-maintenance methylation machinery. All of this is carried out in the central cell, prior to fertilization. However, the ensuing endosperm genome-wide DNA demethylation is not caused by the direct activity of DME, but rather its indirect activity. We suspect loci directly targeted by DME-mediated DNA demethylation in the genome to be present in our high-throughput methylation data, but buried in the large trends caused by the genome-wide demethylation. Indeed, a higher-resolution analysis of our data using the SignalMap™ genome-browser, showed sites considerably more hypomethylated than those of the genome-wide demethylation (data not shown). These sites appeared to be DME targets, since they became hypermethylated in the *dme*-mutant background endosperm (data not shown).

Is there a link between genome-wide DNA demethylation and genomic imprinting?

Similar to the global methylation patterns, each DMR was hypomethylated in the endosperm compared to the embryo, and in the *dme*-endosperm exhibited an increase in CG methylation and a decrease in non-CG methylation. Hence, our observation of imprinting-related DMRs corresponding to genome-wide DNA demethylation patterns, suggests one of the mechanisms involved in plant genomic imprinting - DME-mediated DNA demethylation - might act in a less discreet manner regarding its capacity to target localized regions of the genome for active DNA demethylation. Instead, DEMETER performs a broad demethylation within the endosperm genome, and in doing this, imprinting-related DMRs are inadvertently demethylated as well, possibly resulting in the expression of the adjacent genes they regulate.

Similarly, Gehring et al. also observed the same correlation between genome-wide DNA demethylation and the methylation profiles of imprinted-related DMRs. Moreover, they also saw a significant portion of the imprinted-related DMRs to over lap with transposons and repetitive elements, many of which contained siRNA targeted loci (Gehring et al., 2009a). Indeed, the relationship between transposons and DMRs was considerably strong, as Gering et al used it as a criterion to search for new imprinted genes, of which fifty were discovered (Gehring et al., 2009a).

Thus, both our work and the work of Gehring et al. suggest imprinted genes to not be exceptional sequences specifically targeted for demethylation, but rather part of a nearly universal process that reshapes DNA methylation of the entire maternal genome in the endosperm (Jullien et al., 2008). Imprinted expression of genes regulated by allele-specific DNA methylation could

potentially arise whenever a transposable element insertion or a local duplication near a gene's regulatory sequences induces methylation and gene silencing in other tissues, including the paternal endosperm genome.

What role does DEMETER play in the endosperm genome-wide demethylation?

Our results suggest that the DEMETER DNA glycosylase is at least partially involved in the genome-wide DNA demethylation occurring in the endosperm. Initially, we surmised the *dme*-mutation would cause DNA hypermethylation on all its 5mC targets, regardless of whether it acted in a discreet or global manner, since wild type DEMETER showed no preference for demethylating DNA in a specific context (Gehring et al., 2006; Mok et al., 2010). Instead, the expected mutant phenotype of hypermethylation appeared to be both global (Fig. 2.6, A) and mutually exclusive, taking place on cytosines only in the CG context (Fig. 2.3, A and B; Fig. 2.6, A). For cytosines in the non-CG context there was an unexpected reduction in both CHG and CHH methylation (Fig. 2.3, C-F; Fig. 2.6, B and C). This reduction in non-CG methylation in the *dme*-mutant hinted that DEMETER was somehow involved in regulating *de novo* methylation and RdDM in the endosperm.

On the one hand, the *dme*-induced CG hypermethylation shows one aspect of DME's function, which is to directly target the genome for active DNA demethylation. On the other hand, the reduction in non-CG methylation in the *dme*-endosperm background suggests DME to also be responsible for up-regulating RdDM in the endosperm as well. How then, does DEMETER carryout these two seemingly disparate functions?

A model for the DME-dependent up-regulation of non-CG methylation in the endosperm

We envision a scenario whereby in the central cell prior to fertilization, the DEMETER DNA glycosylase carries out active DNA demethylation on all methylation contexts and throughout the entire maternal genome. Because a large portion of the methylated regions are within transposons and repetitive elements, the DME-mediated DNA demethylation effectively removes the repressive methylation mark from these elements allowing them to be expressed. Many of these non-genic elements have loci targeted by siRNAs, partaking in siRNA-mediated DNA methylation (RdDM). Once expressed, the mRNA transcripts from these non-genic elements are then processed into siRNAs, and are eventually incorporated into the RdDM pathway. Figure 2.8 shows a schematic of our proposed model depicting DME-mediated DNA demethylation in the *Arabidopsis* female gametophyte, and how it might cause the increased activity of non-CG methylation. Notably, our model accounts for the observed reduction in CHH methylation, produced by the *dme*-mutant (Fig. 2.8, B).

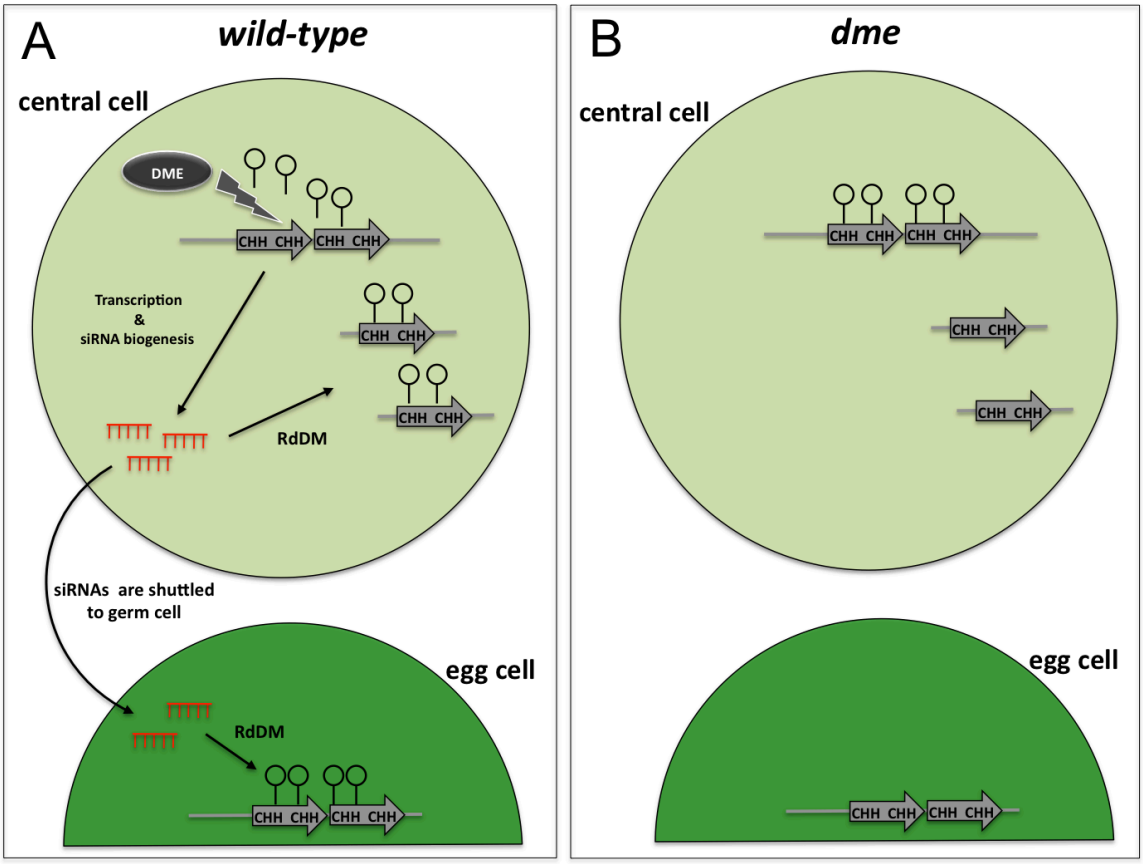


Fig. 2. 8 A model suggesting how DME-mediated DNA demethylation may regulate CHH methylation in the central cell, as well as reinforce the silencing of transposons in the germ cell (i.e. egg cell).

Can genome-wide DNA demethylation in the endosperm reinforce transposon silencing in the plant germline?

Our model of genome-wide DNA demethylation suggests the presence of a large pool of siRNAs residing in the endosperm (Fig. 2.8). Some of these siRNAs appear to feed into the RdDM pathway, resulting in elevated levels of non-CG methylation in the wild-type endosperm (Fig. 2.8, A) compared to *dme*-endosperm (Fig. 2.8, B). In addition to elevated non-CG methylation in the endosperm, high non-CG methylation was also witnessed in the embryo (Table. 2.1). We hypothesize the high-levels of embryo CHH methylation might be a consequence of heightened levels of siRNAs in the embryo, and suspect they might originate from the endosperm/central cell and move to the egg/embryo. As shown in Figure 2.8, a key component to our model is the putative movement of siRNAs from the companion central cell to the germ line egg cell.

This notion of cell-specific transposon-derived siRNAs moving to a neighboring cell in order to reinforce the transposon-supression system may be a fundamental process used by biological systems to regulate inherited deleterious mobile elements between generations. Prior to our work, a study done in the *Arabidopsis* male gametophyte by Slotkin *et al.* (Slotkin et al., 2009) revealed a class of 21 nt transposon-generated siRNAs originating from the vegetative cell, but capable of carrying out transposon silencing in the sperm cells. An elegant model was proposed in which the transposon regulator DDM1 was down-regulated in the vegetative cell causing a global reactivation of transposable elements and subsequent production of transposon-derived siRNAs. These vegetative cell generated siRNAs would then move into the adjacent sperm cells and once inside, carryout siRNA-mediated transposon-silencing. Thus, having a surplus of readily available siRNAs not only guarantees the silencing of deleterious mobile elements, but also reinforces it, which is particularly more important for generative type cells such as sperm cells. We suspect a similar scenario may exists between the central cell and egg cell.

METHODS

Seeds at the mid-torpedo stage to early maturation stage (7-9 days after pollination, ecotype Col-0) were dissected in 0.3 M sorbitol, 5 mM MES (pH 5.7) on a slide under a dissecting microscope. At this stage of development, bilaterally active genes are expressed from both maternal and paternal genomes in embryo and endosperm (Vielle-Calzada et al., 2000). Embryos were washed to remove contaminating endosperm. Because the *Arabidopsis* seed coat has considerable tensile strength, it was possible to separate endosperm from the seed coats and extract DNA from pure endosperm, as previously described (Gehring et al., 2006; Kinoshita et al., 1999). We also extracted DNA from aerial tissues of 4-week old Col-0 plants.

Bisulfite sequencing

For our experiments we adapted the protocol published recently by Joseph Ecker and colleagues (Lister et al., 2008). First, we synthesized custom Illumina ™ adapters in which cytosines were replaced by 5-methylcytosines, so that the adapters would survive bisulfite conversion. We chose to synthesize paired ends (PE) adapters, which allow each molecule to be sequenced from both ends, thus facilitating subsequent alignment to the genomic scaffold. The adapter oligos (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'; 3'-GAGCCGTAAGGACGACTTGGCGAGAAGGCTAGp-5'; oligonucleotide sequences © 2006 and 2008 Illumina™) were annealed as described by Cronn et al. (Cronn et al., 2008), generating a Y-shaped molecule (Fig. A2.1). We isolated 0.5-1 micrograms of genomic DNA from endosperm dissected from wild type and *dme* seeds, as well as control wild type embryos and aerial tissues. We did not use *dme* embryos because they abort at a very early stage. DNA was sheared by sonication to fragments of 100-500 bp, and the adapters were ligated following the Illumina™ protocol. The library was then treated twice with sodium bisulfite (which converts unmethylated Cs to Us) using the Qiagen EpiTect™ kit, and amplified by 18 cycles of PCR using PfuTurboCx DNA polymerase (Stratagene ™), a proofreading enzyme that tolerates uracil in the template strand. PCR amplification results in a library with distinct adapters at each end, so that the 'forward' Illumina™ sequencing primer always produces sequence from the 'original' genomic DNA-derived strand (where Cs represent methylated Cs), and the 'reverse' Illumina™ sequencing primer always produces sequence from the PCR-generated strand (where Gs represent methylated Cs on the opposite strand). Bands around 300 bp were gel-purified and cloned for validation. Traditional sequencing confirmed that the libraries were properly constructed: the clones were fully converted and had correct adapter sequences on each end. The libraries were sequenced at the UC Berkeley Genomic Sequencing Laboratory, generating 45 bp reads from each end (52.3 million for embryo, 46.7 million for WT endosperm, 40.7 million for *dme*-endosperm, and 28.7 million for adult aerial tissues).

Sequence analysis

To align the reads to the TAIR8 *Arabidopsis* genomic scaffold, we adapted the strategy described by Meissner et al. (Meissner et al., 2008). We used custom Perl scripts to convert all the Cs in the 'forward' reads (and in the scaffold) to Ts, and all the Gs in the 'reverse' reads and scaffold to As, and initially aligned the converted reads to the converted scaffold using SeqMap as individual reads (Jiang and Wong, 2008), allowing up to two mismatches per read, with 75% of the mapped reads aligning without mismatches. We subsequently used a Perl script to insure that paired reads mapped to opposite strands within 300 bp of each other, uniquely mapping a total of 207,060,231 reads (61.5% of all reads), 21.7% percent of which were mapped by using pair

information to resolve non-unique matches. We uniquely aligned 2,460,620,115 bases for embryo, 2,187,031,905 bases for wild-type endosperm, 1,959,340,860 bases for *dme*-endosperm, and 1,547,474,265 bases for aerial tissues to the nuclear genome, and 368,429,850 bases for embryo (2,385-fold coverage), 161,853,975 bases for wild-type endosperm (1,048-fold coverage), 215,443,035 bases for *dme*-endosperm (1,395-fold coverage), and 135,402,300 bases for aerial tissues (877-fold coverage) to the chloroplast genome (Table 2.1). We then used Perl scripts to recover the original sequence information and, for each C (on either strand), count the number of times it was sequenced as a C or a T. For each sequence context (CG, CHG, CHH) we calculated bulk fractional methylation (#C/(#C+#T)), as well as fractional methylation within a 50 bp sliding window. Chloroplast bulk fractional methylation was 0.10% for embryo, 0.10% for wild type endosperm, 0.15% for *dme*-endosperm, and 0.35% for aerial tissues (Table 2.1), which we used as an aggregate error measure.

Methylation analysis

Our results closely matched both published conventional bisulfite sequencing (data not shown) and high-throughput bisulfite sequencing data (Fig. 2.2 A-F) (Cokus et al., 2008; Gehring et al., 2006; Jullien et al., 2006)(Penterman et al., 2007; Soppe et al., 2000; Tiwari et al., 2008). We used the high-throughput dataset published by Cokus et al. (Cokus et al., 2008) for comparison because it was derived from very similar tissues (aerial tissues of 5-week-old Col-0 plants) to those used for our aerial dataset (4-week-old Col-0 plants), but utilized different methods for library construction, bisulfite conversion and read alignment. We calculated methylation differences for 50 bp windows containing at least 20 informative sequenced cytosines in each dataset. Mean difference for windows with fractional CG methylation of at least 0.7 in either our aerial dataset or the Cokus et al. dataset was 0.0006, over 300 times smaller than the difference between our embryo and WT endosperm datasets. Mean difference for windows with fractional CHG methylation of at least 0.5 in one of the tissues was 0.029 and mean differences for windows with fractional CHH methylation of at least 0.1 in one of the tissues was -0.006. The same comparison parameters were used to generate the density plots in Fig. 2.6 (A – B) and Fig. A2.3 (A – C). To identify sequences with significant methylation differences between two tissues, adjacent 50 bp windows containing any methylation in the relevant context in at least one of the tissues being compared were concatenated. Concatenated windows with at least 20 informative sequenced cytosines and a methylation difference of at least 0.1 for CG and CHG contexts and at least 0.05 for CHH context at a $p<0.0001$ (Fisher's exact test) are reported (data not shown).

siRNA analysis

We used published siRNA data from *Arabidopsis* flower buds (Lister et al., 2008) to count the number of siRNAs within 50 bp windows covering the genome. Fig. 2.8 depicts all windows, or those matching sequences hypermethylated in either the CHG or CHH context in the endosperm.

Chapter III

DEMETER-mediated DNA demethylation targets similar loci in gamete companion cells

PREFACE

The experiments regarding isolation of vegetative and sperm cells of the *Arabidopsis* male gametophyte by way of FACS, were carried out by Drs. Vera Schoft and Hisashi Tamaru (Gregor Mendel Institute of Molecular Plant Biology, Vienna, Austria).

INTRODUCTION

The DEMETER DNA glycosylase has demonstrated expression in both companion cells of the female and male gametes of *Arabidopsis*. In the central cell, DME-mediated DNA demethylation is critical to *Arabidopsis* development and imprinting, since *dme*-mutant plants give a characteristic mutant phenotype of an over-proliferated endosperm resulting in an aborted seed (Choi et al., 2002). The seed abortion is specifically due to inheriting the maternal *dme*-mutant allele. In the vegetative cell, DME activity is required for proper pollen development, as pollen granules inheriting the *dme*-mutant allele do not germinate properly, nor are they able to form a functional pollen tube (Schoft et al., 2011). Hence, DME-mediated DNA demethylation is an important process occurring in the accessory cells of the male and female gametophytes.

Interestingly, the imprinted genes *FWA* and *MEA* which are known to be targeted by DME in the central cell have been shown to also be targeted by DME in the vegetative cell. Moreover, the DME-mediated DNA demethylation of these two genes in the vegetative cell has led to their expression (Schoft et al., 2011).

In this chapter, our goal was to gain insight into DME-mediated DNA demethylation in the *Arabidopsis* central cell and vegetative cell genomes. Because the ability to isolate adequate amounts of central cells for bisulfite sequencing remains an extreme technical challenge, we chose an SNP approach where ecotype-specific crosses generating allele-specific F1 endosperm was used to distinguish the maternal and paternal genomes of the endosperm. In particular, our interest was in examining the DNA methylation patterns of the maternal endosperm genome, since prior to fertilization it was the central cell genome. Thus, examining the maternal genome of the endosperm served as an indirect approach for examining the central cell genome. We also examined the methylomes of the vegetative and sperm cells of the male gametophyte as part of a collaboration with Professor Hisashi Tamaru and his group from the Gregor Mendel Institute, in Austria (Schoft et al., 2011). By comparing the DNA methylation patterns between the maternal endosperm and vegetative cell genomes, we were able to discover loci that appeared to be targeted by DME-mediated DNA demethylation in both genomes. Moreover, these DME-dependent loci were enriched in transposable elements, particularly smaller ones that appeared frequently at the ends of genes. Our findings suggest the targeting of small transposons by DME results in gene-adjacent demethylation, and adds a new perspective to *Arabidopsis* genomic imprinting.

RESULTS

Allele-specific sequencing and methylation analysis

To examine the maternal and paternal endosperm methylomes, genomic DNA was isolated from wild type F1 endosperm generated from crossing the Col-

0 and Ler ecotypes. The Col-0 and Ler accessions of *A. thaliana* differ by over 400,000 single nucleotide polymorphisms, allowing us to distinguish between the maternal and paternal genomes. Genomic DNA from *dme*-mutant F1 endosperm generated by crossing *dme* female carpel (Ler) with wild type male pollen (Col-0) was also extracted. Genomic DNA libraries were then made and bisulfite converted, followed by high-throughput DNA methylation analysis using the Illumina Hi-seq™ 2000 platform (Hsieh et al., 2009; Lister et al., 2008). We obtained over 50 million reads of sequencing data from each cross and aligned approximately 5 billion bases for the Col-0 x Ler wild type endosperm, 7 billion bases for the Ler x Col-0 wild type endosperm, and 3 billion bases for the *dme*-mutant endosperm (Col-0 x Ler) (data not shown).

Kernel Density plot analysis of the allele-specific endosperm shows a subpopulation of maternal loci potentially targeted by DME-mediated DNA demethylation

To further interrogate the endosperm methylome, kernel density plots were generated by graphing the distribution of density of (CG context) DMRs made by subtracting the paternal methylome from the maternal methylome in wild-type embryos, endosperm, and *dme*-mutant endosperm (Fig. 3.1, B). As shown in Fig. 3.1 (black trace), the wild-type endosperm maternal and paternal genomes produced a plot that was centered almost on zero (0.015 fractional methylation difference), demonstrating the maternal and paternal genomes of the *Arabidopsis* endosperm to be globally similarly methylated. Interestingly, a small subpopulation of cytosines were present (Fig. 3.1, B, black arrow) exhibiting a high fractional methylation difference value (>0.5) and appearing to be maternally hypomethylated. The abundance of these maternal endosperm hypomethylated DMRs were greatly reduced in the *dme*-mutant endosperm background (Fig. 3.1 B, green trace) demonstrating their origins to be DME-dependent. Globally, both maternal and paternal genomes of the embryo were similarly methylated in all DNA contexts (Fig. 3.2, A - C), lacking any distinct subpopulation of strongly hypomethylated (or hypermethylated) loci. This was somewhat expected, as DME is not expressed in this tissue. The allele-specific methylation analysis of the embryo serves as a negative control for DME activity.

Global DNA methylation in the non-CG context was slightly higher on the maternal genome than the paternal genome of the endosperm. As shown in Fig. 3.1 (D and F, black trace), the kernel density plot of the difference between paternal and maternal endosperm CHG methylation was slightly shifted to the left, indicating the maternal endosperm genome to be globally hypermethylated compared to the paternal genome. Similarly, the maternal endosperm genome also displayed higher CHH methylation levels than that of the paternal genome (Fig. 3.1, F), although the maternal CHH hypermethylation was much less than the maternal CHG hypermethylation (e.g. peak closer to zero).

Ends Analysis shows the maternal genome to be hypomethylated in the CG context compared to the paternal genome on the endosperm

Ends analysis of the allele-specific data showed both genes (Fig. 3.3, A – F) and transposable elements (Fig 3.1, A, C, and E**)** to exhibit similar overall methylation profiles as that of our nonallele-specific data (Fig. 2.2, A-F; Fig. 2.3, A-F; Chapter 2), in that gene body methylation was still explicitly found in the CG context (Fig. 3.3, A and B), while methylation in all contexts occurred in the bodies of transposons and repetitive elements (Fig. 3.1, A, C, and E). Moreover, both sets of reciprocal crosses gave equivalent methylation profiles, which confirmed our approach of using ecotype-specifc crosses, to be a useful method for analyzing the maternal and paternal methylomes of the *Arabidopsis* endosperm (Fig. 3.3, A – F).

Similar to the nonallele-specific data in Chapter II, the allele-specific data showed the wild type endosperm (Fig. 3.1, A, C and E, red trace-maternal, blue trace-paternal) to be less methylated the embryo (Fig. 3.1, A, C and E, yellow trace-maternal, turquoise trace-paternal) in all DNA methylation contexts. However, in terms of allele-specific CG methylation, the maternal genome of the wild type endosperm particularly at transposable elements, and somewhat at genes, (Fig. 3.1, A, red trace) was hypomethylated compared to the paternal genome (Fig. 3.1, A, blue trace). In the *dme*-mutant background, DNA methylation in the CG context became elevated on the maternal genome (Fig. 3.1, A, green trace) suggesting the wild type hypomethylation status was caused by DME-mediated DNA demethylation in the central cell. The maternal genome (Fig. 3.1, A, green trace) became hypermethylated to the same level as the paternal genome (Fig. 3.1, A, black trace) in the *dme*-mutant background.

In terms of non-CG methylation (Fig. 3.1, C and E), both wild type maternal (red trace) and paternal (blue trace) alleles had the same methylation levels, but together were uniformly reduced for CHG and CHH methylation in the *dme*-mutant background (Fig. 3.1, C and D, green trace-maternal, black trace-paternal). In particular, allele-specific CHH methylation was almost completely gone (Fig. 3.1, E, green and black traces), an observation that was also seen in our nonallele-specific data (Fig. 2.3, C-D; Chapter 2).

DME-mediated DNA demethylation occurs in all methylation contexts

A biochemical analysis of DEMETER activity shows it actively removes 5mC from all three DNA methylation contexts (unpublished). Hence, we expect loci directly targeted by DME-mediated DNA demethylation to have DNA methylation removed from all neighboring methylated cytosines, regardless of context. To test if the putative maternal endosperm hypomethylated CG DMRs were DME-dependent, we examined the methylation status of their neighboring CHG and CHH sites.

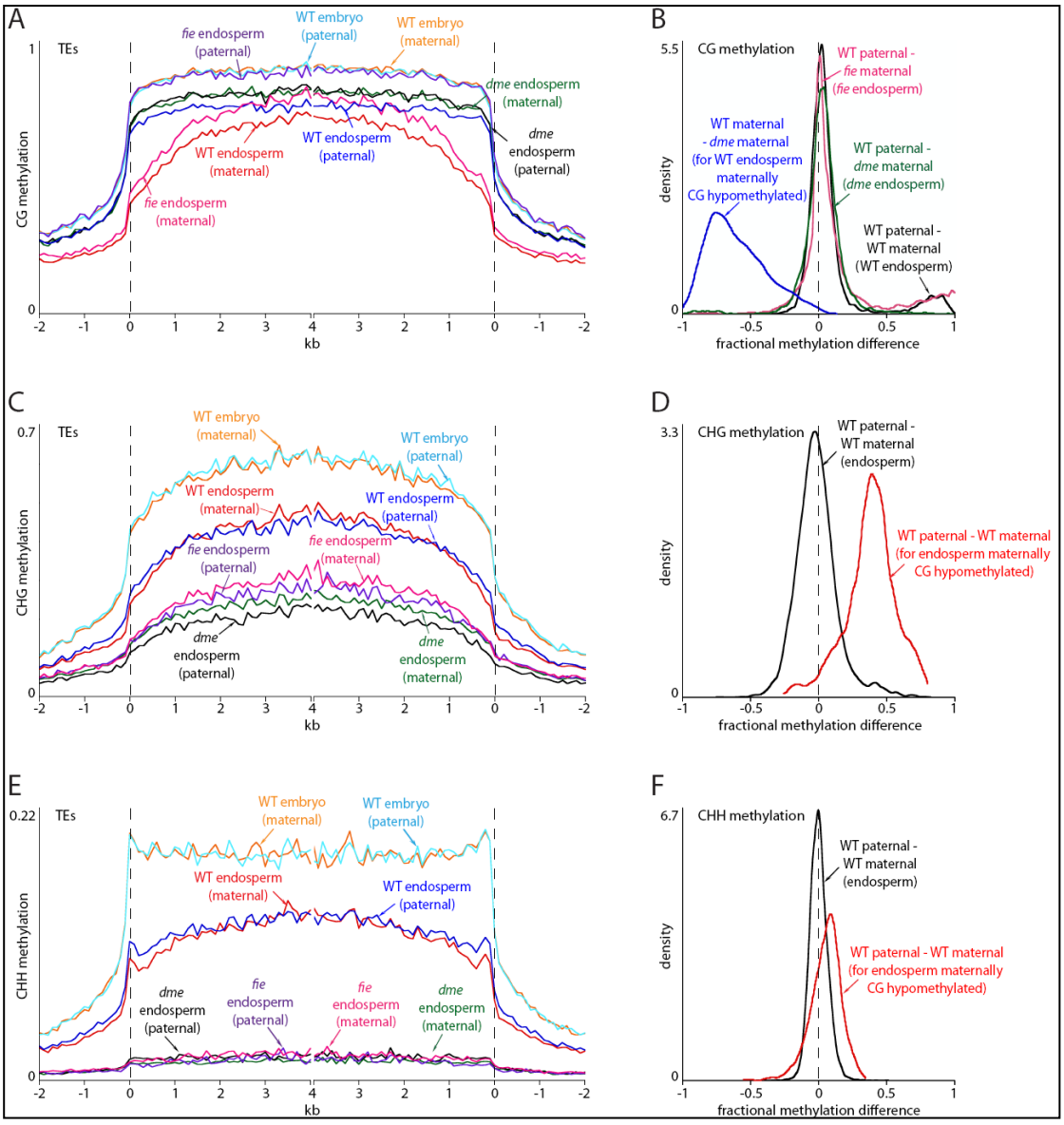


Fig. 3. 1 Allele-specific profiles of *A. thaliana* endosperm and embryo DNA methylation. (A, C, E) Transposons were aligned at the 5' and 3' ends (dashed lines) and average methylation levels for each 100-bp interval are plotted. (B, D, F) Kernel density plots of endosperm methylation differences. A shift of the peak with respect to zero represents a global difference; shoulders represent local differences.

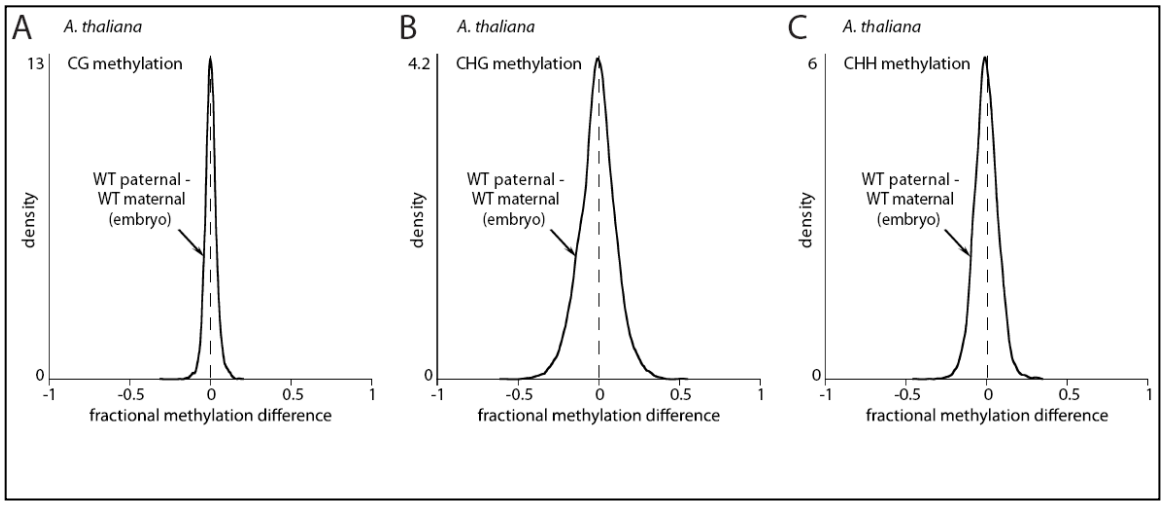


Fig. 3. 2. DNA methylation in *A. thaliana* embryo. (A-F) Kernel density plots, which have the effect of tracing the frequency distribution, of the differences between paternal and maternal embryo methylation in *A. thaliana* (A-C).

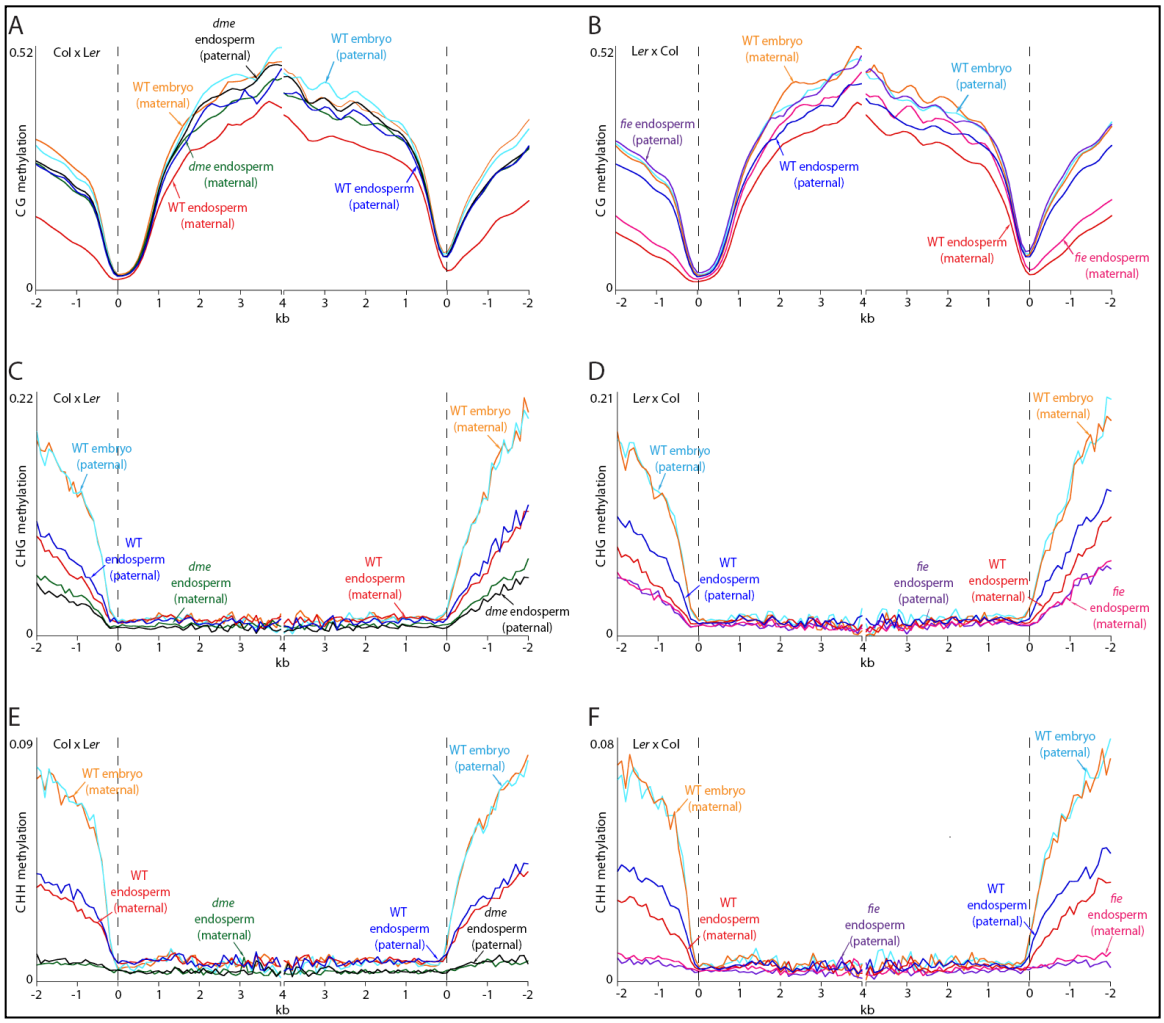


Fig. 3. 3. DNA methylation in *A. thaliana* genes. (A-F) *A. thaliana* genes were aligned at the 5' end (left panel) or the 3' end (right panel) and average methylation levels for each 100-bp interval were plotted for endosperm derived from Col x L*er* and *dme-2* (Col) x L*er* F1 seeds, and embryo derived from Col x L*er* seeds (A, C, E), and for endosperm derived from L*er* x Col and *fie-1* (L*er*) x Col F1 seeds, and embryo derived from L*er* x Col seeds (B, D, F). The dashed line at zero represents the point of alignment. CG methylation is shown in (A-B), CHG in (C-D), CHH in (E-F).

This was done by identifying the genomic coordinates of the maternal hypomethylated CG DMRs (fractional methylation value above 0.5) in each 50 bp window, and then examining the methylation status of the CHG and CHH sites within the window. The resulting non-CG methylation DMRs were then plotted out on a kernel density graph. If the putative maternal endosperm hypomethylated CG DMRs were DME-targets their neighboring CHG and CHH DMRs should also be hypomethylated as well, and when plotted on a kernel density graph, would generate positive values (curve shifted to the right) on the maternal subtracted from paternal endosperm data set. Likewise, non-DME targets would lack non-CG hypomethylation, corresponding to a curve either centered on zero or be negative.

As shown in Figure 3.1 (D, red trace), the bulk of the CHG sites neighboring the maternal endosperm CG hypomethylated loci were also hypomethylated, as the curve was shifted to the right, and centered on a fractional methylation difference of ~0.5. Neighboring CHH sites appeared hypomethylated as well (Fig. 3.1, F, red trace), although not as much as the CHG sites (graph closer to zero and centered on fractional methylation value of 0.1). The reduced number of neighboring CHH hypomethylated DMRs is reflective of the transient nature of siRNA-mediated DNA demethylation.

DEMETER activates PRC2 to promote non-CG methylation in the endosperm

Both our nonallele-specific and allele-specific methlyome studies have shown the *dme*-mutant to cause a reduction in non-CG methylation. A mutant phenotype which displays a reduction in DNA methylation, for an enzyme that functions to demethylate DNA, is counterintuitive to what one would expect, and suggests the presence of other factors exist between DEMETER and its ability to regulate non-CG methylation. One factor that regulates biological processes in the endosperm, which is dependent on DME-mediated DNA demethylation to initiate its formation, is PRC2. Thus, to further study the relationship between DME-mediated DNA demethylation and non-CG methylation in the endosperm, we choose to also carry out an allele-specific analysis of PRC2-mutant endosperm. We used a loss-of-function mutation in *FIE*, a major component of PRC2 which functions as the protein complex's central hub, holding the entire complex together. Hence, mutations in *FIE* obliterate PRC2 formation resulting in abnormal endosperm development, a mutant phenotype very similar to *dme*-mutants. The *fie*-mutant serves as an excellent way to analyze endosperm lacking a functional PRC2. Interestingly, non-CG methylation was reduced in the *fie*-endosperm background (Fig. 3.1, C and E, pink trace-maternal, purple trace-paternal), appearing similar to the *dme*-mutant endosperm (Fig. 3.1, C and E, green trace-maternal, turquoise trace-paternal), suggesting that in wild type endosperm, the PRC2 is somehow involved in up-regulating non-CG methylation.

PRC2 has no effect on CG methylation in the endosperm

As shown in Fig. 3.1 A, in the *fie*-endosperm background, the maternal genome (pink trace) was hypomethylated compared to the paternal genome (purple trace) in the CG context, appearing similar, although slightly higher, to that of the maternal wild type endosperm (Fig. 3.1, A, pink trace). The allele-specific kernel density plot generated from *fie*-endosperm in the CG context (Fig. 3.1, B, pink trace) followed a similar methylation profile as the wild type, and also had a subpopulation of maternally hypomethylated loci (see arrow).

Analysis of the vegetative and sperm cell methylomes

To examine the methylation profiles of the specific cells of the male gametophyte, vegetative cell nuclei and sperm cells were extracted from mature wild type and *dme*-mutant pollen and separated using FACS (Schoft et al., 2011). Genomic DNA libraries were then made and bisulfite converted, following high-throughput DNA methylation analysis using the Illumina Hi-seq™ 2000 platform as previously mentioned (Hsieh et al., 2009; Lister et al., 2008). The resulting methylation data was then investigated using Ends Analysis and kernel density plot examination (Fig. 3.4).

As shown in Figure 3.4 (B, black trace), kernel density plot analysis of CG methylation generated from the difference between wild type sperm and vegetative cell showed that the entire curve was centered at zero. Thus, both the sperm and vegetative cell were similarly methylated on a genome-wide level. However, a distinct subpopulation of vegetative cell hypomethylated loci exhibited a fractional methylation difference above 0.5 (Fig. 3.4, B, black arrow). Notably, these loci had a similar profile to the maternal endosperm hypomethylated loci in Figure 3.1 (B, black arrow). In the *dme*-mutant background (Fig. 3.4, B, green trace) these hypomethylated loci were lost, suggesting they maybe DEMETER targets. This finding hints at the notion that DME may be targeting similar loci in both the male and female gametophyte companion cells. Interestingly, methylation in the non-CG context showed the overall vegetative cell genome to be more methylated than the sperm cell genome (Fig. 3.4, D and F). Both CHG and CHH methylation kernel density plots had peaks that were shifted to the left of zero (the mean fractional methylation for CHG was -0.2, and for CHH was -3.5). This finding was somewhat unexpected, as the proposed model suggests the vegetative cell genome to be less methylated than that of the sperm cell, due a down-regulation of the global transposon repressor DDM1 and heterochromatin-inducing siRNA machinery (Slotkin et al., 2009). A scenario such as this, suggests the vegetative cell genome to be reduced in DNA methylation, at least comparably to the sperm cell.

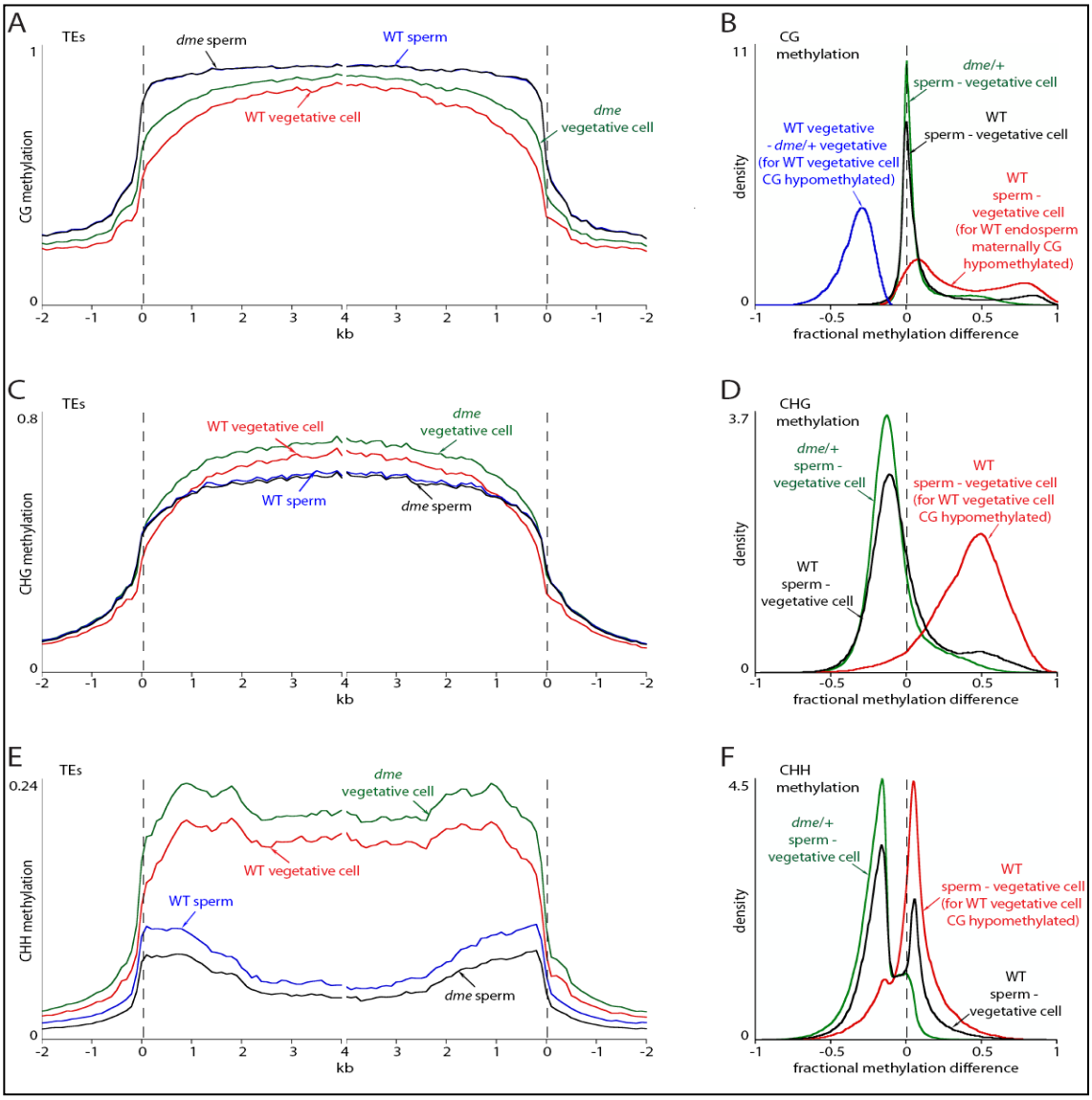


Fig. 3. 4. Profiles of A. thaliana sperm and vegetative cell DNA methylation. (A, C, E) Average methylation in transposons was plotted as in Fig. 1. (B, D, F) Kernel density plots of pollen methylation differences. A shift of the peak with respect to zero represents a global difference; shoulders represent local differences.

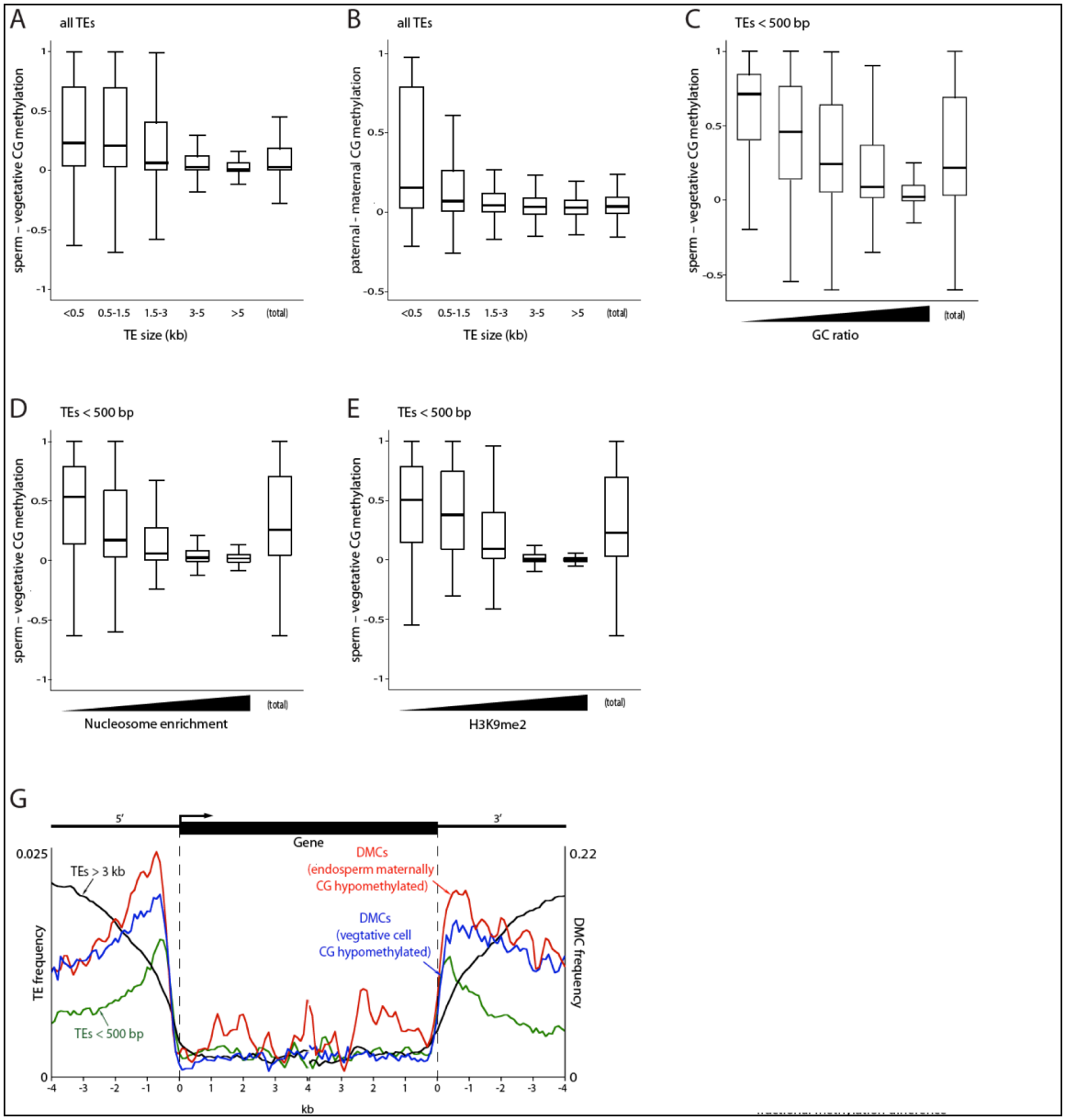


Fig. 3. 5. Demethylation of small TEs in vegetative cells reinforces methylation in sperm. (A-E) Box plots showing the extent of demethylation in transposons. Each box encloses the middle 50% of the distribution, with the horizontal line marking the median, and vertical lines marking the minimum and maximum values that fall within 1.5 times the height of the box. (G) Distribution of transposable elements and significantly differentially methylated cytosines (DMCs; CG context; p-value < 0.0001, Fisher's exact test) with respect to genes.

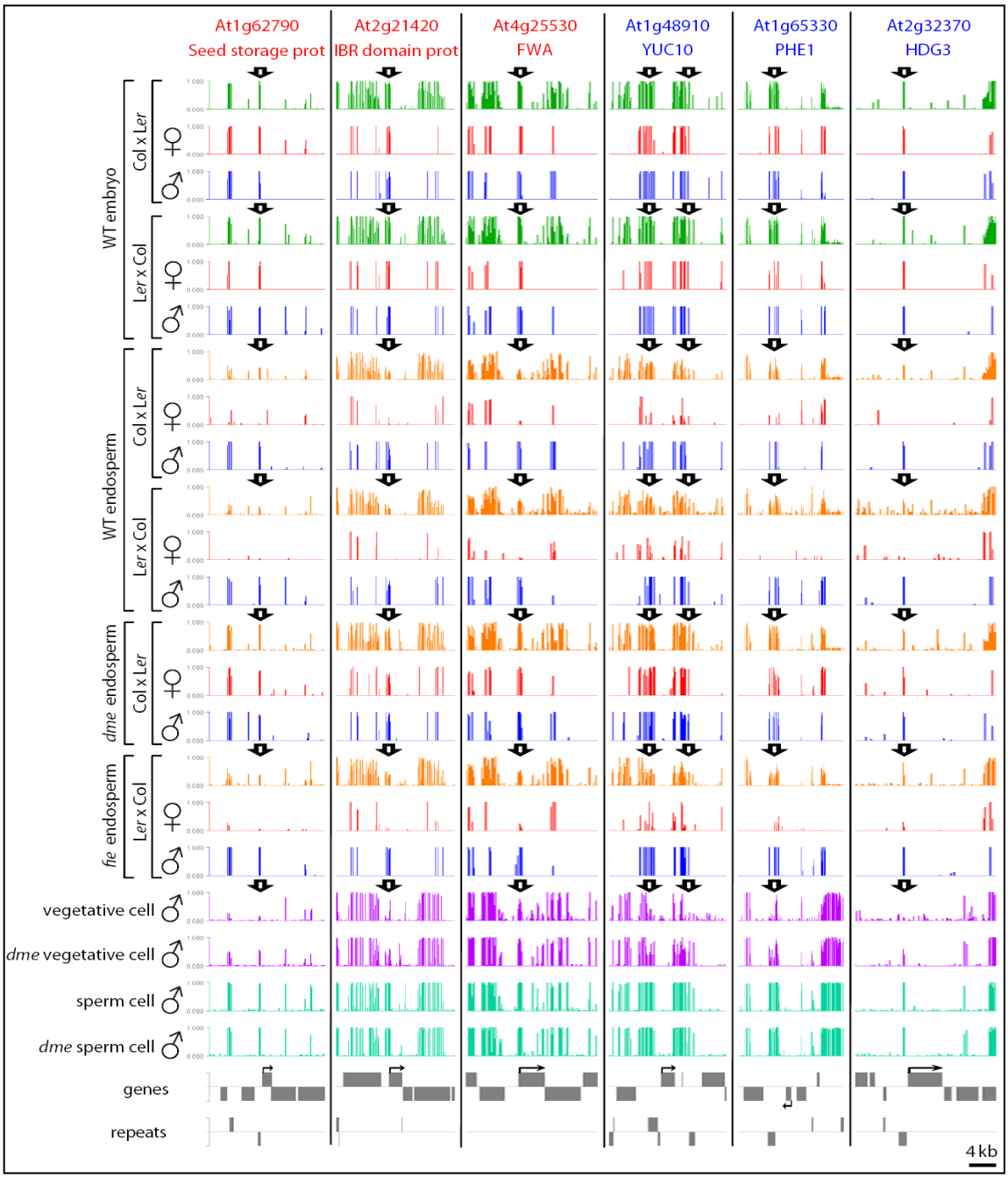


Fig. 3. 6. CG methylation near *A. thaliana* imprinted genes. Snapshots of CG methylation in indicated tissues near maternally expressed (red) and paternally expressed (blue) *A. thaliana* imprinted genes. Green bars represent embryo methylation, orange bars represent endosperm methylation, and red and blue bars represent methylation of the maternal and paternal genomes, respectively. Arrows point out methylation changes in endosperm and vegetative cell.

Does DEMETER target the same loci in the central cell and vegetative cell for active DNA demethylation?

Our data revealed that both the maternal endosperm and the vegetative cell genomes have hypomethylated CG sites that appear to be DME-dependent (Fig. 3.1, B, black trace, and Fig. 3.4, B, black trace). This suggests that DEMETER may be targeting the same loci, but in different cell types. Recent studies have confirmed DME to be expressed and active in the vegetative cell, resulting in the active demethylation of the imprinted *FWA* and *MEA* genes as well as their expression (Schoft et al., 2011). To find out if DEMETER could be targeting the same loci in both the central cell and vegetative cells, we compared the maternal endosperm hypomethylated loci to the vegetative cell hypomethylated loci to see how many of the hypomethylated sites were in fact the same. This was done by checking the methylation status of the maternal endosperm hypomethylated loci in the sperm minus vegetative cell data set, and then plotting out the resulting 'mutually' hypomethylated loci. If the maternal endosperm hypomethylated loci were also hypomethylated in the vegetative cell, than they would register as positive values in the sperm minus vegetative cell background, and when plotted would produce a curve shifted to the right. As shown in Figure 3.4 (B, red trace), the majority of the maternal endosperm hypomethylated loci, appeared to also be hypomethylated in the vegetative cell as demonstrated by the curve being shifted to the right of zero.

Where in the genome are the DME-dependent hypomethylated loci located?

DEMETER and other members of the DML-family have been shown to act at discreet locations within the genome, by specifically targeting the 5' and 3' ends of genes for active DNA demethylation. In the case of DME and its activity in the central cell, the gene-adjacent demethylation leads to genome imprinting (Gehring et al., 2006; Huh et al., 2008). Moreover, many of demethylated sites in *Arabidopsis* are transposons and repetitive elements, such as the repeats found at the 3' end of *FWA*, and three regions next to *MEA* consisting of a helitron DNA transposon element, an RC/helitron (AtREP2), and seven ~182bp direct repeats (labeled the MEA ISR region). In rice, gene adjacent regions seem to be enriched with small transposons in the form of miniature inverted-repeats transposable elements (MITS) (ave length ~189 bp) and short interspersed elements (SINEs) (ave length ~500 bp), while regions more distant from the gene favor larger transposon elements such as long terminal repeats (LTRs) (ave length ~855 bp).

To further investigate the genomic location of the maternal endosperm and vegetative cell DME-dependent hypomethylated loci, we examined their distribution amongst transposons according to transposon size. Figure 3.5 (A and B) shows the resulting box-whisker plots, where the x-axis represents transposons grouped according to increasing size (<0.5 bp, 0.5-1.5 bp, 1.5-3.0 bp, 3-5 bp, >5.0 bp, and 'total' TAIR9 TEs) and the y-axis represents the number

of DME-dependent hypomethylated loci. Both the vegetative cell hypomethylated loci (Fig. 3.5, A) and the maternal endosperm hypomethylated loci (Fig. 3.5, B) are more abundant in smaller transposons than larger ones. Interestingly, these smaller transposons in *Arabidopsis* have chromatin-related features distinct from larger ones. Small transposons tend to be lower in GC content than larger ones (Fig. 3.5, C) (and are therefore higher in AT content). Moreover, they appear to also have low nucleosome occupancy (Fig. 3.5, D), and low enrichment for H3K9me2 (Fig. 3.5, E) marks.

As previous demonstrated in rice (Zemach et al., 2010a), analysis of the distribution of transposons with respect to genes in *Arabidopsis* also showed smaller transposons (Fig. 3.5, G, green trace) to be much closer to genes than larger ones (Fig. 3.5, G, black trace). Moreover, these small transposons were also abundant in the DME-dependent maternal endosperm and vegetative cell hypomethylated loci (Fig. 3.5, G, red and blue traces). A genome-browser view show several of the DME-dependent maternal endosperm and vegetative cell hypomethylated to loci next to known imprinted MEGs and PEGs (Fig. 3.6, arrows).

DISCUSSION

The purpose of this study was to examine the methylomes of the companion cells of the female and male gametes, which are the central cell and the vegetative cell respectively. To circumvent the inability to obtain *Arabidopsis* central cells, we used the maternal genome of the *Arabidopsis* endosperm as a proxy to the central cell genome. Obtaining the cells of the male gametophyte was more feasible and direct, as both the vegetative and sperm cells are amendable to FACS (Schoft et al., 2011).

DME targets loci on the maternal endosperm genome for active DNA demethylation

Our analysis of the maternal and paternal endosperm methylomes revealed a population of significantly hypomethylated loci on the maternal genome compared to paternal genome. Importantly, these hypomethylated loci appeared to be DME-dependent, as they became hypermethylated in the *dme*-mutant endosperm background. However, both maternal and paternal genomes had essentially the same global DNA methylation profiles. This finding sheds light on a particular aspect of the DME DNA glycosylase, which is its targeting range, suggesting DME to demethylate in a more narrow and restricted manner, rather than that of a broad one.

DME targets the same loci in the companion cells for active DNA demethylation

The DEMETER DNA glycosylase has been shown to be active in both accessory cells of the female and male gametophytes, but the extent of its activity in these cells remains to be fully understood. DME-mediated DNA demethylation in the central cell establishes genomic imprinting and is required for proper endosperm development, whereas in the vegetative cell, DEMETER activity is required for proper germination and pollen tube formation (Huh et al., 2008). The cell-specific presence of DME in both of these gametophytic cells, along with the dramatic mutant phenotypes produced in its absence, suggests DME to play a more fundamental role in reproduction than previously thought. Corroborating this notion is the finding of the MEGs, *FWA* and *MEA,* being subject to DME-mediated DNA demethylation in the vegetative cell resulting in expression (Schoft et al., 2011). Thus, a culmination of work done prior this study hinted at the idea of DEMETER potentially demethylating the same loci in both the central cell and vegetative cell.

A critical component to this analysis has been the use of the *dme*-mutant endosperm and pollen (*DME/dme*) tissue, which when coupled to the genomics approach enabled us to gain a better understanding the normal function of DME and the loci it potentially targets. Additionally, the use of kernel density plots has proved to be particularly helpful by being able to reveal both global and local methylation trends.

Our high throughput epigenomics approach suggests the DEMETER DNA gylcosylase targets the same loci in both the central cell and vegetative cell for active DNA demethylation. As shown in Figure 3.6 (arrows), these loci were characterized by being hypomethylated on the maternal genome compared to the paternal genome of the endosperm, becoming maternally hypermethylated in the *dme*-endosperm background. Similarly, these same loci were hypomethylated on the vegetative cell genome compared to the sperm cell genome, and in the *dme*-pollen became hypermethylated on the vegetative cell. Additional evidence suggesting these loci to indeed be targeted by DEMETER in both the central cell and vegetative cell, is their maternal and paternal hypermethylated status in the embryo genome, which is a tissues that lacks DME expression. The ability to examine the same loci in several different tissues, mutants, and reciprocal crosses, allowed for a comprehensive examination of the data resulting in a more accurate identification of loci that were truly targeted by DME-mediated DNA demethylation.

A mechanism for DME-mediated gene adjacent demethylation

We found the DME-dependent mutually hypomethylated loci of the companion cells to be enriched in small transposable elements, that tended to be less than 500 bp in size. Moreover, these small transposons were also shown to occur frequently next to the 5’ and 3’ ends of genes. As a result of these two

findings, we were able to show a correlation between the DME-dependent hypomethylated loci of the companion cells and their proximity to the 5' and 3' ends of genes. In addition to this finding, is our previous model suggesting the direct targeting of transposable elements by DME-mediated DNA demethylation in the central cell, resulting in transposon expression and the subsequent biogenesis of siRNAs for the reinforced silencing of transposons in the egg cell. Thus, the work done in Chapter 2 combined with the work in Chapter 3 allows us to form a model whereby DME in the central cell and vegetative cell targets transposable elements for active DNA demethylation, and because of their proximity to genes, consequently carries out gene adjacent DNA demethylation. In a subtle way, our model also suggests the gene adjacent DNA demethylation observed for the last decade (Gehring et al., 2006; Huh et al., 2008; Kinoshita et al., 2004), has more to do with DME 'chasing down' transposons to demethylate them, than with the specific targeting of the genes themselves for active demethylation.

METHODS

Isolation of Arabidopsis thaliana endosperm

Stage 12-13 flower buds were emasculated and pollinated 48 hours later. Reciprocal crosses were performed using wild-type Col-0 and Ler ecotypes. In addition, dme-2 (Col-gl) (Choi et al., 2002) heterozygous flowers were pollinated with wild-type (Ler) pollen, and fie-1 (Ler) (Ohad et al., 1999) heterozygous flowers were pollinated with wild-type (Col-0) pollen. For wild-type crosses, seven to eight days after pollination, F1 seeds (torpedo-stage to early-bent-cotyledon stage) were immersed in 0.3 M sorbitol and 5 mM Mes (pH 5.7) on a slide under a dissecting microscope. Embryo and endosperm were dissected using a fine needle and forceps. The seed coat was discarded. Wild-type embryos were twice centrifuged and the pellet resuspended in 0.3 M sorbitol and 5 mM Mes (pH 5.7) to remove contaminating endosperm. For crosses with the dme-2 or fie-1 mutations, F1 aborting seeds were identified and mutant endosperm was isolated. Approximately 500 wild-type endosperm, 1000 dme-2 or fie-1 endosperm, and 300 wild-type embryos were collected.

Isolation of vegetative nuclei and sperm cells. Pollen was isolated from wild-type (Col-0), dme-2 heterozygous (Col-gl ecotype), and dme-6 heterozygous flowers (Col-0 ecotype) as described previously (Schoft et al., 2011; Schoft et al., 2009). Vegetative cell nuclei and sperm cells were extracted from mature pollen and fractionated by fluorescence activated cell sorting as described previously (Schoft et al., 2011; Schoft et al., 2009).

Bisulfite sequencing library construction

As described previously, genomic DNA was isolated from vegetative nuclei and sperm cells (Schoft et al., 2011), endosperm, and embryo (Hsieh et al., 2009). Paired-end bisulfite sequencing libraries for Illumina sequencing were constructed as described previously (Hsieh et al., 2009) with minor modifications. In brief, about 150 ng of genomic DNA was fragmented by sonication, end repaired and ligated to custom-synthesized methylated adapters (Eurofins MWG Operon) according to the manufacturer's (Illumina) instructions for gDNA library construction. Adaptor-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (Qiagen) as outlined in the manufacturer's instructions. One quarter of the bisulfite-converted libraries was PCR amplified using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 µl of 10X Extaq reaction buffer, 25 µM dNTPs, 1 µl Primer 1.1, 1 µl Primer 2.1 (50 µl final). PCR reactions were carried out as follows: 95 ºC 3 min, then 12-14 cycles of 95 ºC 30 sec, 65 ºC 30 sec and 72 ºC 60 sec. The enriched libraries were purified twice with solid phase reverse immobilization (SPRI) method using AM-Pure beads (Beckman Coulter) prior to quantification with a Bioanalyzer (Agilent). Sequencing on the Illumina platform was performed at the Vincent J. Coates Genomic Sequencing Laboratory at UC Berkeley and the Genome Center at UC Davis.

Allele-specific determination of DNA methylation

Reads were sorted to the Col and Ler genomes (A. thaliana) as described (Hsieh et al., 2011). DNA methylation of cytosines within sorted reads was calculated as described (Hsieh et al., 2009; Zemach et al., 2010b).

Kernel density plots

All DNA methylation kernel density plots compare fractional methylation within 50 bp windows. For rice, we used windows with at least 5 informative sequenced cytosines and fractional methylation of at least 0.7 (CG), 0.4 (CHG), or 0.025 (CHH) in at least one of the samples being compared. For A. thaliana, we used windows with at least 20 informative sequenced cytosines (10 for CHG in endosperm) and fractional methylation of at least 0.7 (CG), 0.4 (CHG endosperm), 0.5 (CHG pollen), 0.08 (CHH endosperm), 0.15 (CHH vegetative cell), or 0.05 (CHH sperm) in at least one of the samples being compared. For parent-of-origin endosperm plots, only windows with methylation differences between Col and Ler below 0.1 (CG), 0.15 (CHG), or 0.05 (CHH) were used to exclude ecotype-specific differences. Windows in which fractional paternal (or sperm) CG methylation exceeded fractional maternal (or vegetative cell) CG methylation by at least 0.4 were considered demethylated.

Box-whisker plots

Box-whisker plots compare fractional methylation in TEs within 50 bp windows and use the same cutoffs as density plots. To examine the correlation between CG demethylation and chromatin structure, TE windows are separated into five groups in ascending order according to GC ratio, nucleosome enrichment, dimethylation of lysine 9 of histone H3 (H3K9me2), and distance from the closest edge of the TE, respectively. The five groups are as follows: for GC ratio, 0.2-0.3, 0.3-0.35, 0.35-0.4, 0.4-0.45, and >0.45; for nucleosome enrichment, 1-8, 9-16, 17-24, 25-39, and >39 counts (Chodavarapu et al., 2010); for H3K9me2, <0.5, 0.5-1.25, 1.25-2, 2-3, and >3 log2 (IP/input) (Bernatavichute et al., 2008); for distance to TE edge, <50 bp, 50-100 bp, 100-150 bp, 150-200 bp, and 200-250 bp.

Chapter IV

Conclusion

The overall aim of the work carried out in this dissertation was to study the epigenetic landscape of the *Arabidopsis* endosperm. More specifically, it was to understand the role of the DEMETER DNA glycosylase and its ability to actively demethylate DNA within the endosperm and central cell genome. Our rationale for studying DME-mediated DNA demethylation was based on its importance in plant development, particularly in reproduction, as seeds inheriting a maternal *dme*-mutant allele produce abnormal endosperm and aborted embryos. Similarly, DME is critical in *Arabidopsis* imprinting, where it actively removes the repressive methyation marks on the maternal alleles of several genes (i.e. *FWA, MEA, FIS2*) resulting in their imprinted status in the endosperm (Gehring et al., 2006; Huh et al., 2007; Kinoshita et al., 2004).

We took an epigenetics approach examining DME-mediated DNA demethylation by initially obtaining the endosperm methylome (chapter II), and then its specific maternal and paternal methylomes (chapter III). This was done by isolating endosperm tissue via hand-dissection, followed by genomic DNA extraction. The endosperm genomic DNA was then deep-sequenced using the Illumina™ platform. This approach for obtaining methylomes (Methyl Seq) has become a standard method for analyzing large-scale methylome studies (Cokus et al., 2008; Lister et al., 2008). Our interest in obtaining the entire methylome of the endosperm, and eventually the maternal and paternal methylomes, was due to the nature of DME as a DNA demethylase. Indeed, a methylomics approach seemed an almost perfect strategy for examining the effects of an enzyme whose essential function is to remove methylation from the genome.

Our initial non-allele specific study of the endosperm methlyome showed the endosperm to be hypomethylated in all contexts compared to the embryo (Hsieh et al., 2009). Originally, we speculated the genome-wide hypomethylation in the endosperm was directly due to DME-mediated DNA demethylation, since DME has been shown to remove DNA methylation from all contexts. However, in the *dme*-mutant background only a partial elevation in methylation was observed, which unexpectedly appeared to be context-specific occurring only in CG context. In the case of non-CG methylation, both CHG and CHH methylation showed a remarkable drop in methylation in the *dme*-mutant background, an observation hinting that DME maybe regulating non-CG methylation differently than CG methylation.

We suspected the genome-wide DNA demethylation occurring in the endosperm to be the result of an indirect association with DME-mediated DNA demethylation, and assumed the DME DNA glycosylase to function in a more restricted, rather than global manner. Our premise for this was based on work done by us (Hsieh et al., 2011) and others (Jullien et al., 2008), suggesting the pre-fertilized central cell genome lacks the proteins required for maintenance methylation. The hypomethylated endosperm genome could initially arise in the central cell from an RBR1/MS1-mediated down-regulation of MET1 maintenance methyltransferase (Jullien et al., 2008), and/or a PRC2-mediated down-regulation of VIM5 (Hsieh et al., 2011). Fertilization would then occur, adding the paternal genome to the central cell following endosperm proliferation. The lack of DNA methylation machinery from the (previous) central cell would affect both the

maternal and paternal genomes of the nascent endosperm in a uniform manner, resulting in both being globally similar in DNA methylation. In either case, both of these independent studies suggest the central cell is deficient in the critical components of the DNA maintenance methylation machinery, and it is possible that what we are observing as genome-wide DNA demethylation in the endosperm is the remnants of this deficiency.

Our interest in understanding the effects caused by DME ultimately bring us to the central cell, a cell currently unobtainable for isolation and consequently genomic experiments. However, we surmised the maternal genome of the endosperm would provide us with sufficient insight into the DNA methylation landscape of the central cell, since the maternal endosperm genome was previously the central cell genome before fertilization. Thus, we carried out an allele-specific methylome study of the endosperm genome in order to further study the effects of the DME DNA glycosylase in the central cell, in addition to our nonallele-specific study. Our initial examination of the allele-specific endosperm data showed both maternal and paternal genomes to be similarly methylated on a global level. However, a population of maternally hypomethylated loci were present displaying significant hypomethylation, suggesting they were targets of DME-mediated DNA demethylation. Indeed, these hypomethylated loci were completely lost in the *dme*-mutant endosperm background confirming their origins were DME-dependent. We explored the distribution of the maternal endosperm hypomethylated loci, finding them highly enriched in transposable elements, in particular smaller ones less than 500 bp in size. Work done by us examining the distribution of transposable elements in *Arabidopsis* (unpublished) show small transposons to be frequently found right-up close next to genes. Indeed, an over lay of the maternal endosperm hypomethylated sites with small transposons next to genes showed enrichment of hypomethylated sites, and the emergence of ‘5 and ‘3 gene-adjacent demethylation. Importantly, this enrichment of maternal endosperm hypomethylated loci was lost when an over lay of the hypomethyated loci with larger transposons was examined. Hence, our examination of the maternal endosperm methylome revealed a phenomenon of DME-mediated gene-adjacent DNA demethylation, giving us new insights into DME’s function and role in central cell and endosperm development.

Excitingly, we also discovered DME-mediated hypomethylated loci in the vegetative cell genome in the male gametophyte. Moreover, the hypomethylated loci on the vegetative cell genome, were the same as those on the maternal endosperm genome. Thus, we found the DME DNA glycosylase to target similar loci in both the central cell and vegetative cell for active DNA demethylation. This finding challenges the traditional understanding of DME, which sees DME as an enzyme mainly active in the female gametophyte, and particularly focused on imprinting. The notion of DME acting in both companion cells, and at the same places in the genome, suggests DME’s function to be more ancient and primitive then previously thought.

Our methylome studies of the endosperm show the gene-adjacent demethylation carried out by DME, correlating with the presence of small

transposable elements residing there. The active removal of the repressive methylation mark next to genes has been described as DME's modus operandi causing mono-allelic expression and the eventual imprinting of the gene. However, our findings showing DME targeting small transposons that frequently appear next to genes, suggests DME is drawn to transposons and not genes. This suggests the DME-mediated expression of a gene may arise more as a by-product of DME targeting a gene-adjacent transposon, than of DME explicitly targeting the gene for demethylation. Hence, it is possible that what has been witnessed for the last decade as imprinting in *Arabidopsis* is more of an indirect outcome of DME targeting transposons than an innate drive to demethylate particular genes.

The complete loss of non-CG methylation in the *dme*-mutant endosperm suggested DME to play a role in increasing the activity of the RdDM pathway. We propose a model where DME demethylates transposons in the central cell, resulting in their expression and the eventual biogenesis of siRNAs. These siRNA are then incorporated into the RdDM pathway in the central cell, and are also shuttled to the neighboring egg cell, where they carry out RdDM as well. We believe the reinforced transposon silencing in the egg cell, due to the DME-mediated proliferation of transposon expression in the central cell, exemplifies the nature of the companion cell in its evolutionary drive to nurture and protect the generative germ cell.

References:

Allis, C.D., Jenuwein, T., and Reinberg, D. (2007). Epigenetics, 1st edn (Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press).

Arnaud, P., and Feil, R. (2006). MEDEA takes control of its own imprinting. Cell *124*, 468-470.

Barlow, D.P. (2011). Genomic Imprinting: A Mammalian Epigenetic Discovery Model. Annual Review of Genetics *45*, 379-403.

Baroux, C.l., Pecinka, A., Fuchs, J.r., Schubert, I., and Grossniklaus, U. (2007). The Triploid Endosperm Genome of Arabidopsis Adopts a Peculiar, Parental-Dosage-Dependent Chromatin Organization. The Plant Cell Online *19*, 1782-1794.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F.*, et al.* (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. Nature *445*, 671-675.

Bartee, L., Malagnac, F., and Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev *15*, 1753-1758.

Bastock, R., and Strutt, D. (2007). The planar polarity pathway promotes coordinated cell migration during Drosophila oogenesis. Development *134*, 3055-3064.

Bauer, M.J., and Fischer, R.L. (2011). Genome demethylation and imprinting in the endosperm. Curr Opin Plant Biol *14*, 162-167.

Berger, F. (1999). Endosperm development. Curr Opin in Plant Biol *2*, 28-32.

Berger, F., Fitz Gerald, J.N., and Ingouff, M. (2007). Arabidopsis as a model for undertanding the basics of endosperm development. In Endosperm: Developmental and Molecular Biology, O.-A. Olsen, ed. (Berlin, Springer-Verlag), pp. 91-110.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA Demethylation Dynamics. Cell *146*, 866-872.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D.b., Xu, G.-L., Lin, C.-S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. Science *294*, 2536-2539.

Bourc,Äôhis, D.b., and Voinnet, O. (2010). A Small-RNA Perspective on Gametogenesis, Fertilization, and Early Zygotic Development. Science *330*, 617-622.

Brown, R.C., and Lemmon, B.E. (2007). The developmental biology of cereal endosperm. In Endosperm: Developmental and Molecular Biology, O.-A. Olsen, ed. (Berlin, Springer-Verlag), pp. 1-20.
Cedar, H., and Bergman, Y. (2012). Programming of DNA Methylation Patterns. Annual Review of Biochemistry, *81*.
Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet *6*, 351-360.
Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J.*, et al.* (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388-392.
Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in *Arabidopsis*. Cell *110*, 33-42.
Cokus, S.J., Fing, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.
Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res *36*, e112, 1-11.
Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet *12*, 19-31.
De Smet, I., Lau, S., Mayer, U., and Jürgens, G. (2010). Embryogenesis - the humble beginnings of plant life. The Plant Journal *61*, 959-970.
Drews (2011). The Female Gametophyte. The Arabidopsis Book.
Eissenberg, J.C. (2001). Molecular biology of the chromo domain: an ancient chromatin module comes of age. Gene *275*, 19-29.
Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E.*, et al.* (2010a). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.
Feng, S., Jacobsen, S.E., and Reik, W. (2010b). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.
Franke, A., DeCamillis, M., Zink, D., Cheng, N., Brock, H.j., and Paro, R. (1992). Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. EMBO J *11*, 2941-2950.
Frost, J.M., Monk, D., Moschidou, D., Guillot, P.V., Stanier, P., Minger, S.L., Fisk, N.M., Moore, H.D., and Moore, G.E. (2011). The effects of culture on genomic imprinting profiles in human embryonic and fetal mesenchymal stem cells. Epigenetics *6*, 52-62.
Frost, J.M., Monk, D., Stojilkovic-Mikic, T., Woodfine, K., Chitty, L.S., Murrell, A., Stanier, P., and Moore, G.E. (2010). Evaluation of Allelic Expression of Imprinted Genes in Adult Human Blood. PLoS One *5*, e13556.

Gehring, M., Bubb, K.L., and Henikoff, S. (2009a). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science *324*, 1447-1451.
Gehring, M., and Henikoff, S. (2007). DNA methylation dynamics in plant genomes. Biochimica et Biophysica Acta *1769*, 276-286.
Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.
Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PLoS One *6*, e23687.
Gehring, M., Reik, W., and Henikoff, S. (2009b). DNA Demethylation by DNA Repair. Trends in Genetics *25*, 82-90.
Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. (1997). A polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. Nature *386*, 44-51.
Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science *329*, 643-648.
Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a polycomb-group gene in *Arabidopsis*. Science *280*, 446-450.
Haig, D. (2000). The Kinship Theory of Genomic Imprinting. Annual Review of Ecology and Systematics *31*, 9-32.
Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. American Naturalist *134*, 147-155.
Hayashi, K., and Surani, M.A. (2009). Resetting the Epigenome beyond Pluripotency in the Germline. Cell Stem Cell *4*, 493-498.
He, X.-J., Chen, T., and Zhu, J.-K. (2011a). Regulation and function of DNA methylation in plants and animals. Cell Res *21*, 442-465.
He, X.-J., Chen, T., and Zhu, J.-K. (2011b). Regulation and function of DNA methylation in plants and animals. Cell Research *21*, 442-465.
He, X.-J., Hsu, Y.-F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.-L., Wang, C.-S., Jin, H., and Zhu, J.-K. (2009). An Effector of RNA-Directed DNA Methylation in Arabidopsis Is an ARGONAUTE 4- and RNA-Binding Protein. Cell *137*, 498-508.
Hemberger, M., Dean, W., and Reik, W. (2009). Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. Nat Rev Mol Cell Biol *10*, 526-537.
Henderson, I.R., and Jacobsen, S.E. (2007). Epigenetic inheritance in plants. Nature *447*, 418-424.
Hsieh, T.-F., Ibarra, A.C., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L., and Zilberman, D. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. Science *324*, 1451-1454.
Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D.*, et al.* (2011). Regulation of

imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A *108*, 1755-1762.
Huh, J.H., Bauer, M.J., Hsieh, T.-F., and Fischer, R.L. (2007). Endosperm gene imprinting and seed development. Curr Opin in Genet and Devel *17*, 480-485.
Huh, J.H., Bauer, M.J., Hsieh, T.-F., and Fischer, R.L. (2008). Cellular Programming of Plant Gene Imprinting. Cell *132*, 735-744.
Ikeda, Y. (2012). Plant Imprinted Genes Identified by Genome-wide Approaches and Their Regulatory Mechanisms. Plant and Cell Physiology 53 (5), 809-816.
Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science *333*, 1300-1303.
Iwano, M., Shiba, H., Miwa, T., Che, F.-S., Takayama, S., Nagai, T., Miyawaki, A., and Isogai, A. (2004). Ca2+ Dynamics in a Pollen Grain and Papilla Cell during Pollination of Arabidopsis. Plant Physiology *136*, 3562-3571.
Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle *8*, 1698-1710.
Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. Chromosoma *121*, 1-20.
Jahnke, S., and Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol *19*, 1677-1681.
Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature *449*, 248-251.
Jiang, H., and Wong, W.H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics *15*, 2395-2396.
Jullien, P.E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., and Berger, F. (2008). Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol *6*, e194.
Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr Biol *13*, 421-426.
Kawashima, T., and Goldberg, R.B. (2010). The suspensor: not just suspending the embryo. Trends in Plant Science *15*, 23-30.
Kim, K.-P., Thurston, A., Mummery, C., Ward-van Oostwaard, D., Priddle, H., Allegrucci, C., Denning, C., and Young, L. (2007). Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. Genome Research *17*, 1731-1742.
Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science *303*, 521-523.
Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm. Plant Cell *11*, 1945-1952.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J., Koornneef, M., and Kakutani, T. (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. Plant J *49*, 38-45.
Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.
Kohler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. Genes and Development *17*, 1540-1553.
Kohler, C., Wolff, P., and Spillane, C. (2012). Epigenetic Mechanisms Underlying Genomic Imprinting in Plants. Annual Review of Plant Biology *63*, 331-352.
Krueger, F., Kreck, B., Franke, A., and Andrews, S.R. (2012). DNA methylome analysis using short bisulfite sequencing data. Nat Meth *9*, 145-151.
Lan, L., Lin, S., Zhang, S., and Cohen, R.S. (2010). Evidence for a Transport-Trap Mode of Drosophila melanogaster gurken mRNA Localization. PLoS One *5*, e15448.
Lauria, M., Rupe, M., Guo, M., Kranz, E., Pirona, R., Viotti, A., and Lund, G. (2004). Extensive Maternal DNA Hypomethylation in the Endosperm of Zea mays. The Plant Cell *16*, 510-522.
Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics *11*, 204-220.
Leyser, O., and Day, S. (2003). Mechanisms in Plant Development (Oxford, UK, Blackwell Science Ltd.).
Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science *292*, 2077-2080.
Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the Arabidopsis genome. Cell *133*, 395-397.
Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., No, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.
Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development, and function. Plant Cell *5*, 1383-1399.
Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol *21*, 367-376.
Matzke, M.A., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J.M. (2007). Targets of RNA-directed DNA methylation. Curr Opinion in Plant Biol *10*, 512-519.
McKeown, P.C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M.W., Donoghue, M.T., Fort, A., Duszynska, D., Comte, A., Lao, N.T.*, et al.* (2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. BMC Plant Biol *11*, 113.
McLaren, A., and Lawson, K.A. (2005). How is the mouse germ-cell lineage established? Differentiation *73*, 435-437.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, k.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.
Mok, Y.G., Uzawa, R., Lee, J., Weiner, G.M., Eichman, B.F., Fischer, R.L., and Huh, J.H. (2010). Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. Proc Natl Acad Sci U S A *107*, 19225-19230.
Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues. Journal of Biological Chemistry *279*, 52353-52360.
Nguyen, H., Brown, R.C., and Lemmon, B.E. (2000). The specialized chalazal endosperm in *Arabidopsis thaliana* and *Lepidium virginicum* (Brassicaceae). Protoplasma *212*, 99-110.
Niehrs, C., and Schafer, A. (2012). Active DNA demethylation by Gadd45 and DNA repair. Trends in Cell Biology *22*, 220-227.
Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Mutations in *FIE*, a WD polycomb group gene, allow endosperm development without fertilization. Plant Cell *11*, 407-415.
Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A.*, et al.* (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207-213.
Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell *99*, 247-257.
Olsen, O.-A. (2001). Endosperm development: Cellularization and cell fate specification. Annu. Rev. Plant Physiol. Plant Mol. Biol. 52, 233-67.
Olsen, O.-A. (2004). Nuclear endosperm development in cereals and Arabidopsis thaliana. Plant Cell *16*, S214-S227.
Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. Cell *133*, 1145-1148.
Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.
Palanivelu, R., and Tsukamoto, T. (2012). Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. Wiley Interdisciplinary Reviews: Developmental Biology *1*, 96-113.
Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*, 6752-6757.
Pirrotta, V. (1998). Polycombing the genome: PcG, trxG, and chromatin silencing. Cell *93*, 333-336.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.
Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences *97*, 5237-5242.
Schoft, V.K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.S., Park, G.T.*, et al.* (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A *108*, 8042-8047.
Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.
Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by Polycomb and Trithorax Proteins. Cell *128*, 735-745.
Schwartz, Y.B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. Curr Opin Cel Biol *20*, 266-273.
Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.
Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements. Cell *136*, 461-472.
Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell *6*, 791-802.
Surani, M.A. (2001). Reprogramming of genome function through epigenetic inheritance. Nature *414*, 122-128.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science *324*, 930-935.
Teixeira, F.K., and Colot, V. (2009). Gene body DNA methylation in plants: a means to an end or an end to a means. EMBO J *28*, 997-998.
Tiwari, S., Schulz, R., Ikeda, Y., Dytham, L., Bravo, J., Mathers, L., Spielman, M., Guzman, P., Oakey, R.J., Kinoshita, T.*, et al.* (2008). MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. Plant Cell *20*, 2387-2398.

Tiwari, S., Spielman, M., Schulz, R., Oakey, R., Kelsey, G., Salazar, A., Zhang, K., Pennell, R., and Scott, R. (2010). Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. BMC Plant Biology *10*, 72.
Twell, D. (2011). Male gametogenesis and germline specification in flowering plants. Sex Plant Reprod *24*, 149-160.
Vielle-Calzada, J.-P., Baskar, R., and Grossniklaus, U. (2000). Delayed activation of the paternal genome during seed development. Nature *404*, 91-94.
Wang, X., Soloway, P.D., and Clark, A.G. (2011). A Survey for Novel Imprinted Genes in the Mouse Placenta by mRNA-seq. Genetics *189*, 109-122.
Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567-576.
Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes & Development *2*, 1136-1143.
Weaver, J., Susiarjo, M., and Bartolomei, M. (2009). Imprinting and epigenetic changes in the early embryo. Mammalian Genome *20*, 532-543.
Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding Transcription by RNA Polymerase Pol IVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. Cell *135*, 635-648.
Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.
Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Kohler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. PLoS Genet *7*, e1002126.
Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genetics *4*, e1000156.
Woo, H.R., Pontes, O., Pikaard, C.S., and Richards, E.J. (2007). VIM1, a methylcytosine-binding protein requiered for centromeric heterochromatinization. Genes and Development *21*, 267-277.
Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Research *34*, 667-675.
Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. (2005). Methylation as a Crucial Step in Plant microRNA Biogenesis. Science *307*, 932-935.
Zemach, A., Kim, M.Y., Silva, P., Rodrigues, J.A., Dotson, B., Brooks, M.D., and Zilberman, D. (2010a). Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci U S A *107*, 18729-18734.
Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010b). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.
Zemach, A., and Zilberman, D. (2010). Evolution of Eukaryotic DNA Methylation and the Pursuit of Safer Sex. Current biology *20*, R780-R785.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Shan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide high resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell *126*, 1189-1206.
Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.-K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. EMBO J *26*, 1691-1701.
Zhu, J., Kapoor, A., Sridhar, V.V., Agius, F., and Zhu, J.K. (2007). The DNA glycosylas/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis. Curr Biol *17*, 54-59.
Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.
Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nature Genet *39*, 61-69.
Zilberman, D., and Henikoff, S. (2007). Genome-wide analysis of DNA methylation in plants. Development *134*, 3959-3965.

Appendix A

Figures:

Chapter. 1: No Figures

Chapter. 2:

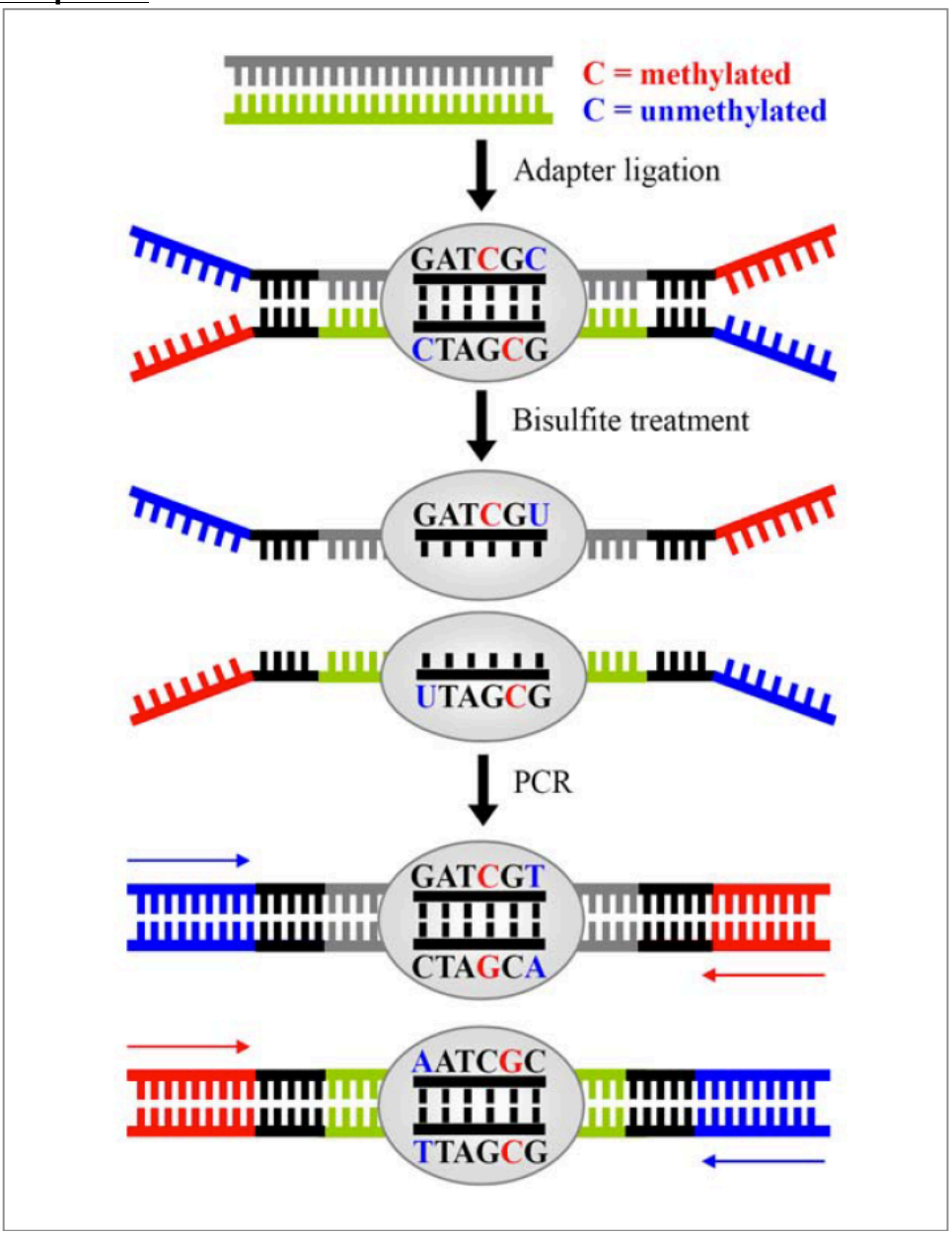


Fig. A2.1. A schematic of library construction for Illumina™ sequencing of bisulfite-converted DNA.

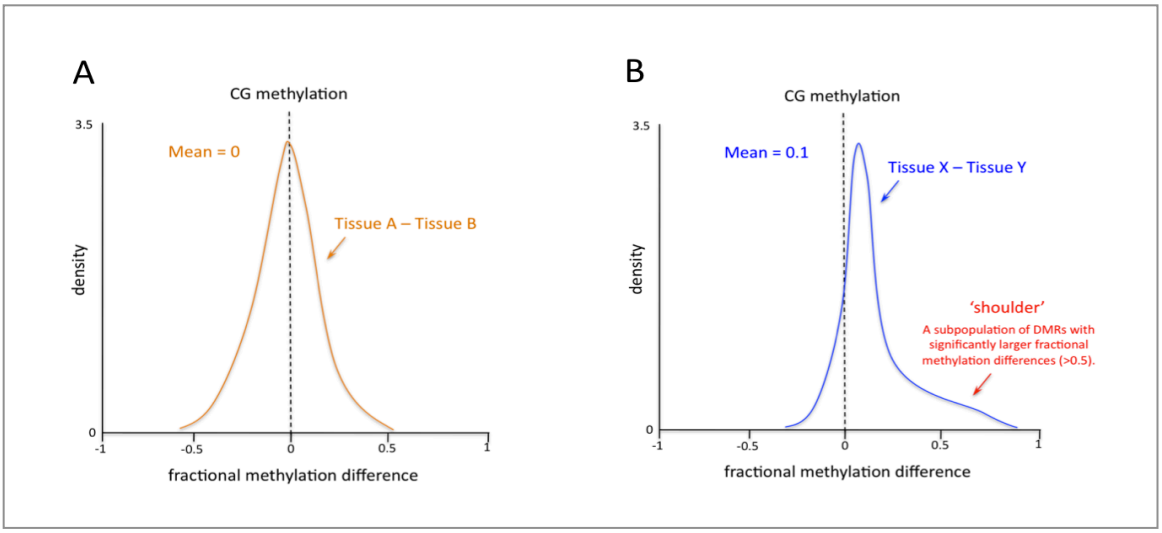


Fig. A2.2 (A and B). Illustration of kernel density plots from two fictitious tissues.

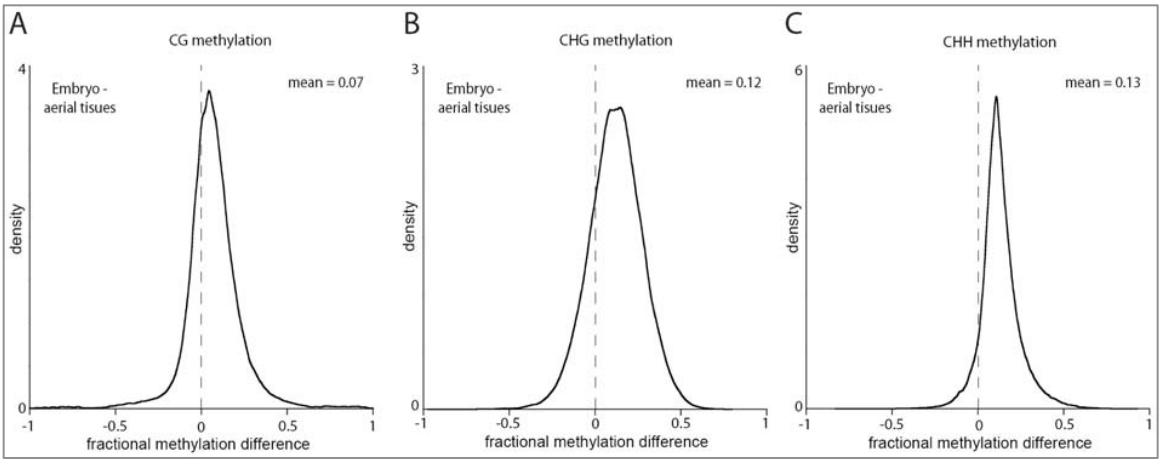


Fig. A2.3 (A – C). Kernel density plots tracing the frequency distribution of the differences between embryo and aerial tissue methylation. Methylation differences for 50 bp windows containing at least 20 informative sequenced cytosines are shown. Differences for windows with fractional CG methylation of at least 0.7 in one of the tissues are shown in (**A**), differences for windows with fractional CHG methylation of at least 0.5 in one of the tissues are shown in (**B**) and differences for windows with fractional CHH methylation of at least 0.1 in one of the tissues are shown in (**C**).

Chapter. 3: No Figures

Tables: **Chapter. 1: No tables**
Chapter. 2:

tissue	chr	bp		C count	T count	C ratio	C count*	T count*	C ratio*
WT embryo	Chr 1	568929150		8100672	91131683	0.0816	7826337	89846185	0.0801
WT embryo	Chr 2	593927055		11542687	103300284	0.1005	6590250	57572101	0.1027
WT embryo	Chr 3	456339330		790122	73691770	0.0968	6705595	68732114	0.0889
WT embryo	Chr 4	348287310		5953574	55096646	0.0975	5371624	53205556	0.0917
WT embryo	Chr 5	493137270		6939913	79060448	0.0807	6444776	77576772	0.0767
WT embryo	Chr C	368429850		65673	66330891	0.0010	65673	66330891	0.0010
WT embryo	Chr M	134482455		407053	28246756	0.0142	407053	28246756	0.0142
WT embryo	Chr 1-5	**2460620115**		40438138	402280831	0.0913	32938582	346932728	0.0867
WT endosperm	Chr 1	519334110		5258717	78924342	0.0625	5130421	78023936	0.0617
WT endosperm	Chr 2	477313830		7444184	83607198	0.0818	4693198	59490810	0.0731
WT endosperm	Chr 3	417591585		5406774	69564019	0.0721	4706791	65263111	0.0673
WT endosperm	Chr 4	318999060		4083906	52172735	0.0726	3727261	50402192	0.0689
WT endosperm	Chr 5	453793320		4818403	74849271	0.0605	4505556	73397577	0.0578
WT endosperm	Chr C	161853975		30369	29253964	0.0010	30369	29253964	0.0010
WT endosperm	Chr M	73043370		282324	15279872	0.0181	282324	15279872	0.0181
WT endosperm	Chr 1-5	**2187031905**		27011984	359117565	0.0700	22763227	326577626	0.0652
dme endosperm	Chr 1	473471280		3866155	78205445	0.0471	3752167	77126483	0.0464
dme endosperm	Chr 2	408568275		4898112	71844948	0.0638	3200457	54572318	0.0554
dme endosperm	Chr 3	375417945		3670247	62883907	0.0551	3210375	58930372	0.0517
dme endosperm	Chr 4	289063260		2886113	47523488	0.0573	2583741	45496525	0.0537
dme endosperm	Chr 5	412820100		3504514	68140113	0.0489	3223554	66441738	0.0463
dme endosperm	Chr C	215443035		57907	38593565	0.0015	57907	38593565	0.0015
dme endosperm	Chr M	42596055		192192	8862634	0.0212	192192	8862634	0.0212
dme endosperm	Chr 1-5	**1959340860**		18825141	328597901	0.0542	15970294	302567436	0.0501
aerial tissue	Chr 1	365412735		4343022	58104421	0.0695	4193559	57245155	0.0683
aerial tissue	Chr 2	346716855		6620646	59395185	0.1003	3444154	40187727	0.0789
aerial tissue	Chr 3	297334800		4125735	48412207	0.0785	3418007	43751567	0.0725
aerial tissue	Chr 4	223121250		3014958	35353278	0.0786	2707618	33843002	0.0741
aerial tissue	Chr 5	314888625		3606377	50236596	0.0670	3296869	48950469	0.0631
aerial tissue	Chr C	135402300		84568	24157791	0.0035	84568	24157791	0.0035
aerial tissue	Chr M	31992210		216638	6520041	0.0322	216638	6520041	0.0322
aerial tissue	Chr 1-5	**1547474265**		21710738	251501687	0.0795	17060207	223977920	0.0708

Table 2.1. Methylation statistics for embryo, WT endosperm, *dme* endosperm and aerial tissues. *Data were filtered to exclude Cs that were sequenced >20 times in a given sequencing run (10 times the expected coverage) on Chr1-5. Values in bold are reported in the main text.

Table 2.1 Continued

tissue	chr	bp		CG count	TG count	CG ratio	CG count*	TG count*	CG ratio*
WT embryo	Chr 1	568929150		3599331	10733704	0.2511	3486859	10662693	0.2464
WT embryo	Chr 2	593927055		5176488	15291623	0.2529	2844342	6392909	0.3079
WT embryo	Chr 3	456339330		3636331	9277480	0.2816	3026290	8046204	0.2733
WT embryo	Chr 4	348287310		2725687	6165902	0.3065	2445781	6110133	0.2859
WT embryo	Chr 5	493137270		3326702	9324266	0.2630	3064881	9245121	0.2490
WT embryo	Chr C	368429850		14473	11996754	0.0012	14473	11996754	0.0012
WT embryo	Chr M	134482455		167383	4861684	0.0333	167383	4861684	0.0333
WT embryo	Chr 1-5	**2460620115**		18464539	50792975	0.2666	14868153	40457060	**0.2687**
WT endosperm	Chr 1	519334110		2505628	9959942	0.2010	2448093	9914786	0.1980
WT endosperm	Chr 2	477313830		3541349	12844801	0.2161	2140387	7536584	0.2212
WT endosperm	Chr 3	417591585		2638537	9454838	0.2182	2254907	8308863	0.2135
WT endosperm	Chr 4	318999060		2014737	6415945	0.2390	1815349	6318183	0.2232
WT endosperm	Chr 5	453793320		2469766	9560360	0.2053	2291688	9437475	0.1954
WT endosperm	Chr C	161853975		6807	5355079	0.0013	6807	5355079	0.0013
WT endosperm	Chr M	73043370		121529	2641829	0.0440	121529	2641829	0.0440
WT endosperm	Chr 1-5	**2187031905**		13170017	48235886	0.2145	10950424	41515891	**0.2087**
dme endosperm	Chr 1	473471280		2689770	9550813	0.2197	2613086	9508886	0.2156
dme endosperm	Chr 2	408568275		3374184	10544507	0.2424	2161740	6673175	0.2447
dme endosperm	Chr 3	375417945		2563966	8152870	0.2392	2232724	7213327	0.2364
dme endosperm	Chr 4	289063260		2038053	5550042	0.2686	1802225	5457021	0.2483
dme endosperm	Chr 5	412820100		2520094	8352243	0.2318	2306407	8231574	0.2189
dme endosperm	Chr C	215443035		11707	7188394	0.0016	11707	7188394	0.0016
dme endosperm	Chr M	42596055		114125	1512307	0.0702	114125	1512307	0.0702
dme endosperm	Chr 1-5	**1959340860**		13186067	42150475	0.2383	11116182	37083983	**0.2306**
aerial tissue	Chr 1	365412735		2336050	7092601	0.2478	2254057	7054692	0.2421
aerial tissue	Chr 2	346716855		4033026	8809144	0.3140	1866107	4857035	0.2776
aerial tissue	Chr 3	297334800		2401101	6551553	0.2682	1910624	5352611	0.2631
aerial tissue	Chr 4	223121250		1718425	4134125	0.2936	1529951	4071831	0.2731
aerial tissue	Chr 5	314888625		2105253	6205153	0.2533	1912254	6110345	0.2384
aerial tissue	Chr C	135402300		14910	4519364	0.0033	14910	4519364	0.0033
aerial tissue	Chr M	31992210		104797	1108647	0.0864	104797	1108647	0.0864
aerial tissue	Chr 1-5	**1547474265**		12593855	32792576	0.2775	9472993	27446514	**0.2566**

Table 2.1 Continued: Methylation statistics for embryo, WT endosperm, *dme* endosperm and aerial tissues. *Data were filtered to exclude Cs that were sequenced >20 times in a given sequencing run (10 times the expected coverage) on Chr1-5. Values in bold are reported in the main text.

Table 2.1 Continued

tissue	chr	bp		CHG count	THG count	CHG ratio	CHG count*	THG count*	CHG ratio*
WT embryo	Chr 1	568929150		1506895	14386742	0.0948	1481504	14305698	0.0938
WT embryo	Chr 2	593927055		2464625	16267498	0.1316	1380440	8689595	0.1371
WT embryo	Chr 3	456339330		1555194	11611769	0.1181	1353975	10877571	0.1107
WT embryo	Chr 4	348287310		1106889	8403845	0.1164	1071553	8316478	0.1141
WT embryo	Chr 5	493137270		1237781	12465302	0.0903	1194898	12362822	0.0881
WT embryo	Chr C	368429850		10326	9885728	0.0010	10326	9885728	0.0010
WT embryo	Chr M	134482455		117094	4988577	0.0229	117094	4988577	0.0229
WT embryo	Chr 1-5	**2460620115**		7871384	63135156	0.1109	6482370	54552164	**0.1062**
WT endosperm	Chr 1	519334110		1122946	12660557	0.0815	1107603	12627601	0.0806
WT endosperm	Chr 2	477313830		1823446	13168413	0.1216	1118064	9236929	0.1080
WT endosperm	Chr 3	417591585		1212250	11154493	0.0980	1071914	10524238	0.0924
WT endosperm	Chr 4	318999060		877523	8115160	0.0976	846481	8038784	0.0953
WT endosperm	Chr 5	453793320		985696	12000016	0.0759	947642	11905911	0.0737
WT endosperm	Chr C	161853975		4797	4421054	0.0011	4797	4421054	0.0011
WT endosperm	Chr M	73043370		92030	2703418	0.0329	92030	2703418	0.0329
WT endosperm	Chr 1-5	**2187031905**		6021861	57098639	0.0954	5091704	52333463	**0.0887**
dme endosperm	Chr 1	473471280		696201	12710260	0.0519	683101	12672471	0.0511
dme endosperm	Chr 2	408568275		1016656	11542971	0.0809	676439	8708905	0.0721
dme endosperm	Chr 3	375417945		697759	10236847	0.0638	628040	9662325	0.0610
dme endosperm	Chr 4	289063260		523497	7466793	0.0655	499848	7372154	0.0635
dme endosperm	Chr 5	412820100		591072	11014537	0.0509	563325	10899358	0.0491
dme endosperm	Chr C	215443035		8780	5755850	0.0015	8780	5755850	0.0015
dme endosperm	Chr M	42596055		55636	1566252	0.0343	55636	1566252	0.0343
dme endosperm	Chr 1-5	**1959340860**		3525185	52971408	0.0624	3050753	49315213	**0.0583**
aerial endosperm	Chr 1	365412735		936323	9405999	0.0905	911431	9366332	0.0887
aerial endosperm	Chr 2	346716855		1486807	9545355	0.1348	818995	6336979	0.1144
aerial endosperm	Chr 3	297334800		874973	7889926	0.0998	762562	7153559	0.0963
aerial endosperm	Chr 4	223121250		624952	5542303	0.1013	601303	5471861	0.0990
aerial endosperm	Chr 5	314888625		711787	8133277	0.0805	678269	8042366	0.0778
aerial endosperm	Chr C	135402300		12317	3700086	0.0033	12317	3700086	0.0033
aerial endosperm	Chr M	31992210		65258	1169820	0.0528	65258	1169820	0.0528
aerial endosperm	Chr 1-5	**1547474265**		4634842	40516860	0.1027	3772560	36371097	**0.0940**

Table 2.1 Continued: Methylation statistics for embryo, WT endosperm, *dme* endosperm and aerial tissues. *Data were filtered to exclude Cs that were sequenced >20 times in a given sequencing run (10 times the expected coverage) on Chr1-5. Values in bold are reported in the main text.

Table 2.1 Continued

tissue	chr	bp		CHH count	THH count	CHH ratio	CHH count*	THH count*	CHH ratio*
WT embryo	Chr 1	568929150		2994446	66011237	0.0434	2857974	64877794	0.0422
WT embryo	Chr 2	593927055		3901574	71741163	0.0516	2365468	42489597	0.0527
WT embryo	Chr 3	456339330		2709767	52802521	0.0488	2325330	49808339	0.0446
WT embryo	Chr 4	348287310		212099	40526899	0.0497	1854290	38778945	0.0456
WT embryo	Chr 5	493137270		2375430	57270880	0.0398	2184997	55968829	0.0376
WT embryo	Chr C	368429850		40874	44448409	0.0009	40874	44448409	0.0009
WT embryo	Chr M	134482455		122576	18396495	0.0066	122576	18396495	0.0066
WT embryo	Chr 1-5	**2460620115**		14102215	288352700	0.0466	11588059	251923504	**0.0440**
WT endosperm	Chr 1	519334110		1630143	56303843	0.0281	1574725	55481549	0.0276
WT endosperm	Chr 2	477313830		2079389	57593984	0.0348	1434747	42717297	0.0325
WT endosperm	Chr 3	417591585		1555987	48954688	0.0308	1379970	46430010	0.0289
WT endosperm	Chr 4	318999060		1191646	37641630	0.0307	1065431	36045225	0.0287
WT endosperm	Chr 5	453793320		1362941	53288895	0.0249	1266226	52054191	0.0237
WT endosperm	Chr C	161853975		18765	19477831	0.0010	18765	19477831	0.0010
WT endosperm	Chr M	73043370		68765	9934625	0.0069	68765	9934625	0.0069
WT endosperm	Chr 1-5	**2187031905**		7820106	253783040	0.0299	6721099	232728272	**0.0281**
dme endosperm	Chr 1	473471280		480184	55944372	0.0085	455980	54945126	0.0082
dme endosperm	Chr 2	408568275		507272	49757470	0.0101	362278	39190238	0.0092
dme endosperm	Chr 3	375417945		408522	44494190	0.0091	349611	42054720	0.0082
dme endosperm	Chr 4	289063260		324563	34506653	0.0093	281668	32667350	0.0085
dme endosperm	Chr 5	412820100		393348	48773333	0.0080	353822	47310806	0.0074
dme endosperm	Chr C	215443035		37420	25649321	0.0015	37420	25649321	0.0015
dme endosperm	Chr M	42596055		22431	5784075	0.0039	22431	5784075	0.0039
dme endosperm	Chr 1-5	**1959340860**		2113889	233476018	0.0090	1803359	216168240	**0.0083**
aerial endosperm	Chr 1	365412735		1070649	41605821	0.0251	1028071	40824131	0.0246
aerial endosperm	Chr 2	346716855		1100813	41040686	0.0261	759052	28993713	0.0255
aerial endosperm	Chr 3	297334800		849661	33970728	0.0244	744821	31245397	0.0233
aerial endosperm	Chr 4	223121250		671581	25676850	0.0255	576364	24299310	0.0232
aerial endosperm	Chr 5	314888625		789337	35898166	0.0215	706346	34797758	0.0199
aerial endosperm	Chr C	135402300		57341	15938341	0.0036	57341	15938341	0.0036
aerial endosperm	Chr M	31992210		46583	4241574	0.0109	46583	4241574	0.0109
aerial endosperm	Chr 1-5	**1547474265**		4482041	178192251	0.0245	3814654	160160309	**0.0233**

Table 2.1 Continued: Methylation statistics for embryo, WT endosperm, *dme* endosperm and aerial tissues. *Data were filtered to exclude Cs that were sequenced >20 times in a given sequencing run (10 times the expected coverage) on Chr1-5. Values in bold are reported in the main text.]

Chapter. 3: No Tables