Title
Regulation of Imprinted Gene and Small RNA Expression in the Arabidopsis Endosperm

Author
Shin, Juhyun

Publication Date
2012

Peer reviewed|Thesis/dissertation

Regulation of Imprinted Gene and Small RNA Expression in the Arabidopsis Endosperm

By

Juhyun Shin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Environmental Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert Fischer, Chair

Professor Bob B. Buchanan

Professor Daniel Zilberman

Fall 2012

Copyright page
Some of the experiments presented in this Thesis were published in:

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in Arabidopsis endosperm, Proc Natl Acad Sci USA 108:1755-1762

Mosher RA, Tan EH, Shin J, Fischer RL, Pikaard CS, Baulcombe DC (2011) An atypical epigenetic mechanism affects uniparental expression of Pol IV-dependent siRNAs, PLos One 6: e25756

Abstract

Regulation of Imprinted Gene and Small RNA Expression in the Arabidopsis Endosperm

By

Juhyun Shin

Doctor of Philosophy in Agricultural and Environmental Chemistry

University of California, Berkeley

Professor Robert Fischer, Chair

Imprinted genes are expressed primarily or exclusively from either the maternal or paternal allele, a phenomenon that occurs in flowering plants and mammals. Flowering plant imprinted gene expression has been described primarily in endosperm, a terminal nutritive tissue consumed by the embryo during seed development or after germination. Imprinted expression in Arabidopsis thaliana endosperm is orchestrated by differences in cytosine DNA methylation between the paternal and maternal genomes as well as by Polycomb group proteins. Currently, only 11 imprinted A. *thaliana* genes are known. Here, we use extensive sequencing of cDNA libraries to identify 9 paternally expressed and 34 maternally expressed imprinted genes in A. thaliana endosperm that are regulated by the DNA-demethylating glycosylase DEMETER, the DNA methyltransferase MET1, and/or the core Polycomb group protein FIE. These genes encode transcription factors, proteins involved in hormone signaling, components of the ubiquitin protein degradation pathway, regulators of histone and DNA methylation, and small RNA pathway proteins. We also identify maternally expressed genes that may be regulated by unknown mechanisms or deposited from maternal tissues.

Small RNAs generated by RNA polymerase IV (Pol IV) are the most abundant class of small RNAs in flowering plants. In *Arabidopsis thaliana*, PolIV-dependent short interfering (p4-si) RNAs accumulate specifically in endosperm specifically from maternal chromosome. To identify epigenetic factors required for maternal-specific expression of p4-siRNAs, we analyzed the effect of a series of candidate mutations, including those required for genomic imprinting of protein-coding genes, on uniparental expression of a representative p4-siRNA locus. We demonstrate that the repression of paternal p4-siRNA expression at locus 08002 is not controlled by any of these proteins. Similarly, loss of several chromatin modification enzymes, including a histone acetyltransferase, a histone methyltransferase, and two nucleosome-remodeling proteins, does not affect maternal expression of locus 08002. Maternal alleles of imprinted genes are hypomethylated by DEMETER DNA glycosylase, yet expression of p4-siRNAs occurs irrespective of demethylation by DEMETER or related glycosylases. These data indicate that there may be an unknown mechanism controlling these p4-siRNAs, or they are deposited by maternal tissue in the endosperm.

Our results show that imprinted expression is an extensive mechanistically complex phenomenon that likely affects multiple aspects of seed development.

Table of Contents

Chapter I
Introduction

SUMMARY

Imprinted genes are expressed in a parent-of-origin specific manner. This phenomenon is only observed in mammals and flowering plants. Although there is variation, the mechanisms that control imprinted expression are similar in mammals and plants at the molecular level, involving DNA methylation/demethylation and/or histone modification by the Polycomb Repressive Complex 2 (PRC2). These two mechanisms, acting independently or together, regulate the expression or repression of alleles in a parent-of-origin specific manner that is independent of their sequence, resulting in their imprinted expression.

The endosperm in angiosperms is similar to the placenta in mammals, since both function to nourish the embryo. Endosperm is uniquely triploid, resulting from the fertilization of a homodiploid maternal central cell and a haploid paternal sperm cell. In Arabidopsis, endosperm is the main tissue where imprinting has been discovered, and mutations of imprinted genes like MEA or FIS2 cause seed abortion. This suggests that in plants, imprinting is a mechanism needed for proper development of the endosperm and for seed viability. Several theories have been proposed to explain the evolution of gene imprinting, with the parental conflict theory being the most widely accepted.

FERTILIZATION AND ENDOSPERM DEVELOPMENT IN ARABIDOPSIS

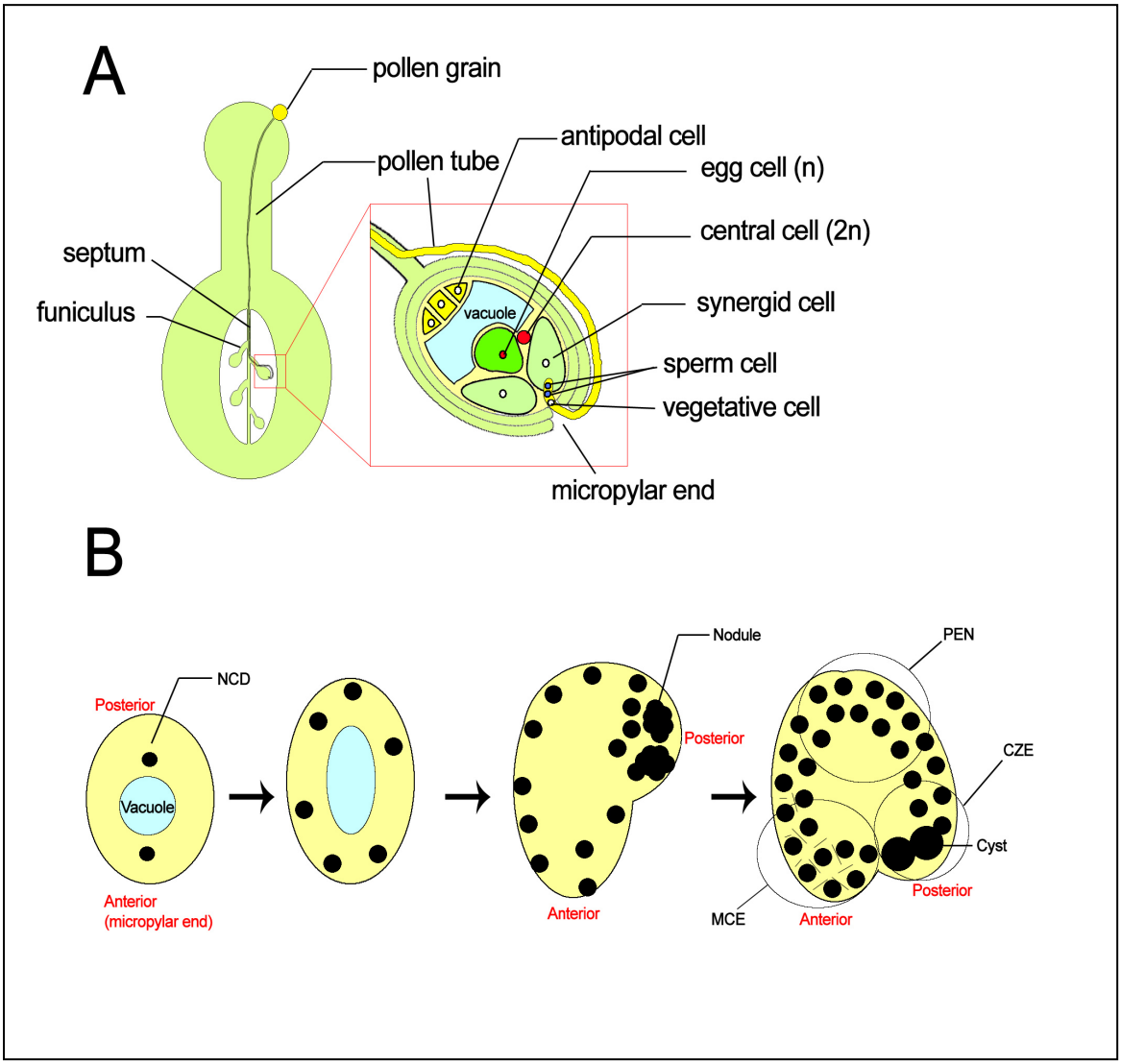


Figure 1. Arabidopsis gametophyte fertilization.
A. Double fertilization process, the red dots represent the two maternal cells that are fertilized by the blue paternal sperm cell.
B. Endosperm development after fertilization.

The seeds of flowering plants (angiosperms) are composed of distinctive compartments. Three different tissues are present in the seed: the embryo, endosperm and the seed coat. The seed coat is maternal tissue, while the embryo and endosperm are the result of two separate fertilization events, which is unique to angiosperms, called double fertilization.

In contrast to the animal life cycle, the plant life cycle consists of a multi-cellular haploid (gametophyte) generation that alternates with a diploid multicellular (sporophyte) generation that arises from the fusion of gametes. Gametogenesis occurs after meiosis and creates haploid spores, which, after several cell mitotic divisions, develop into the gametophyte. The gametophyte generation comprises most of the life cycle in fungi, algae or lower plants. However in higher plants, the gametophyte generation is relatively short and enclosed within

sporophytic reproductive tissues. Angiosperms produce two different kinds of gametophytes, female and male gametophytes, which are present in the flower.

The female gametophyte is embedded in the ovule, which is connected by the funiculus to the septum of the ovary. Both the ovary and the ovule are of sporophytic origin. The morphology of the female gametophyte differs among angiosperms, but its most common form, including that of Arabidopsis, consists of eight nuclei in seven cells - three antipodal cells, two synergid cells, one egg cell and one central cell. All the cells are haploid except the central cell, which has a homodiploid nucleus formed by the fusion of two polar nuclei prior to the cellularization of the female gametophyte (Mansfield et al, 1991).

At the start of the female gametogenesis, one diploid megaspore mother cell undergoes meiosis and forms four haploid megaspores. Only the chalazal-oriented megaspore survives and the three other megaspores undergo cell death. The surviving megaspore undergoes three more mitoses without cell division, giving rise to eight haploid nuclei. Shortly after nuclear division, two nuclei from opposite poles of the female gametophyte migrate toward the center and fuse, generating the large homodiploid nucleus of the central cell. Next, cellularization occurs, creating seven cells. The final product of this process is called the embryo sac or megagametophyte (Drews et al, 2002).

The male gametophyte arises from a diploid sporogenous initial cell (pollen mother cell). The pollen mother cell, and its sister tapetum initial cell, are embedded in the tissue that will ultimately become the anthers of the flower. The pollen mother cell undergoes meiosis and produces a tetrad of attached haploid cells. An enzyme, callase, produced by the adjacent tapetum, degrades the cells walls, freeing the haploid microspores. Microspores undergo synchronized mitoses. The first mitosis is asymmetrical - resulting in two haploid cells, a small generative cell that is embedded in a vegetative cell (McCormick et al, 1993). The vegetative and germ cell fate differ cytologically (McCormick et al, 1993) and molecularly (Twell, 2011). The generative cell undergoes a second mitosis to form two sperm cells that are likewise embedded in the vegetative cell, thereby generating the mature, tri-cellular pollen grain.

After the pollen grain hydrates and germinates upon a receptive stigma, the vegetative cell generates a pollen tube, which grows through and along the maternal septum and funiculus, attracted by the two synergids cells near the micropylar-end of the ovule. The pollen tube will penetrate one of the synergid cells, releasing the two sperm cells carried by the pollen tube. The overall growth and attraction of the pollen tube is carefully orchestrated by both maternal sporophyte, maternal gametophyte, and paternal gametophyte tissue (Ma et al, 2010, Kessler et al, 2011). One sperm cell fertilizes the haploid egg cell and becomes the embryo; while the other sperm fertilizes the diploid central cell and gives rise to the triploid endosperm.

In Arabidopsis, the primary endosperm nucleus, surrounded by cytoplasm (called a nuclear cytoplasmic domain, NCD), initially undergoes mitoses without cell division. This results in a large cell with multiple NCDs (coenocyte), which has a large vacuole in the center (Dumas et al, 2008). Even at that coenocyte stage, the endosperm shows distinct compartmentalization and polarity. At the anterior pole, where the pollen tube entered, the region that surrounds the embryo forms the micropylar endosperm (MCE), the central region becomes the peripheral endosperm (PEN) and the posterior pole becomes the chalazal endosperm (CZE). Experiments using reporter genes show that MCE-, PEN-, and CZE-specific expression occurs prior to cellularization, suggesting that the endosperm has distinct compartments during the coenocyte stage (Boisnard-Lorig et al, 2001, Stangeland et al, 2003, Berger et al, 2007). NCDs move from the anterior to the posterior pole, forming nodules that fuse into a large multinucleate

cyst that becomes the CZE. Formation of the CZE is a process influenced by Polycomb group (PcG) proteins MEA, FIS2 and FIE (Guitton et al, 2004), which might repress the expression of posterior-specific genes, as it was previously shown that mutations in PcG genes leads to ectopic expression of posterior markers (Sørensen et al, 2001). The size of the chalazal endosperm and overall seed size is also affected by genome dosage. For example, higher paternal genome ploidy results in an enlarged CZE and seed (Scott et al, 1998). Cellularization of the endosperm occurs shortly after the final round of mitosis, in a wave-like manner that starts at the micropylar endosperm (MCE), proceeds to the peripheral endosperm (PEN), and finally to the chalazal endosperm (CZE) (Brown et al, 1999).

IMPRINTED GENES

The term imprinting refers to the phenomenon of differential gene expression between alleles in an individual, dictated by the parental origin of the allele. From a broad perspective, parental origin differences have been observed in many organisms, including yeast, mollusks, fish and insects. In these organisms, parent of origin differences may be the result of nonrandom chromosome segregation, heterochromatization, destruction of the chromosome from one parent, or differences in DNA methylation and chromatin organization occurring without a change in transcription (de la Casa-Esperón and Sapienza, 2003). However, the term gene imprinting refers only to the mechanisms that result in differential expression (transcription) of a gene based on its parent-of-origin. This phenomenon is distinctive because the activation or repression of a gene is not based on sequence differences. Rather, it is based on epigenetic differences between alleles. Its phylogenic occurrence is restricted to angiosperms and viviparous animals – i.e. eutherian and marsupial mammals (Renfree et al, 2009).

Imprinting in mammals

In mammals, imprinting was first observed in the non-random inactivation of the X chromosome. In all cells of female marsupials and in extraembryonic tissues of the mouse, the paternal X chromosome is specifically silenced. This is in contrast to the random X-inactivation seen in other mouse tissues and in humans (Cooper et al, 1971). Later, the more narrow transcriptional differential expression was observed by studying why uniparental embryos were not viable. That phenomenon suggested that certain genes are exclusively expressed from only one parental genome, and their correct mono-allelic expression is indispensible for viable embryogenesis (Surani et al, 1984, McGrath et al, 1984). This suggestion proved to be true as it was shown that it was possible to make a viable mouse of pure maternal origin (Kono et al, 2004) if the imprinted gene dosage of a key growth factor is normalized to the wild-type level. These data provide evidence that imprinting may be an important barrier to parthenogenesis in mammals (Kawahara et al 2007).

Several genetic studies were done by mating mice with chromosomal translocations and observing the resulting parent-of-origin-specific defects, such as growth abnormalities, behavioral abnormalities, or death (Searle and Beechey, 1978, 1990, Cattanach 1982, 1986, Cattanach and Kirk 1985). These studies were not able to identify subtle or tissue-specific phenotypes, or to locate specific genes. However, they allowed researchers to focus on a portion of a chromosome, which enabled the preliminary identification of an imprinted region. In mice, more than 90% of the imprinted genes found to date are located in these regions (Bartolomei and Ferguson-Smith, 2011).

Imprinting in plants

In plants, scientists discovered imprinting by studying the R allele in maize. The anthocyanin distribution pattern in the aleurone layer of the endosperm in maize is dependent on the R allele. Genetic crosses revealed that, independent of its dosage, if a specific R allele is transmitted from the maternal parent, maize kernels are mottled, whereas if the same allele is paternally transmitted, solid colored kernels are observed. This contrasted with the inheritance pattern of another allele, R-stippled, which resulted in unstable spotting, because its expression was controlled in a dosage dependent manner (Kermicle, 1970). Subsequently, all of the imprinted genes identified in the Arabidopsis or maize have been found specifically in endosperm tissue, with the exception of the maize mee1 gene, which was found to be imprinted in both the endosperm and embryo (Jahnke et al, 2009).

Key Arabidopsis imprinted genes and their identification

FWA, an imprinted gene in Arabidopsis, is maternally expressed and paternally silenced in the endosperm. The genomes of various strains (ecotypes) of Arabidopsis are distinguished by single nucleotide polymorphisms (SNPs). One SNP that distinguishes the Col and Ler ecotypes is present in the coding region of *FWA* and was used for measuring allele-specific expression by RT-PCR in progeny from reciprocal crosses between two ecotypes. A PCR primer pair was designed to amplify a region including this polymorphic SNP so that the product sequence will differ based on the parental ecotype origin of the *FWA* transcripts. Following enzyme restriction, the parental origin of the sequence can be visualized by gel electrophoresis. These primers were used to generate cDNA from dissected F1 seeds from reciprocal crosses between the two strains (Col and Ler). This experiment showed that *FWA* expression in endosperm is of maternal origin. Also, visual analysis of GFP fluorescence revealed that the *pFWA::FWA-GFP* transgene is expressed only in the central cell prior to fertilization and in the endosperm after fertilization, which is consistent with its maternal imprinted expression. However, it should be noted that the function of the FWA protein in seeds is still not known (Kinoshita et al, 2004). Mutations in the *FWA* gene have no effect on seed development, which may be due to the expression of multiple related *FWA* genes in seeds. However, ectopic expression of the FWA protein, a homeodomain-containing transcription factor, causes a late flowering phenotype (Soppe et al, 2000).

MEDEA (*MEA*) is another gene shown to be maternally expressed in the endosperm using the allele-specific RT-PCR method described above. *MEA* encodes a SET-domain Polycomb group protein homologous to Enhancer of Zeste (E(z)) in *Drosophila*. In mammals, insects and fungi, SET domain Polycomb group proteins have been shown to repress transcription of specific genes by altering chromatin conformation (Pirrotta, 1998). In Arabidopsis, the phenotype resulting from mutations in the maternal *MEA* allele include excess endosperm cell proliferation, seed abortion and loss of embryo viability (Grossniklaus et al. 1998; Luo et al, 1999; Kiyosue et al, 1999).

FERTILIZATION INDEPENDENT SEED2 (*FIS2*) is another imprinted gene in Arabidopsis. Analysis of *FIS2::GUS* expression in seeds revealed that its expression was very similar to *MEA* and *FWA*. That is, *FIS2::GUS* was expressed in the central cell and only its maternal allele was expressed in endosperm (Luo et al, 2000). *FIS2* encodes a zinc-finger transcription factor homologous to Suppressor of Zeste 12 (Su(z)12) (Luo et al., 1999). In *Drosophila*, mutations in the Su(z)12 gene strongly influence position-effect variegation (PEV), a phenomenon observed in regions close to heterochromatin, suggesting it functions in heterochromatin regulation (Birve et al., 2001). In *Arabidopsis*, mutations in FIS2 cause phenotypes very similar to *mea* mutations – excess endosperm cell proliferation, seed abortion

and loss of embryo viability. This is not surprising given that FIS2 and MEA are both components of the Polycomb Regulatory Complex 2 (PRC2).

Using allele specific RT-PCR, *PHERES1* (*PHE1*) was the first imprinted gene identified that is paternally expressed in the Arabidopsis endosperm (Kohler et al., 2005). PHE1 is a MADS-box transcription factor. MADS-box proteins comprise a large family of transcription factors that control many developmental processes in plants (Parenicová et al, 2003). There are two types of MADS-box transcription factors, Type I and Type II, which can be distinguished on the basis of their sequences. Among the 61 Type I genes in the Arabidopsis genome, *PHERES1*, *DIANA*, *AGL62*, *AGL23* and *AGL80* have been shown to function in female gametophyte and early seed development (Ng and Yanofsky, 2001, Köhler et al, 2003, Bemer et al, 2010). *PHE1* is expressed after fertilization in the chalazal endosperm and is repressed in other parts of the seed by PRC2 proteins. By themselves, mutations in the *PHE1* gene do not have an easily detectable phenotype. However, *phe1* mutations partially rescue seed abortion caused by the *mea* mutation, showing that the seed abortion phenotype of *mea* is partially due to the failure of the PRC2 complex to repress the expression of *PHE1* (Köhler et al, 2003).

IMPRINTING BY METHYLATION AND HISTONE MODIFICATION

In mammals, imprinted genes are organized in large chromosomal clusters controlled by *cis*-acting imprinting control regions (ICRs) that regulate the imprinted status of genes several kilobases away (Reik et al, 2001). By contrast, the known imprinted genes found in Arabidopsis, rice or maize are not clustered and exist as singletons (Hsieh et al, 2011; Gehring et al, 2011; Luo et al, 2011; Waters et al, 2011). However, the molecular regulation of gene imprinting is remarkably similar in mammals and plants. To-date, two key mechanisms have been identified that epigenetically modify genomic DNA to control gene imprinting in plants and mammals: DNA methylation and histone modification.

DNA methylation controls imprinting in mammals

In mammals, ICR sequences are rich in CG dinucleotides. During gametogenesis, most of the ICRs in the genome are methylated in the female gametes, with only a subset of ICRs methylated in the male germline. Following fertilization, these methylation differences are maintained, and resulting in allele-specific gene expression. In the germ line within the embryo, by active demethylation that subsequently enables establishment of oocyte- or sperm-specific methylation and imprinting in the next generation (Feil et al, 2007). The mechanism of active DNA demethylation in the germ line is still not clear but is vital for imprinting, and therefore the growth and development of mammals (Morgan et al 2005).

In gametes, the Dnmt3 family of de novo DNA methyltransferases mediates the establishment of DNA methylation at ICRs. There are three Dnmt3 proteins; Dnmt3a, Dnmt3b and Dnmt3L. Dnmt3a and b are both needed for proper *de novo* oocyte- or sperm-specific methylation during gametogenesis (Okano et al, 1999). Dnmt3a was found to be necessary for methylation at differentially methylated regions (DMRs) near imprinted genes *Snrpn, Igf2r* and *Peg1* in oocyte gametogenesis (Lucifero et al, 2004). In male gametogenesis, Dnmt3a is needed for the methylation of H19, Dlk1/Gtl2 and short interspersed repeats, SineB1 DMRs. Both Dnmt 3a and Dnmt3b are needed for Rasgfr1 DMR and long interspersed IAP and Line1 methylation. This indicates that Dnmt3a and b have both distinct and overlapping targets for establishing the correct methylation patterns in mammalian oocytes and sperm. Dnmt3L sequence suggests that it

is not a catalytically active DNA methyltransferase. However this protein that is expressed during gametogenesis is essential for establishing maternal imprints via methylation, which show that it is an important regulator of gene imprinting (Bourc'his et al, 2001).

Dnmt1 maintains CG methylation in somatic cell lineages throughout development (Howell et al, 2001). Dnmt1 interacts with Ubiquitin-like PHD and RING finger domain 1 (UHRF1). UHFR1 specifically binds to hemimethylated CG dinucleotides. This suggests that UHFR1 directs conversion of hemimethylated CG to fully methylated CG by Dnmt1 (Arita et al, 2008).

The way DNA methylation at ICRs controls gene expression differs among mammalian imprinted genes. One of the best-studied examples is the ICR close to two imprinted genes, Igf2 and H19. The ICR between these two genes is only methylated on the paternal allele. This methylation blocks the binding of an enhancer-blocking zinc finger protein CTCF, which therefore only binds to the maternal allele. This creates a differential chromatin boundary that prevents interaction between Igf2 gene and enhancers downstream of H19. As a result, expression of the maternal Igf2 allele is repressed. By contrast, paternal H19 allele expression is repressed, as the methylation at the ICR spreads to the nearby H19 promoter (Delaval et al, 2004).

Plant DNA methyltransferases and other proteins maintain and establish DNA methylation

In angiosperms, DNA methylation is also an important mechanism for controlling gene imprinting. There are three different families of methyltransferases that have distinctive roles in plants: The METHYLTRANFERASE (MET) family, the CHROMOMETHYLASE (CMT) family and the DOMAINS REARRANGED METHYLTRANSFERASES (DRM) family (Chan et al, 2005).

MET1 is one of the four Arabidopsis Dnmt1 homologs in Arabidopsis (Goll et al, 2005) and is the primary DNA methyltransferase for maintaining CG methylation. Mutations in the MET1 gene result in global demethylation of cytosine in the CG sequence context, as well as partial loss of cytosine methylation in the CHG (H = A, C or T) context (Finnegan et al, 1996; Kankel et al, 2003). Although MET1's main role seems to be maintaining symmetric CG methylation, it may also play a role in *de novo* CG methylation, as it was observed that in met1 mutant plants, *de novo* CG methylation at the NOS promoter was not fully established despite normal levels of *de novo* methylation mediated by RNA-directed DNA methylation (RdDM) in other sequence contexts (Werner et al, 2004). MET1 mutations result in phenotypes that are morphologically visible, such as late flowering due to the ectopic expression of the *FWA* transcription factor gene (Kankel et al, 2003), or abnormal flower structure due to altered expression of *AP3* and *AG* transcription factor genes (Finnegan et al, 1997, Kishimoto et al, 2001).

The pattern of inheritance of phenotypes caused by a *met1* mutation illustrates one of the differences in methylation dynamics that distinguish plants from mammals. Plants do not strip off and re-establish DNA methylation every generation as mammals do. Plant CG methylation is stably maintained and inherited from generation to generation. When CG methylation is lost due to a *met1* mutation, it is not regained even when a wild-type *MET1* gene is inherited. That is, its hypomethylated state and phenotypic abnormalities persist in *MET1/met1* heterozygous plants. (Finnegan et al, 1996). This continuity contrasts with the mammalian methylation system, where methylation is reset in the early embryonic stage in the stem cell lines and re-established later during gametogenesis (Kafri et al, 1992).

In addition to MET1, there are three additional related DNA methyltransferases in Arabidopsis: MetIIa (MET2), MetIIb (MET3) and METIII (MET4) (Genger et al, 1999). A mutation in *MET1* does not completely abolish CG methylation (Kankel et al, 2003), suggesting that these others members may have a role in maintaining some CG methylation. Of the four homologues, METI is predominantly expressed in all tissues, with higher expression in meristematic tissue. However, *MET2* and *MET3* are also transcribed in all tissue at lower levels (Genger et al, 1999), and may be active. Similar to the mammalian model involving Dnmt1, the MET family provides methyltransferase activity, but other proteins are also needed for forming a complex to maintain CG methylation. Arabidopsis *VARIANT IN METHYLATION* (*VIM*) genes, that have an SRA domain and bind both to hemimethylated DNA and MET1, have been shown to be required for maintaining CG methylation (Woo et al 2008), possibly by recognizing hemimethylated DNA and directing MET1 to CG sites for methylation.

The CHROMOMETHYLASE (CMT) DNA methyltransferase family of proteins in Arabidopsis contain both a methyltransferase domain and a chromodomain. Chromodomains are protein regions that mediate interactions between chromatin proteins (Ingram et al, 1999). The CMT family in Arabidopsis consists of 3 members (CMT1, 2, 3) and is unique to plants (Henikoff et al, 1998). CMT3 is the main enzyme responsible for maintaining CHG methylation in Arabidopsis (Lindroth et al 2001). CMT3 also maintains certain non-symmetric CHH methylation (Batee et al, 2001). CMT3 methylates transposon related sequences (Tompa et al, 2002) and is necessary for transposon silencing (Kato et al, 2003, Lipman et al 2003).

The DOMAINS REARRANGED METHYLTRANSFERASES (DRM) family in Arabidopsis is related to the mammal methyltransferase Dnmt3. However, the DRM structure is unique to flowering plants (Goll et al, 2005). The Arabidopsis DRM family consists of two proteins, DRM1 and DRM2. It was shown that both are needed for de novo methylation in the GC, CHG and CHH contexts (Cao et al, 2002). By an experiment using two transgenes, producing a double-stranded RNA (dsRNA) and the other representing a 'target', whose promoter region can be silenced by the dsRNA, it was shown that DRM1 and 2 maintain and establish DNA methylation by the RdDM pathway (Cao et al, 2003).

As described above, the MET, CMT, and DRM DNA methyltransferases have distinct targets and roles for establishing and maintaining DNA methylation patterns in plants. For example, a triple mutant of DRM1, DRM2 and CMT3 is needed to completely abolish the establishment of *de novo* DNA methylation in all sequence contexts (Cao et al, 2003).

DNA demethylation and its role in controlling imprinting in Arabidopsis

A mutation in MET1 disrupts maintenance of CG methylation and causes passive DNA demethylation, which is due to a failure to convert hemimethylated DNA to fully methylated DNA during S-phase of the cell cycle. Passive demethylation disrupts the expression of two imprinted genes, *FWA* and *FIS2*. It was shown that loss of methylation at upstream regions caused expression of *FWA* and *FIS2* paternal alleles in endosperm, which are normally silenced in wild type endosperm.

DEMETER (DME), a DNA glycosylase/lyase protein, initiates active DNA demethylation. DME initiates the base excision repair (BER) pathway, excising 5-methylcytosine, and replacing it with unmethylated cytosine (Gehring et al, 2006). DME is expressed in the central cell where it demethylates and activates expression of *FWA* and *FIS2* maternal alleles (Kinoshita et al, 2004; Jullien et al, 2006). When a mutant maternal *dme* allele is inherited, *FWA* and *FIS2* expression is not activated in the central cell, and the maternal *FWA*

and *FIS2* alleles are not expressed in the endosperm. Likewise, maternal *MEA* allele expression requires DME activity (Choi et al, 2002; Gehring et al, 2006).

Histone modification in controlling imprinting

Besides methylation of genomic DNA, another epigenetic mechanism that controls the expression of genes is the post-translational modification of histone proteins. Histones are proteins that package eukaryote genomes into nucleosomes. A nucleosome consists of 167 bp of DNA wrapped in two left-handed turns around a core of eight histones: two H3 and two H4 plus two H2A/H2B dimers. The histone/DNA complex exists in two different states; a compacted heterochromatin state or a relatively open state euchromatin state and each are associated with different histone modifications (Elgin et al, 2003; Kouzarides, 2007). For example, histone H3 methylated at lysine 9 (H3K9me) allows the binding of repressive HP1 proteins, which results in heterochromatin formation and gene silencing (Lachner et al, 2001).

Mammalian ICRs are affected by DNA methylation and histone modifications

In mammals, several studies have shown that ICR CG DNA methylation and histone modification are linked processes. As described earlier, Dnmt3a is responsible for establishing *de novo* methylation at ICR in stem cell lineages, whereas Dnmt3b is required for the methylation of centromeric minor satellite repeats and some ICR elements (Okano et al, 1999, Kato et al, 2007). Dnmt3L is necessary in this process, by possibly binding to histones and recruiting Dnmt3a to its target (Jia et al, 2007). However, it was shown that Dnmt3L cannot bind to H3 if it is methylated at lysine 4 (H3K4me) (Ooi et al, 2007), which suggests that histone modification can affect DNA methylation at ICRs. Another example is the maternally human ICR for the SNRPN gene, where DNA methylation occurs after fertilization. This suggests that another mechanism (e.g., histone modifications) may mark the ICR prior to fertilization, which targets its parental-specific methylation after fertilization (Kantor et al, 2004).

DNA methylation is associated with specific histone modifications. After embryogenesis, Dnmt1 maintains the differential methylation established at ICRs. It was observed that the chromatin organization is different in these ICRs. In DNA hypermethylated alleles, histone H3 and H4 are hypoacetylated, and methylated at H3K9me3, H4K40me3 and H4H2AR3me2. HP1 proteins bind to these alleles, which generates heterochromatin. In hypomethylated counterpart alleles, it was shown that H3 and H4 are hyperacetylated, and are enriched in H3K4me2/3 (Kacem et al, 2009). In mutant Dnmt3L cells that are devoid of maternal DNA methylation imprints, histone H3K9me3, H4K20m3 and H2A/H4AR3me2 are greatly reduced in these ICRs, showing that DNA methylation and histone methylation are tightly connected (Henckel at al, 2009).

An interesting example of the interplay between DNA methylation and histone modification in imprinting can be seen in the placenta. The placenta connects the embryo to the maternal tissue and is an extra-embryonic tissue where a large number of mammalian imprinted genes function and have a profound effect on embryonic growth and development (Wasgschal et al, 2006). One imprinted domain, Kcnq1, is methylated on the maternal allele at the promoter of the non-coding RNA, *Kcnq1ot1*. Subsequent paternal-specific *Kcnq1ot1* expression results in recruitment of Polycomb group Proteins Eed, Suz12 and Esh2, which form the Polycomb repressive complex 2, conferring enrichment of H3K9 and H3K27 methylation on the paternal allele, leading to silencing. This conformation is established during early embryo and extra-embryonic development. Notably, in the placenta, the methylated histone status is maintained in the absence of allelic DNA methylation (Umlauf et al, 2004, Lewis and Reik, 2004).

Histone methylation and Polycomb complexes control plant gene imprinting

As described above, the maternally expressed *MEA* allele is activated by active DNA demethylation mediated by DME prior to fertilization in the central cell. However, the mechanism for MEA imprinting, and in particular, the mechanism for silencing the paternal allele, is more complex than *FWA* imprinting. MEA, a PRC2 complex component related to Drosophila E(z), self-regulates its imprinting. That is, maternally expressed MEA silences paternal *MEA* allele expression by methylating histones on the promoter of the paternal *MEA* allele (Ghering et al, 2006). In support of this model, a maternal mutation of PcG proteins, MEA or FIE, causes biallelic *MEA* expression in the endosperm (Gehring et al, 2006). Thus, MEA is an example of how paternal silencing, mediated by Polycomb proteins, is controlled independently of maternal allele activation, mediated by DNA demethylation. *PHERES1* (*PHE1*), a paternally expressed imprinted gene, is another example of silencing controlled by histone modification. It was shown that the maternal *PHE1* allele is silenced by H3K27 methylation mediated by the PRC2 complex. Interestingly, DNA hypomethylation of the maternal allele, likely catalyzed by the DME DNA glycosylase, is required for H3K27 methylation and maternal allele silencing (Makarevich et al, 2006).

Similar to examples shown above in mammals, in Arabidopsis, histone modification and DNA methylation are interdependent processes. The *SUPERMAN* and *PAI2* loci lose their CHG and CHH methylation in plants with mutations in the *KYP* gene, which encodes an H3K9-histone methyltransferase (Jackson et al, 2002; Malagnac et al, 2002). Also, DNA methyltransferase CMT3 binds to H3K9 methylated histones and methylates target CHG sites (Lindroth et al, 2004). Furthermore, CMT3 is shown to interact with homologs of HP1 protein, LPH1. Taken together, this show that non-CG methylation can be affected by H3K9 modification. By contrast, CG methylation is not affected by mutations in the *KYP* gene (Jackson et al, 2002; Malagnac et al, 2002). However, in *met1* mutant plants with almost no CG methylation, there is a dramatic loss of H3K9 methylation at heterochromatic centromeric and pericentromeric regions (Soppe et al, 2002; Tariq et al, 2003).

THEORIES FOR THE ROLE OF GENE IMPRINTING

Several theories have been proposed to explain the occurrence and perpetuation of gene imprinting during evolution. Monoallelic expression of genes is inherently dangerous, exposing the organism to deleterious phenotypes that would be complemented by a second gene copy had the gene displayed biallelic expression (Orr, 1995). Therefore imprinting must have a selective advantage that outweighs this problem. The following theories have been derived to try to explain imprinting phenomenon. Among these theories, the parental conflict theory is the most prominent because it explains many of the parent-of-origin transcriptional differences in angiosperms and viviparous mammals, and also may cover some of other parent-of-origin differences witnessed in insect or unicellular organisms (Moore and Mills, 2008).

Up to now, more than 14 different theories have been proposed, which can be grouped into three different, but not mutually exclusive, categories; 1) imprinting is a conflict of interest between parents, 2) imprinting is a beneficial mechanism that benefits the maternal parent, and/or 3) imprinting is a side-effect of other phenomena.

The parental conflict theory states that in species where maternal investment is high (angiosperms and viviparous mammals) and females can have offspring from multiple males, the interest toward nutrition allocation to the offspring is different between the parents (Haig, 2000).

Following the selfish gene theory that envisions genes promoting their own sexual transmission without cooperation from other genes (Doolittle and Sapienza, 1980), it states that certain paternally expressed imprinted genes promote the growth of their own progeny at the expense of progeny from different fathers and of the maternal parent. By contrast, maternally expressed imprinted genes restrict the growth of all offspring equally and reserve nutrition for the maternal parent and future progeny (Haig and Trivers, 1995, Hurst et al, 1996). This paradigm results in an arms race between maternally- and paternally-expressed imprinted genes (Haig and Trivers, 1995).

The parental conflict theory explains why imprinting is restricted to mammals and angiosperms, and why most imprinted genes are usually expressed in tissues that influence nutrition allocation to the embryo: the placenta and fetus in mammals and the endosperm in plants. This theory also explains many phenomena associated with imprinted pairs of growth promoters and inhibitors, like IGF2 and IGF2R, which are silenced at their maternal and paternal alleles, respectively. That is, single null mutations produce deleterious phenotypes, whereas mutants in both result in viable offspring (Haig and Trivers, 1995, Haig, 1997). It also explains why crosses between parents of different ploidy in Arabidopsis results in seed size differences based on the parent origin, where an increase of maternal genome dosage decreases seed size whereas increasing paternal genome dosage increases (Scott et al, 1998). Moreover, changes in DNA methylation caused by parent-of-origin inheritance of a *met1* mutation, affects seed size in a way predicted by the parental conflict theory. That is, as maternal *met1* mutants have larger seeds while paternal *met1* mutations result in smaller seeds (Xiao et al, 2006). As described above, DNA methylation is an important mechanism for controlling the expression of imprinted genes. Therefore, the effect of hypomethylated alleles on seed size can be explained by the disruption of maternally- and paternally-expressed imprinted genes.

Some other theories describe the driving force for the evolution of imprinting as a mechanism that is beneficial to the mother and/or the offspring. For example, imprinting has been viewed as way of controlling gene expression (Solter, 1988), cellular differentiation (Holliday, 1990), allowing gene expression during cell division (Hall, 1990), suppressing chromosome loss or gain (Thomas, 1995), preventing Ovarian trophoblast disease and overly invasive placentas (Varmuza and Mann, 1994; Hall, 1990) or inhibiting parthenogenesis (Solter, 1988). However, all these theories, while being able to explain some phenomena, cannot explain many features of imprinted genes, for example, why imprinting evolved in both mammals and angiosperms; or, they fail to predict the effect of mutations in imprinted genes.

It also has been speculated that imprinting is a side effect of the regulation of expression of imprinting genes, which may have some other important functions (Chandra and Nanjundia, 1990) besides controlling imprinted genes expression. However, if this is the case, there is no reason why imprinted genes did not 'escape' their imprinted status since biallelic expression is more advantageous, instead of persevering in the restricted phylogenic arena of mammals and angiosperms. For example, one theory that falls in this category asserts that imprinting is a side effect of a host defense system, which silenced genes nearby an area 'infected' by foreign DNA, such as a transposon (Barlow, 1993). This theory may explain how the mechanism of imprinting (transcriptional control via DNA methylation and histone modification) has evolved, but it doesn't explain why the phenomenon of imprinting has become fixed in mammals and angiosperms, and controls parent-of-origin gene transcription of genes that influence growth control of offspring.

Chapter II
Identification of imprinted genes in the Arabidopsis seed

INTRODUCTION

SNPs are a powerful tool to detect gene imprinting in plants

Single nucleotide polymorphisms (SNP) are differences of one nucleotide in two different sequences at a defined location. They are generated during evolution by mutations. They may be silent, provide slight phenotypic variation between organisms within a species, or may be detrimental. In Arabidopsis, SNPs exist between different inbred ecotypes and can be used as powerful markers for imprinted gene expression. They occur throughout the genome (on average 1 SNP per 100 to 300 bp) providing researchers with reasonable coverage of the whole genome (Appleby et al, 2009).

In Arabidopsis, the full-length genome sequence is available for the Col-0 ecotype (The Arabidopsis Genome Initiative, 2000). Based on Col-0 reference genome, SNPs distinguishing about 20 other ecotypes were investigated using high-density oligonucleotide arrays (Clark et al, 2007) or resequencing microarrays (Zeller et al, 2008). As described below, high throughput sequencing of the Ler genome has identified over 400,000 SNPs that distinguish Ler from Col-0.

High-throughput sequencing technology as a tool for studying the transcriptome

Recently a powerful technique called high throughput sequencing, or Next Generation Sequencing has been developed. Double stranded DNA is fragmented and amplified using a forked, polymorphic adaptor. As a result, from one double strand DNA fragment, two double-stranded DNA molecules will be produced with a different adaptor attached to each end. This product can be attached to the surface of a Flow Cell covered with a lawn of oligonucleotide strands that are complimentary to the adaptor molecules. Annealed fragments are repeatedly amplified until each adapter-DNA hybrid forms a large and detectable cluster, which is sequenced one base at a time using the addition of fluorescent deoxynucleotides with reversible terminators. The reversible nature of the terminators enables multiple sequencing cycles to take place on the flow cell (Bentley et al, 2008).

The Next Generation Sequencing (NGS) platform developed by the company Illumina® is utilized very widely in the research community. Illumina technology is evolving continually and has proven to be very efficient and relatively inexpensive (considering the sample throughput and genome coverage) compared to conventional sequencing methods. The sequences (reads) produced are relatively short, from 25-30 bp (improved now to 100 bp, with up to 250 bp planned) (Forrest et al, 2011). However, even short 25-35 bp reads are sufficient for sequencing small genomes like C. *elegans* (Hillier et al, 2007), or for detecting SNPs in already sequenced larger genomes of model organisms (e.g., *A. thaliana*). In 2008, the Illumina NGS platform was used to compare the sequences of Col-0, Bur-0 and Tsu-1 and revealed 823,325 SNPs, proving that this new method can be very efficient in discovering new SNPs between ecotypes in the Arabidopsis (Ossowski et al, 2008). Additionally, the Illumina NGS platform has been modified for bisulfite sequencing, and for sequencing cDNAs and small RNAs (smRNAs), making it a powerful tool for generating a highly integrated map of the Arabidopsis epigenome. In 2008, using Col-0 genomic DNA (gDNA) as a scaffold, a genome-wide profile of DNA methylation (methylome) was determined in Arabidopsis wild-type plants, CG methylation mutant plants (*met1*), non-CG methylation mutant plants (*drm1*/*drm2*/*cmt3*), and DNA demethylation mutant plants (*ros1*/*dml2*/*dml3*). In addition, wild type and mutant DNA methylation profiles were correlated with changes in smRNA and mRNA profiles (Lister et al, 2008). These studies demonstrate the power of genome-wide research on the Arabidopsis

epigenome using the Illumina system. Recently, DNA methylation profiles for Arabidopsis wild type and demethylase mutant (*dme*) endosperm were also elucidated (Hsieh et al, 2009). As previously described in chapter I, DNA methylation is an important regulatory mechanism that can control gene expression in a parent-of-origin manner. This study demonstrated that the small, ephemeral Arabidopsis endosperm was amenable to being analyzed by the Illumina system. As described below, to understand the extent and mechanism(s) of endosperm gene imprinting, I generated and analyzed parent-of-origin expression profiles of the Arabidopsis endosperm.

Identification of imprinted genes by genome-wide sequencing

In mammals, the discovery of imprinted genes has evolved from the identification of parentally non-equivalent chromosomal regions using linkage analyses, to the use of RT-PCR to identify discrete clusters of linked imprinted genes (Barlow et al. 1991, Ferguson-Smith et al 1991; DeChiaria et al, 1991) and by 2005, a list had been compiled comprising research from many different groups to describe a total of approximately 90 imprinted genes in mice, and 50 in humans (Morison et al 2005). More recently, genome-wide transcriptome analysis using high-throughput sequencing in mammals is facilitating discovery of additional novel imprinted genes and non-coding RNAs (Babak et al, 2008). In Arabidopsis, prior to my study, only eleven imprinted genes were known to be expressed in the endosperm and only three (*MEA, FIS2, FWA*) were extensively studied in terms of their function and imprinting mechanism. Our main goal was to identifying new imprinted genes by high-throughput sequencing cDNAs from F1 seeds generated from reciprocal crosses between wild type Col and Ler ecotypes, and using SNPs to elucidate parental contributions to gene expression. To gain an understanding of imprinting mechanisms, these experiments were also carried out in crosses between wild type and DNA methylation mutants (*met1*), DNA demethylation mutants (*dme*) and Polycomb group mutants (*fie*).

RESULTS

Read alignment and validation based on known genes

Two independently generated cDNA library pairs from reciprocal crosses of wild type Col and Ler ecotypes were sequenced using the Illumina GA2 platform. Sequence reads were aligned to the Col and Ler genomic scaffolds and assigned to ecotypes based on minimal mismatches. Each gene received a Col and Ler expression score equal to the number of reads assigned to respective ecotypes. Expression scores for each gene were obtained by calculating the number of reads per kb of sequence per 10 million aligned reads.

To validate our methods, the scores for the eleven known imprinted genes were examined. (Table1). Among these imprinted genes, the expressed potions of the genes *MEA* and *PHE1* lacked any SNPs between Col and Ler, therefore reads could not be assigned to the Ler and Col ecotypes. Also, the reads generated for the *MPC* gene did not cover any SNPs. Among the eight remaining imprinted genes, the number of reads for three genes (*HDG8*, *FWA* and *At5g62110*) were too few and not statistically significant. However, the trends observable from the available data concurred with previously published results; *HDG8* and *FWA* seemed to be maternally expressed in Col X Ler, and *At5g62110* seemed to be paternally expressed. Among the remaining statistically significant five genes, *HDG3*, *HDG9* and *MYB3R2* were scored as imprinted ($p<0.001$). *FIS2* also scored as imprinted ($p<0.05$). *FH5*, which was maternally expressed in reciprocal crosses of Col and C24 (Gerald and Berger, 2009), was maternally expressed in Col X Ler but was scored as biallelic in Ler X Col.

Table 1. Previously described imprinted gene reads
Total maternal (M) and paternal (P) reads are shown for the indicated genotypes, as well as transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm (Endo exp), embryo (Emb exp);*Transcriptional scores derived from hand-dissected (H) and LCM(L) tissue are shown; # = scored as imprinted with $p < 0.001$; $ = scored as imprinted with $p < 0.05$; NA = no SNPs between Col and L*er*.

Number	Annotation	CxL M/P	LxC M/P	**Endo exp H(L)***	**Emb exp H(L)***
AT1G02580	MEA	NA	NA	**15(0)**	**1(0)**
AT1G65330	PHE1	NA	NA	**0(0)**	**0(0)**
AT2G32370	HDG3 #	43/126	42/216	**108(256)**	**8(0)**
AT2G35670	FIS2 $	9/0	0/4	**9(13)**	**1(1)**
AT3G03260	HDG8	17/0	2/4	**2(5)**	**0(0)**
AT3G19350	MPC	0/0	0/0	**0(0)**	**0(0)**
AT4G00540	ATMYB3R2 #	100/0	75/2	**18(28)**	**0(1)**
AT4G25530	FWA	3/0	2/0	**6(8)**	**1(3)**
AT5G17320	HDG9 #	45/2	208/7	**75(203)**	**13(1)**
AT5G54650	FH5	340/53	186/160	**45(8)**	**29(11)**
AT5G62110	-	2/5	1/2	**1(3)**	**0(0)**

At a p-value < 0.05, 1,801 genes were scored as maternally expressed and 25 genes scored as paternally expressed. These numbers drop to 739 and 9, respectively, if the p-value cutoff is adjusted to 0.001. One of the previously known imprinted genes, *FIS2*, was detected at a $p < 0.05$, but not at $p < 0.001$, demonstrating that a significance value of < 0.001 excludes some biologically relevant data, at least in some cases. However, 1,801 genes is a rather striking value, significantly higher than previous estimates or indeed the number of estimated mammalian genes. Therefore we suspected that at this value of p, a large number of false positives are included, so we chose to filter our genes at a more stringent p-value of < 0.001.

Filtering data of endosperm from possible contamination using LCM data

The high maternal contribution in our database is likely to be partly an artifact due to the method used for isolating the endosperm tissue from the seed coat. As endosperms were separated from seed coats using manual hand-dissection under a light microscope, we could not absolutely rule out the possibility of seed coat RNA contamination of our endosperm. As the seed coat is of maternal origin, some of the genes scored as maternally expressed imprinted genes may be, in fact, non-imprinted genes expressed in the seed coat. For example, *TRANSPARENT TESTA 10* (*TT10*), a gene highly expressed in the seed coat (Pourcel et al, 2005) is present in our dataset of imprinted genes ($p < 0.001$). Therefore, we filtered our data using an independent cDNA library of Col X Ler endosperm tissue isolated by laser capture microscopy (LCM). This RNA pool, while yielding less reads than the libraries generated from hand dissected endosperm, was less prone to seed coat contamination because the fixed and sliced endosperm is separated from the seed coat using a highly accurate laser (Kerk et al, 2003). If, for a specific gene, the high maternal contribution in the hand-dissected dataset is coming from seed coat RNA, then, the fraction of maternal reads should drop dramatically in the LCM data. We considered genes to be imprinted if their expression levels in both of our manually dissected endosperm library pairs were no more than fourfold greater than those from the LCM dataset. We also included genes close to the above cutoff if their imprinted status was significantly altered by mutations that would not be predicted to affect their expression in maternal tissues such as the seed coat, but that are known to vastly alter the regulation of imprinting mechanisms, i.e. *met1, fie,* and *dme*. These filtering steps reduced the number of maternally expressed imprinted genes to 114 ($P < 0.001$) (Tables 1 and 2), which includes two previously reported new imprinted genes, *MYB3R2* and *HDG9* (Gehring et al, 2009). We focused further analyses on the LCM-filtered maternal $P < 0.001$ dataset and the paternal $P < 0.001$ dataset, because we believe that these are most likely to represent genes with truly imprinted endosperm expression (Tables 1, 2 and 3).

Validation and addition of some genes to the endosperm dataset by Sanger sequencing

We examined allele-specific expression of 52 genes by RT-PCR followed by conventional DNA sequencing, and the results agreed closely with those obtained by sequencing cDNA libraries using the Illumina GA2 platform. Primers were designed for spanning one or more SNPs, and cDNA from reciprocally crossed endosperm was used for RT-PCR. Amplified products were sequenced, and intensity peaks at SNP site were used for evaluating the parental contribution to the RNA pool. From the 52 genes analyzed, 43 genes were validated as imprinted. For 9 genes, expression was clearly monoallelic in one cross, but the peak intensity in the reciprocal cross was approximately equal for both bases of the SNP, representing biallelic expression. Overall, our validation data suggest that the majority of our newly identified genes are indeed imprinted, with a smaller number of genes displaying imprinting only in one of the reciprocal crosses. Similar

effects of ecotypes on parent-of-origin expression have been reported for other imprinted genes in A. *thaliana* and mammals.

In addition, we validated the imprinting status of three genes, whose p-value was close to 0.001, *SUVH8* (*At2g24740*), *JMJ15* (*At2g34880*) and *SUVH7* (*At1g1770*) and confirmed that these genes are imprinted; *SUVH8* and *JMJ15* being maternally expressed and *SUVH7* being paternally expressed in both reciprocal crosses. (Fig 1) This indicates that there are likely to be more imprinted genes in our dataset, but that were not included in our list due to our stringent cut-off. In total, we identified 116 new maternally expressed genes including 2 that were previously described (*HDG9* and *MYB3R2*), and 10 new paternally expressed genes, including previously described *HDG3*.

Figure 1. RT-PCR sequencing chromatograph of three genes with p-value close to 0.001

Selected SNP regions (center base in figure) were amplified by RT-PCR and sequenced in reciprocal crosses of Col and Ler seeds. All three genes show maternal RNA in reciprocal crosses.

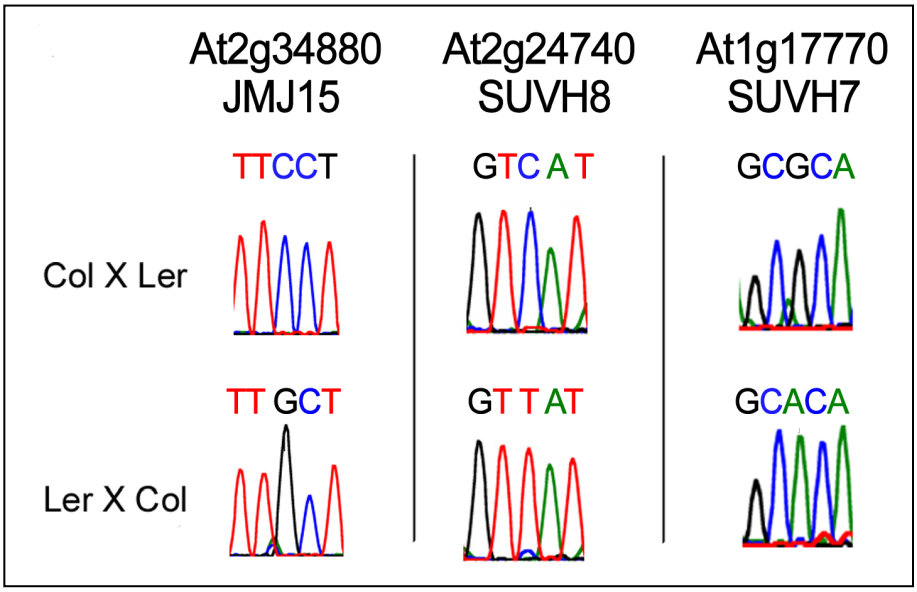


Table 2. Newly found maternally expressed imprinted genes list, at p<0.001 and LCM-filtered
CXL_M = sum of CXL 'pure' and 'full' endosperm maternal score (pure: hand-dissected with the less contamination of seed coat possible, but lacking most of the chalazal endosperm tissue; full: hand-dissected so that chalazal endosperm is mostly present, but has more potential to be contaminated by seed coat) **CXL_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **LXC_M**= sum of LXC 'pure' and 'full' endosperm maternal score, **LXC_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **M/P**=log(2) of maternal score divided by paternal score, **Fischer p**= Fischer's two tailed p value, **end exp**= endosperm #reads per sequence per 10 million aligned reads score average between full and pure data, **LCM exp**= exp of using LCM data (please see method)

Gene	Annotation	S	CxL_M	CxL_P	LxC_M	LxC_P	CxL_M/P	LxC_M/P	p	end_exp	LCM_exp
AT1G05280	Fringe-related protein	+	745	67	753	87	3.5	3.1	3.7E-175	388	217
AT1G05570	CALS1; 1,3-beta-glucan synthase/ transferase	-	49	6	34	2	3.0	4.1	1.0E-163	75	45
AT1G06470	Phosphate translocator-related	-	76	19	129	27	2.0	2.3	1.4E-137	65	52
AT1G07230	Hydrolase, acting on ester bonds	-	218	46	261	62	2.2	2.1	5.9E-116	79	34
AT1G08050	Zinc finger (C3HC4-type RING finger) family protein	+	24	1	42	4	4.6	3.4	2.8E-115	81	36
AT1G08830	CSD1; superoxide dismutase	-	144	2	117	20	6.2	2.5	3.4E-115	177	75
AT1G09380	Integral membrane family; nodulin MtN21-related	+	1036	89	1349	126	3.5	3.4	2.1E-108	3373	2596
AT1G11590	Putative pectin methylesterase	-	139	7	12	3	4.3	2.0	1.3E-87	31	10
AT1G13900	calcineurin-like phosphoesterase family	-	86	15	112	21	2.5	2.4	5.7E-73	53	28
AT1G17840	WBC11; ATPase, transmembrane transport of fatty acid	-	2816	355	2747	493	3.0	2.5	2.0E-66	284	198
AT1G21790	Unknown protein	-	66	3	57	6	4.5	3.2	2.0E-62	500	94
AT1G22180	SEC14; cytosolic phosphoglyceride transfer family	-	296	65	309	35	2.2	3.1	2.8E-62	121	70
AT1G24030	Protein kinase family	-	187	8	186	16	4.5	3.5	6.3E-56	34	16
AT1G28050	Zinc finger (B-box type) family	-	37	0	29	5	6.2	2.5	2.4E-47	155	85

AT1G29730	ATP binding / kinase protein	-	48	3	18	4	4.0	2.2	5.7E-44	10	6
AT1G31290	PAZ/piwi domain protein	+	324	33	280	22	3.3	3.7	1.3E-41	160	169
AT1G34180	anac016; transcription factor	-	11	0	99	15	4.5	2.7	6.1E-37	15	15
AT1G35580	CINV1; beta-fructofuranosidase	-	102	11	79	10	3.2	3.0	2.2E-34	60	36
AT1G35630	Protease-associated zinc finger family	-	144	0	84	3	8.2	4.8	1.9E-30	48	18
AT1G42470	Patched family	+	24	5	67	3	2.3	4.5	1.0E-28	34	12
AT1G44750	ATPUP11; purine transmembrane transporter	-	66	12	97	23	2.5	2.1	4.6E-27	291	93
AT1G54570	Esterase/lipase/thioesterase family	-	27	2	44	3	3.8	3.9	1.3E-24	215	302
AT1G54575	Unknown protein	+	538	104	517	22	2.4	4.6	3.1E-23	485	258
AT1G59930	Unknown protein	+	495	1	465	0	9.0	9.9	6.8E-23	419	1123
AT1G61090	Unknown protein	-	244	0	419	2	8.9	7.7	1.2E-22	125	127
AT1G61097	Expressed unknown protein	-	8	0	25	0	4.0	5.6	2.6E-21	18	19
AT1G62660	BFRUCT3; vacuolar beta-fructofuranosidase/invertase	+	14	2	123	0	2.8	7.9	2.7E-21	9	0
AT1G63690	Protease-associated (PA) domain	-	1168	285	1690	354	2.0	2.3	3.1E-21	574	312
AT1G64610	WD-40 repeat family protein	-	24	1	30	1	4.6	4.9	3.1E-20	14	5
AT1G69900	Unknown protein	+	61	4	64	10	3.9	2.7	6.9E-18	75	54
AT1G72470	ATEXO70D1; protein binding	-	43	5	36	3	3.1	3.6	1.5E-17	8	3
AT1G72790	Hydroxyproline-rich glycoprotein family	+	52	12	90	18	2.1	2.3	2.2E-17	271	98
AT1G72810	Putative threonine synthase	-	137	22	173	34	2.6	2.3	3.9E-16	91	42
AT1G73390	Unknown protein	+	53	11	101	14	2.3	2.9	1.4E-15	24	9
AT1G76250	Unknown protein	+	140	3	36	1	5.5	5.2	1.6E-15	44	29
AT1G76820	GTP binding / GTPase	-	45	4	50	12	3.5	2.1	5.0E-15	71	19

AT1G77000	SKP2B; ubiquitin-protein ligase	+	279	64	264	60	2.1	2.1	2.8E-14	150	119
AT1G77850	ARF17; auxin response transcription factor	+	114	25	196	32	2.2	2.6	4.3E-14	104	59
AT1G77960	Unknown protein	-	27	1	104	2	4.8	5.7	1.3E-13	31	11
AT1G78830	Curculin-like lectin family	-	6	1	41	0	2.6	6.4	3.4E-13	6	2
AT1G79520	Cation efflux family	-	58	9	57	0	2.7	6.8	2.5E-12	35	24
AT1G80510	Amino acid transporter family	-	50	2	37	6	4.6	2.6	2.6E-12	28	20
AT2G04620	Cation efflux family	-	66	14	105	22	2.2	2.3	4.5E-12	296	191
AT2G11810	MGDC; galactosyltransferase	-	20	0	23	0	5.3	5.5	3.0E-11	14	7
AT2G13560	Putative malate oxidoreductase	-	75	18	73	8	2.1	3.2	4.0E-10	162	82
AT2G17690	SDC; Suppressor of DRM1 DRM2 CMT3	+	124	1	119	2	7.0	5.9	4.9E-10	59	96
AT2G17990	Unknown protein	+	311	32	185	45	3.3	2.0	5.3E-10	430	263
AT2G28380	DRB2; double-stranded RNA binding	+	806	198	737	171	2.0	2.1	1.1E-09	936	343
AT2G29730	UGT71D1; UDP-glycosyltransferase	-	38	1	20	0	5.2	5.3	2.1E-09	9	4
AT2G31360	ADS2; oxidoreductase	+	369	5	174	28	6.2	2.6	4.0E-09	2448	507
AT2G36310	URH1; uridine nucleosidase	-	319	66	527	87	2.3	2.6	4.1E-09	186	63
AT2G39650	Unknown protein	-	184	45	165	40	2.0	2.0	6.2E-09	176	84
AT2G40020	Unknown protein	-	81	16	123	27	2.3	2.2	8.4E-09	170	54
AT3G05700	Unknown protein	-	86	17	66	14	2.3	2.2	1.2E-08	27	15
AT3G06860	MFP2; 3-hydroxyacyl-CoA dehydrogenase/ enoyl-CoA hydratase	-	113	23	159	30	2.3	2.4	1.6E-08	199	117
AT3G10590	Myb family transcription factor	+	47	3	69	3	4.0	4.5	2.9E-08	62	57
AT3G17250	Protein phosphatase 2C-related	+	33	8	64	8	2.0	3.0	3.2E-08	220	255

AT3G21830	ASK8; ubiquitin-protein ligase	+	252	1	318	1	8.0	8.3	3.3E-08	149	96
AT3G21860	ASK10; ubiquitin-protein ligase	+	34	0	99	0	6.1	7.6	4.0E-08	34	48
AT3G22810	Phosphoinositide binding	+	263	48	328	32	2.5	3.4	4.1E-08	138	100
AT3G22968	CPuORF59; Conserved peptide upstream open reading frame 59	-	90	6	50	9	3.9	2.5	4.5E-08	74	51
AT3G23570	Dienelactone hydrolase family	-	19	0	26	2	5.2	3.7	1.5E-07	9	4
AT3G25290	auxin-responsive family	+	301	25	238	42	3.6	2.5	1.7E-07	401	182
AT3G27300	G6PD5; glucose-6-phosphate dehydrogenase	-	44	4	58	12	3.5	2.3	2.3E-07	19	17
AT3G28960	Amino acid transporter family	+	40	6	172	9	2.7	4.3	2.4E-07	40	22
AT3G51895	SULTR3;1; sulfate transmembrane transporter	-	45	3	41	8	3.9	2.4	2.5E-07	188	132
AT3G53410	Zinc finger family	-	44	8	59	8	2.5	2.9	4.0E-07	199	65
AT3G54100	Unknown protein	-	164	31	157	35	2.4	2.2	6.3E-07	85	40
AT3G54740	Unknown protein	-	35	1	37	4	5.1	3.2	7.7E-07	22	11
AT4G00220	JLO; JAGGED LATERAL ORGANS	+	666	11	836	15	5.9	5.8	7.7E-07	478	186
AT4G00540	ATMYB3R: MYB domain transcription factor	+	100	0	75	2	7.6	5.2	1.2E-06	18	28
AT4G00570	Putative malate oxidoreductase	-	73	12	157	25	2.6	2.7	2.5E-06	65	51
AT4G01840	KCO5 ; calcium activated potassium channel	+	96	3	68	5	5.0	3.8	2.5E-06	30	12
AT4G12080	DNA-binding family	-	89	8	42	10	3.5	2.1	3.6E-06	45	13
AT4G15080	Zinc finger family	-	66	6	70	16	3.5	2.1	3.8E-06	34	13
AT4G16180	Unknown protein	-	155	37	214	48	2.1	2.2	8.5E-06	50	43
AT4G16760	ACX1; acyl-CoA oxidase	+	898	92	645	86	3.3	2.9	9.3E-06	76	38
AT4G18150	Unknown protein	+	203	2	871	3	6.7	8.2	1.3E-05	147	70

AT4G18320	Unknown protein	+	342	1	159	0	8.4	8.3	1.4E-05	90	297
AT4G18650	Transcription factor-related	+	807	89	92	14	3.2	2.7	2.5E-05	434	32
AT4G29570	putative cytidine deaminase	+	34	4	82	5	3.1	4.0	2.8E-05	49	75
AT4G29580	putative cytidine deaminase	-	49	1	45	0	5.6	6.5	2.8E-05	20	121
AT4G29640	Putative cytidine deaminase	+	32	8	127	14	2.0	3.2	3.4E-05	87	73
AT4G29860	EMB2757; embryo defective, nucleotide binding	-	54	6	51	10	3.2	2.4	3.5E-05	26	12
AT4G31060	AP2 domain; putative transcription factor	+	719	9	229	7	6.3	5.0	4.9E-05	498	150
AT4G39140	Protein binding / zinc ion binding	+	81	10	92	21	3.0	2.1	6.1E-05	131	62
AT4G39955	Hydrolase, alpha/beta fold family	+	35	5	145	1	2.8	7.2	6.7E-05	48	25
AT5G02630	Unknown protein	-	31	0	26	0	6.0	5.7	7.1E-05	6	0
AT5G02880	UPL4; ubiquitin-protein ligase	-	198	44	287	64	2.2	2.2	7.2E-05	115	170
AT5G02970	Hydrolase, alpha/beta fold family	-	334	33	284	71	3.3	2.0	7.6E-05	119	48
AT5G03280	EIN2; ethylene related transporter	+	1346	207	1725	169	2.7	3.4	7.8E-05	730	610
AT5G03370	Acylphosphatase family	+	168	13	83	16	3.7	2.4	9.0E-05	42	21
AT5G13820	TBP1; telomeric dsDNA binding	-	45	4	40	6	3.5	2.7	9.1E-05	146	88
AT5G15470	GAUT14; polygalacturonate 4-alpha-galacturonosyltransferase	-	134	11	85	21	3.6	2.0	1.1E-04	119	63
AT5G17320	HDG9; homeodomain protein	-	45	2	208	7	4.5	4.9	1.1E-04	75	203
AT5G20280	ATSPS1F; sucrose-phosphate synthase	-	185	34	229	35	2.4	2.7	1.1E-04	405	229
AT5G21150	PAZ / piwi domain containing	+	74	18	76	6	2.0	3.7	1.2E-04	29	20
AT5G22200	Harpin-induced family protein	+	387	2	43	2	7.6	4.4	1.2E-04	394	189
AT5G22920	Zinc finger	+	64	7	55	10	3.2	2.5	1.6E-	62	48

	family								04		
AT5G23340	Unknown protein	-	162	33	157	27	2.3	2.5	1.9E-04	90	36
AT5G24460	Unknown protein	+	342	79	278	43	2.1	2.7	2.8E-04	194	181
AT5G33290	XGD1; UDP-xylosyltransferase	-	500	93	587	93	2.4	2.7	3.1E-04	206	88
AT5G35490	Unknown protein	+	43	0	30	0	6.4	5.9	3.6E-04	32	16
AT5G36940	CAT3; cationic amino acid transmembrane transporter	-	43	10	106	17	2.1	2.6	3.6E-04	16	20
AT5G42235	Expressed protein	+	2226	160	4438	949	3.8	2.2	3.8E-04	1907	457
AT5G44350	Ethylene-responsive nuclear protein - related	-	121	7	94	20	4.1	2.2	3.9E-04	45	45
AT5G47560	TDT; malate transmembrane transporter/ sodium:dicarboxylate symporter	-	44	2	12	1	4.5	3.6	3.9E-04	24	16
AT5G47770	FPS1; dimethylallyltrans transferase	+	45	6	111	15	2.9	2.9	4.0E-04	66	60
AT5G49890	CLC-C; voltage-gated chloride channel	-	164	24	127	26	2.8	2.3	4.4E-04	39	20
AT5G53250	AGP22; arabinogalactan protein	+	58	2	28	0	4.9	5.8	4.5E-04	24	6
AT5G53870	Plastocyanin-like domain-containing	+	241	21	222	13	3.5	4.1	8.3E-04	38	9
AT5G57900	SKIP1; SKP1 interacting	-	74	8	63	5	3.2	3.7	9.4E-04	82	31
AT5G64400	Unknown protein	-	88	21	132	32	2.1	2.0	9.8E-04	121	59
AT5G64440	AtFAAH; N-(long-chain-acyl)ethanolamine deacylase/ amidase	-	61	14	163	13	2.1	3.6	9.8E-04	21	13

Table 3. Newly found paternally expressed imprinted genes list, at p<0.001
CXL_M = sum of CXL 'pure' and 'full' endosperm maternal score, **CXL_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **LXC_M**= sum of LXC 'pure' and 'full' endosperm maternal score, **LXC_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **M/P**=log(2) of maternal score divided by paternal score, **Fischer p**= Fischer's two tailed p value (please see method)

Gene	Annotation	CxL_M	CxL_P	LxC_M	LxC_P	CxL_M/P	LxC_M/P	p
AT1G31640	AGL92; transcription factor	1	6	1	14	-2.5	-3.8	0.0
AT1G48910	YUC10; FAD binding / monooxygenase/ oxidoreductase	36	70	30	279	-0.9	-3.2	2.5E-52
AT1G57800	VIM5; Variant in methylation5, protein binding	249	2818	343	3513	-3.5	-3.3	0
AT1G60410	F-box family protein	4	33	6	32	-3.0	-2.4	2.0E-11
AT2G21930	F-box family protein	2	15	0	18	-2.9	-5.1	1.5E-07
AT2G32370	HDG3; homeodomain containing transcription factor	43	126	42	216	-1.5	-2.3	4.9E-45
AT2G36560	DNA-binding protein-related	0	55	3	42	-6.7	-3.8	9.5E-24
AT4G11940	Unknown protein	1	6	1	15	-2.5	-3.9	0.0
AT5G63740	Zinc finger protein-related	16	34	16	59	-1.0	-1.8	1.2E-10

Table 4. VIM and MET genes

A list of all A. *thaliana* VIM and MET genes. Total maternal (M) and paternal (P) reads are shown for the indicated genotypes, as well as transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm (Endo exp), embryo (Emb exp), and the indicated mutant genotypes.

*Transcriptional scores derived from manually-dissected and LCM tissue are shown before and after the slash (/), respectively.

Table 4.

Number	Annotation	CxL M/P	LXC M/P	*met1* M/P	*fie* M/P	*dme* M/P	Endo exp*	Emb exp*	*met1* exp	*fie* exp	*dme* exp
AT1G57820	VIM1	233/144	168/442	23/51	215/178	289/72	114/47	717/632	55	136	95
AT1G66050	VIM2	22/7	28/24	1/0	206/126	81/85	69/73	74/67	11	856	436
AT5G39550	VIM3	2/17	30/13	1/0	151/20	14/58	35/38	138/141	19	329	111
AT1G66040	VIM4	NA	NA	NA	NA	NA	50/62	56/53	8	678	380
AT1G57800	VIM5	249/2,81	343/3,51	10/246	3,734/3,966	1,066/694	458/20	22/2	157	1,965	324
AT5G49160	MET1	1/0	2/1	2/0	7/0	6/6	87/87	327/375	34	264	287
AT4G08990	MET2	93/13	60/55	1/1	141/81	326/81	36/78	6/4	3	136	130
AT4G13610	MET3	12/17	59/11	2/0	1,204/461	115/479	13/19	1/1	1	434	121
AT4G14140	MET4	176/72	199/117	11/14	619/344	681/222	66/205	7/5	29	261	172

Imprinted genes in the embryo

Embryo cDNA libraries from reciprocal crosses were prepared and sequenced using the same methods as described for endosperm. At $p < 0.001$, 37 maternally expressed and one paternally expressed imprinted gene(s) were identified. Since the one paternally expressed imprinted gene we identified, *VIM5*, and 30 of the maternally expressed imprinted genes identified in the embryo were also present in the $p < 0.001$ list of endosperm imprinted genes, we suspected that our embryo tissue might be contaminated with endosperm tissue.

Therefore, the embryo data set was filtered by LCM data using similar methods as described above. Only two potentially maternally expressed genes remained after filtering, *At1g70830* and *At5g47150*. These two genes were also in the endosperm-imprinted list but were discarded as their LCM data indicated that their maternal-specific expression might be due to seed coat contamination. That is, in endosperm, their imprinted score was more than four times higher in the hand-dissected data compared to the LCM data. As it indicates that *At1g70830* and *At5g47150* are highly expressed in seed coat and/or the endosperm, we discarded them, as their imprinting status in the embryo is likely to be an artifact of RNA contamination from the endosperm and/or seed coat.

In summary, at $p < 0.001$, the number of potentially imprinted genes found in embryo was much lower than the endosperm. After LCM filtering only two maternally highly expressed imprinted genes remained, but were discarded since they might be artifacts from tissue contamination. Therefore, in this study, no genes were identified as being imprinted in the embryo.

DISCUSSION

In plants, the endosperm tissue of the seed has a major influence on embryogenesis. Endosperm proliferation and cellularization define seed size and possibly future embryo size, especially in dicot plants like Arabidopsis, where endosperm growth precedes embryo growth. Also, several transcription factors, signal molecules and hormones produced in the endosperm are known to affect seed size and viability (Sun et al, 2010). Finally, the embryo uses nutrients stored in the endosperm for growth and maturation (Fiume and Fletcher 2012). Therefore, genes expressed in the endosperm are important for understanding the process of seed formation. One particular set of genes, the imprinted genes were our focus, as their elaborate expression mechanism suggests that they may have an important functional role in seed formation and may reflect the parental conflict of nutrient allocation to the embryo.

Using high-throughput sequencing of endosperm cDNA libraries, our study has significantly expanded the number of known imprinted genes in Arabidopsis. Based on the density of SNPs that distinguish the Ler and Col ecotypes, and the depth of sequencing, we estimate that our data set represents roughly one-half of the endosperm transcriptome, which is about 10,755 genes assuming that two thirds of the 28,244 genes in *A. thaliana* are expressed in the endosperm (Day et al, 2007; Le et al, 2010). Based on our stringent cutoff ($p < 0.001$), we found a total of 114 genes that were maternally expressed and 10 genes that were paternally expressed. However, whilst the stringency of our cutoff reduces the number of potential false positives in our list, it also may result in our underestimation of the true number of imprinted genes in our data set. Therefore we conservatively estimate that we have found around 20 – 30% of the total number of imprinted genes. Hence, we estimate that the total number in Arabidopsis may be 200 – 500 maternally expressed imprinted genes and 30 – 50 paternally expressed imprinted genes. Consistent with our estimate, in a separate study, another set of 208 imprinted genes have been found, with only 20 genes overlapping with our data (Gehring et al, 2011). This low level of overlap can be explained by (i) the small number of genes found respectively in the two studies compared to the true number of potentially imprinted genes and (ii) by the differences in statistical criteria used. In addition to these differences, the growth stage and techniques used for isolating tissue from endosperm varied between our two studies, which could have an effect if there are highly tissue specific or transient imprinting effects. In another study, a reciprocal cross between Col-0 and Bur-0 gift generated a different set of imprinted genes, again with only few overlaps to our data and those of Gehring et al. In addition to the reasons outlined above, the difference in ecotypes used in this study may add further level of complexity and difference to the imprinted genes identified (Wolf et al, 2011). It will be interesting to gather all the existing raw data and process them using a uniform set of statistical criteria as it will allow us to have a more complete understanding of the imprinted transcriptome in the Arabidopsis endosperm.

While many of the imprinted genes do not overlap in the different data sets generated by different investigators, it is interesting to note that some features are still in conserved among the data sets. For example, in all three studies, endosperm is essentially the only tissue where imprinting occurs, and it is shown to be nearly non-existent in the embryo (Hsieh et al, 2011; Ghering et al, 2011; Wolf et al, 2011). An analysis in the rice seeds derived from reciprocal crosses revealed that rice imprinting occurs mainly in endosperm and that the few embryo imprinted candidate genes were almost all discarded when more stringent statistical criteria were used (Luo et al, 2011). In a study that focused on the transcriptome of the very early Arabidopsis

embryo (1-2 cells), it was found that maternal and paternal contribution to the transcriptome is almost equivalent. Thus, parent-of-origin-specific expression in the Arabidopsis embryo is very short lived compared to the Arabidopsis and rice endosperm and the mammalian embryo (Nodine and Bartel, 2012).

Imprinting has been found not only in the mammalian placenta (the equivalent of the endosperm tissue in plants) but also extensively in the embryo, throughout fetal development and also in the adult in a large number of tissues, particularly the brain of mammals (Bartolomei and Ferguson-Smith, 2011). By contrast, plant gene imprinting seems to be mainly restricted to the endosperm, a tissue that is not transmitted to the next generation. Also, it is noteworthy to state that in the four studies cited above, while the imprinted gene list is not complete, imprinted genes were not clustered in the genome, as is found in mammals. These two features represent significant differences between plant and mammalian imprinting.

The parental conflict theory proposes that the parents antagonistically influence nutrient flow to the embryo. In our study, we found that this antagonistic relationship may occur at many different regulatory levels. For example, for conflict at the level of chromatin, it was previously shown that genes regulating histone modification can be imprinted, such as maternally expressed *MEA* and *FIS*. In our study, we found that *VIM5*, encoding a protein needed for maintaining CG DNA methylation and *SUVH7*, encoding a protein involved in H3K9 methylation are paternally expressed. In addition, we found genes encoding proteins regulating RdDM that are maternally expressed (*DRB2, SUVH8, JMJ15*). We detected genetic conflict at the post-translational level of gene expression – protein degradation mediated by the ubiquitin-26S proteosome system. In Arabidopsis, nearly 6% of the total proteosome consists of proteins related to the ubiquitin-26S proteosome system. In Arabidopsis, the number of F-box and SKP proteins that direct protein degradation of specific target proteins rivals the number of transcription factors. Thus, degradation of proteins via the ubiquitin-26S proteosome system is a well-established, robust mechanism for controlling gene expression (Vierstra, 2009). In our study, we found imprinted genes encoding proteins that target protein degradation – maternally expressed SKP2B and paternally expressed F-box genes (*At1g60410* and *At2g21930*). Finally, imprinting of hormone synthesis (*YUC10* and *ACX1*) and response (*JLO* and *EIN2*) genes suggests that hormone action may also be involved in parental conflict. All in all, our study suggests that imprinted genes encode antagonistic regulatory proteins that function at many different levels. It will be interesting to follow up each of these genes and understand their function in their respective networks, as this will help us understand their role in endosperm nutrient allocation, embryo development, and seed yield.

Our data provide insights into the regulation of maintaining CG methylation in endosperm. In this study, we found that one of the components of the VIM family needed for CG maintenance, *VIM5*, is paternally expressed. Interestingly, *VIM5* is not expressed highly in somatic tissues and organs, leaves or inflorescences, where *VIM1/2/3* genes are highly expressed and have overlapping functions in maintaining of global CG methylation (Woo et al, 2008). We also found that in our endosperm transcriptome data, the main methyltransferase that maintains CG methylation in *A. thaliana, MET1*, is not the most highly expressed (Genger et al, 1999) in endosperm. Rather, *MET2* and *MET4* are unexpectedly expressed at higher levels (Table 4). These data suggest that in endosperm, different members from the canonical CG methyltransferase complex are active in the endosperm. These results may help explain why the *A. thaliana* endosperm genome is hypomethylated in the CG context relative to the embryo genome. (Hsieh et al, 2009). That is, partial suppression of the maintenance methylation

machinery in early endosperm development may be responsible for its global CG hypomethylation compared to the embryo.

METHODS

Plant growth conditions

All seeds were sterilized in 100% isopropanol and 50% bleach/0.05% Tween 20 solution. Then, they were planted on 0.5X MS plates and grown at 4º C in a dark chamber for 2 days, and germinated in a continuous light growth chamber for approximately 10 days. Seedlings were transplanted on soils and growth in a 16:8 light:dark, 22º C greenhouse environment.

Crosses and RNA isolation from tissue

Stage 12-13 flower buds were emasculated two days prior to being reciprocally crossed. Siliques were collected at 6 to 8 DAP (days after pollination), which corresponds to the embryo walking stick stage. Endosperm and embryo were isolated on a slide using a light microscope. Two sets for both endosperm and embryo in reciprocal crosses were dissected. One set of hand-dissected tissue contained endosperms that were dissected with the least amount of endosperm tissue loss possible whilst removing the vast majority of the seed coat, and was called 'full'. Another set contained endosperm that have been dissected with the least seed coat contamination possible, but may have lost some endospermal tissues, especially in the chalazal region. This set was called 'pure'. Then embryo tissue was also separately isolated from the seeds of the two separate pools and labeled as 'full' or 'pure', even if the method for isolating embryo were identical.

For each pool, approximately 7-10 siliques were dissected and tissues were isolated in 20ul of RNA later solution (Quiagen). Total RNA was extracted from dissected endosperm and embryo tissue using RNAeasy kit (Quiagen) plus in-column DNase digestion.

Laser Capture Microscopy

Laser capture microdissection was done essentially as described (GEO accession no. GSM311287). Siliques at stages that contained seeds with embryos from the linear to the bent cotyledon stage were fixed according to the methods previously used (Kerk et al. 2003). The embryo proper or the entire endosperm was dissected and captured from seed sections using LMD 6000 system (Leica Microsystems), and RNA was isolated using the RNaqueous microkit (Ambion).

Library construction

Approximately 30-50ng of total RNA were converted to double-stranded cDNA using the Ovation RNA-seq System(NuGen Technologies) according to the manufacturer's protocol. This system is based on a single primer, isothermal amplification (SPIA) technology using both poly-T oligos and random hexamers. This method was shown to be adequate for cDNA amplification for the next-generation sequencing platform with good correlation to the transcriptome using a small amount (from 500pg) of total RNA (Tariq et al, 2011).

Identification of SNP between Col and Ler

SNPs between Columbia (Col) and Landsberg erecta (Ler) were identified by sequencing Ler genomic libraries using the Illumina GA2 platform and these sequences mapped onto the TAIR8 Col scaffold. Using MAQ aligner, 402,226 SNPs with an average coverage of 8.7-fold per SNP, minimum 2-fold coverage per SNP and consensus quality of 30, were identified.

Identification of imprinted genes

76 bp reads from Illumina were aligned to Col and Ler scaffolds using Bowtie allowing up to three mismatches per alignment. Reads were assigned to Col or Ler based on their mismatch score. That score was called maternal or paternal score respectively in reciprocal

crosses. Also, all reads were assigned to corresponding gene, independently of their ecotype for computing the transcriptional score normalized per 1kb sequence per 10 million reads as gene length and library sequencing intensity can affect the read number per gene. Pure and full library reads score were summed for making the final reads score.

The probability that a gene's expression deviates from expectation was calculated by Fisher's two-tailed exact test. The genes' parental score with a low p-value ($p<0.05$ or $p<0.001$), that is, with a low statistical probability that the null hypothesis (biallelic expression) is true have been selected. Within this cutoff, only genes that maternally score four times greater than the paternal score in reciprocal crosses were selected as maternally imprinted. For the paternally imprinted genes, only genes with 1.5 fold higher paternal score in both crosses were considered as imprinted.

Validation of Imprinted genes

Primer sets were designed to amplify cDNA fragments that have one SNP between Col and Ler. RNA from reciprocally crossed F1 endosperm and embryo tissue were converted to cDNA by random-primed first-strand synthesis (Ambion) or by using Ovation RNA-seq System (NuGen Technologies) according to manufacturer's protocol. cDNAs were sequenced by the Sanger method and the cDNA SNP Col and Ler contribution were determined by Phred score and peak intensity at the site of interest.

Chapter III
Regulation of imprinting by DME, MET1 and/or FIE

INTRODUCTION

Up to now, only a few imprinted genes had been studied carefully enough to understand the mechanisms that control their parent-of-origin expression. Based on expression changes caused by mutations in DNA methyltransferase genes, DNA demethylation genes, and/or genes encoding Polycomb group complex components, three distinct mechanisms have been proposed to control imprinted genes expression in *A. thaliana* endosperm. Using genome-wide methods, we not only increased the number of known imprinted genes, but also identified groups of imprinted genes that are regulated by these and other mechanisms.

Paternal allele silencing by DNA methylation

FWA, an imprinted gene in Arabidopsis, is maternally expressed and paternally silenced in the endosperm. Analysis of *pFWA::FWA-GFP* expression revealed that *FWA* is expressed in central cell and endosperm (Kinoshita et al, 2004). *FWA* encodes a homeodomain-containing transcription factor, but its role in seed tissue is still not clear (Kinoshita et al, 2004). However, ectopic expression of *FWA* causes late flowering in transgenic Arabidopsis plants (Soppe et al, 2000).

In the endosperm, 5'-CG sites in the promoter of the maternal *FWA* allele are hypomethylated compared to the paternal allele, and to other tissues (embryo, seed coat, leaf and pollen). Analysis of *FWA* gene imprinting in DNA methylation mutant backgrounds revealed its imprinting mechanism. Biallelic expression of *FWA* in F1 endosperm is observed when wild type plants are pollinated with *met1* pollen. By contrast, expression of maternal and paternal *FWA* alleles is not detected when *dme* mutant plants are pollinated with wild type pollen (Kinoshita et al, 2004). These results suggest that MET1 maintains CG methylation in the *FWA* promoter of the silenced paternal allele, while DME demethylates 5' CG methylation that activates the maternal *FWA* allele. Similar results indicate that the imprinted expression of the *FIS2* Polycomb group gene is regulated by the same mechanism (Jullien et al, 2006).

Paternal allele silencing by histone modifications

Another maternally expressed gene, *MEA* shows a different imprinting mechanism. *MEA* encodes a SET-domain Polycomb group (PcG) protein homolog to Drosophila Enhancer of Zeste (E(z)) that is maternally expressed in the endosperm. Mutations in *MEA* cause excess endosperm cell proliferation, loss of embryo viability, and seed abortion (Grossniklaus et al., 1998; Luo et al, 1999; Kiyosue et al, 1999). DME demethylates sequences flanking the *MEA* gene, which activates *MEA* maternal expression in the central cell and endosperm (Gehring et al, 2006). However, the paternal *MEA* allele was silenced when wild type plants were pollinated with *met1* pollen. Thus, DNA methylation is not required for stable silencing of the paternal *MEA* allele. Additional genetic crosses revealed that PRC2 silences the paternal *MEA* allele. That is, biallelic *MEA* expression in endosperm was detected when a plant with mutations in genes encoding PRC2 components (MEA or FIE) was pollenated with wild type pollen. (Gehring et al, 2006). PRC2 methylates histone H3 at amino acid K27 (H3K27). Indeed, the *MEA* paternal allele is enriched for H3K27 methylation, and this enrichment requires maternal-derived PRC2 activity (Gehring et al, 2006). Maternal PRC2 proteins also silence the paternal allele of the actin regulator, ARABIDOPSIS FORMIN HOMOLOG 5 (FH5) (Fitz Gerald et al, 2009).

The model proposed for these two genes is that before fertilization, DME demethylates and activates *MEA* expression in the central cell, which binds FIE and other Polycomb group proteins

to form the PRC2. After fertilization, in the endosperm, PRC2 maintains silencing of the paternal allele of imprinted genes by H3K27 methylation

Maternal allele silencing by DNA demethylation and histone modifications

Analysis of the expression of the imprinted gene, *PHERES1* (*PHE1*), helped to elucidate a general mechanism for maternal allele silencing. PHE1 is a MADS-box type I transcription factor that is expressed after fertilization in the chalazal endosperm and repressed in other parts of the seed by PRC2. Phenotypes are not detected in *phe1* mutant seeds. However, a *phe1* mutation partially rescues *mea* mutations, showing that the seed abortion phenotype associated with *mea* mutations is partially due to the failure of PRC2 to repress *PHE1* expression (Köhler et al, 2003).

PHE1 is a paternally expressed imprinted gene that is biallelically expressed in endosperm with maternally inherited mutations in genes encoding PRC2 proteins (Köhler et al 2003, Köhler et al 2005). Maternal demethylation of tandem repeats downstream of *PHE1* is also thought to be required for maternal *PHE1* allele silencing. This idea is supported by the observation that loss of methylation in the paternal genome because of a *met1* mutation reduced expression of the paternal *PHE1* allele (Villar et al, 2009, Makarevich et al 2008). Indeed, DME-dependent endosperm hypomethylation of these tandem repeats has been reported (Hsieh et al 2009). Thus, the current model explaining regulation of *PHE1* gene imprinting proposes that maternal DNA demethylation near the gene exposes a PRC2 binding site, thereby allowing PcG-mediated silencing of the maternal allele (Villar et al, 2009). Supporting this model, it was recently reported that DNA hypomethylation allows targeting by PcG proteins in endosperm (Weinhofer et al 2010). This model predicts that demethylation of the paternal genome by a *met1* mutation should silence similarly regulated genes by exposing the paternal allele to PRC2-mediated repression, whereas a maternal *fie* mutation should cause biallelic expression by disabling PRC2.

RESULTS

Maternally imprinted genes controlled by DNA methylation

A maternally expressed imprinted gene that is regulated by DNA methylation as described in the model above will undergo biallelic expression in a paternal *met1* background and neither allele will be expressed in a maternal *dme* background. In our study of the endosperm transcriptome, nine new imprinted genes have been identified that, for the most part, adhere to this model (Table1). Interestingly, while the number of new imprinted genes is relatively small, they have potentially an important role in the endosperm because they are predicted to encode regulatory proteins. *MYB3R2* and *ERF/AP2* (*At4g31060*), a truncated *PHE1*-related *MADS* box transcription factor gene (*At1g59930*), and three genes known to be regulated by DNA methylation: *SDC* (*At2g17690*) and *MRU1* (*At5g35490*), which are overexpressed in lines lacking non-CG methylation (Henderson et al, 2008; Kurihara et al, 2008), and *At4g18650*, a transcription factor gene down-regulated by mutation of the DME homolog *REPRESSOR OF TRANSCRIPTIONAL GENE SILENCING 1* (*ROS1*) in seedlings (Zhu et al, 2007). *SDC* encodes an F-box protein that is predicted to confer specificity to the E3 ligase complex that ubiquitylates proteins targeted for degradation by the 26S proteasome (Vierstra, 2009). Among *met1*-affected genes are two regulators of hormone signaling: *JAGGED LATERAL ORGANS* (*JLO; AT4G00220*), a transcription factor that affects transport of the plant hormone auxin by regulating the expression of *PINFORMED* auxin-efflux carrier genes (Bureau et al, 2010), and *ETHYLENE INSENSITIVE 2* (*EIN2; At5g03280*), a membrane protein crucial for perception of the gaseous hormone ethylene that is also required for proper auxin, abscisic acid, jasmonic acid, salicylic acid, and cytokinin signaling (Stepanova and Alonso, 2009). *DOUBLE-STRANDED RNA BINDING 2* (*DRB2; At2g28380*) is a predicted component of the small RNA pathway (Chen, 2009). Available microarray data (http://seedgenenetwork.net; GEO accession no. GSE12404) shows that the *met1*-affected genes are expressed primarily in endosperm. Although small in number, many maternally expressed imprinted genes affected by *met1* are likely endosperm-specific key regulators that activate or repress other genes.

Table 1. 9 new imprinted genes with maternal allele expression controlled by demethylation (DME) and paternal allele silencing controlled by methylation (MET1).
CXL_M = sum of CXL 'pure' and 'full' endosperm maternal score, **CXL_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **LXC_M**= sum of LXC 'pure' and 'full' endosperm maternal score, **LXC_P**= sum of CXL 'pure' and 'full' endosperm paternal score, **M/P**=log(2) of maternal score divided by paternal score, **Fischer p**= Fischer's two tailed p value, also result for **CX*met1*** and ***dme*XL** using same statistic calculation(please see method), NA# for SUVH8 as it was validated by RT-PCR(Chapter1, Figure 1.)

Table 1.

Gene	Annotation	CxL_M	CxL_P	LxC_P	LxC_M	CxL_M/P	LxC_M/P	P	end exp	met1_M	met1_P	met1_M/P	P	met1 exp	dme_M	dme_P	dme_M/P	P	dme exp
AT1G59930	PHERES-like imprinted transcription factor-related (truncated)	495	1	0	465	8.95	9.9	2.1E-108	419	202	189	0.1	2.3E-118	1497	32	1	5.0	0.1	93.1
AT2G17690	SDC; F-box protein targets protein for ubiquitylation sending it to the 26S proteosome	124	1	2	119	6.95	5.9	1.3E-24	59	33	87	-1.4	4.1E-52	150	10	0	4.3	1.0	8.5
AT2G24740	SUVH8; SET-domain protein, H3K9 methyltransferase	48	0	19	50	6.58	1.4	2.6E-03	_	3	1	1.58	5.2E-01	-	77	0	7.3	0.0	-
AT2G28380	DRB2; DS RNA-binding, component of sRNA patway	806	198	171	737	2.03	2.1	6.8E-23	936	94	60	0.65	7.5E-08	454	69	11	2.6	0.2	272.1
AT4G00220	JLO; transcription factor, regulate PIN expression for auxin transport	666	11	15	836	5.92	5.8	1.4E-137	478	116	45	1.37	6.3E-31	379	36	1	5.2	0.5	36.8
AT4G00540	MYB3R2, transcription factor	100	0	2	75	7.64	5.2	6.9E-18	18	2	7	-1.81	9.3E-10	6	36	2	4.2	0.1	7.4
AT4G31060	ERF/AP2 transcription factor	719	9	7	229	6.32	5.0	1.3E-87	498	45	62	-0.46	7.4E-56	382	17	0	5.1	1.0	13.0
AT5G03280	EIN2, ethylene hormone signaling	1346	207	169	1725	2.7	3.4	2.8E-115	730	155	80	0.95	2.1E-19	317	269	65	2.0	0.0	141.8
AT5G35490	MRU1, unknown function, overexpressed in non-CG mutant	43	0	0	30	6.43	5.9	6.2E-09	32	4	6	-0.58	5.6E-07	18	0	0	0.0	1.0	1.2

Maternally imprinted genes controlled by Polycomb complexes

A maternally expressed imprinted gene that is regulated by Polycomb group proteins as described in the above model will display maternal expression in a paternal met1 background and biallelic expression in a maternal *fie* background. In our data, 20 new genes were found to follow that expression pattern (Table 2). Other than the SKP2B F-box gene (*At1g77000*) and two zinc-finger genes (*At1g08050* and *At5g22920*), most of these genes function in intermediary metabolism or signaling. Metabolism genes encode the *ADS2* lipid desaturase (*At2g31360*), an acylphosphatase (*At5g03370*) that might function in glycolysis, the *TPK5* potassium channel gene (*At4g01840*), and the *FPS1* farnesyl diphosphate synthase (*At5g47770*) that is in the isoprenoid biosynthesis pathway. Signaling genes encode the *PP2C*-related protein phosphatase gene (*At3g17250*), which may negatively regulate protein kinase pathways, a phosphoinositide binding protein gene (*At3g22810*) potentially involved in lipid signaling, and the ACX1 acyl-CoA oxidase (*At4g16760*) that is in the jasmonate hormone biosynthesis pathway (Schilmiller et al, 2007). Available microarray data (http://seedgenenetwork.net; GEO accession no. GSE12404) shows that these genes are expressed primarily in endosperm. Two of twenty imprinted genes affected by *fie*, *At1g69900* and *At5g47770* (FPS1), display biallelic expression caused by a maternally inherited *dme* mutation, consistent with the role of DME in activating maternal expression of the core PRC2 components FIS2 and MEA. One possible explanation for a more limited effect of *dme* compared with *fie* mutations might be that *FIE* is a single copy gene required for all PRC2 molecule formation, whereas MEA and FIS2 are members of gene families, with other family members possibly able to compensate for a loss. PcG proteins related to MEA (SWINGER and CURLY LEAF) and FIS2 (VERNALIZATION 2 and EMBRYONIC FLOWER 2) are expressed in endosperm (http:// seedgenenetwork.net). These proteins can interchangeably form PRC2 and might provide redundant PRC2 functionality in a *dme* mutant background.

Table 2. 20 new imprinted genes with paternal allele silencing controlled by polycomb group protein (FIE, MEA) or indirectly by DME.
CXL_M = sum of CXL 'pure' and 'full' endosperm maternal score; **CXL_P**= sum of CXL 'pure' and 'full' endosperm paternal score; **LXC_M**= sum of LXC 'pure' and 'full' endosperm maternal score; **LXC_P**= sum of CXL 'pure' and 'full' endosperm paternal score; **CXL_M/P** and **LXC_M/P**=log (2) of maternal score divided by paternal score; **P**= Fischer's two tailed p value**; end_exp=** transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm; **fie**= result for *fie*XCol-0 endosperm using same statistic calculation; **mea**= result for meaXCol-0 endosperm using same statistic calculation; **dme**= result for *dme*XLer endosperm using same statistic calculation(please see method)

Table 2.

gene	annotation	CxL_M	CxL_P	LxC_P	LxC_M	CxL_M/P	LxC_M/P	P	end_exp	fie_M	fie_P	fie_M/P	P	fie_exp	mea_M	mea_P	mea_M/P	P	dme_Me_P	dme_M/P	P
AT1G08050	Zinc finger family	24	1	4	42	4.6	3.4	0.0	81	35	34	0.0	1.4E-08	476	17	18	-0.1	5.6E-07	25	4	2.6	0.4
AT1G21790	TRAM; LAG1 and CLN8 lipid-sensing domain containing protein	66	3	6	57	4.5	3.2	0.0	500	144	38	1.9	6.4E-04	2282	25	9	1.5	2.9E-03	222	16	3.8	1.0
AT1G69900	Actin cross-linking protein	61	4	10	64	3.9	2.7	0.0	75	1501	567	1.4	1.3E-06	3897	444	159	1.5	1.9E-05	767	322	1.3	0.0
AT1G76250	Unknown protein	140	3	1	36	5.5	5.2	0.0	44	36	31	0.2	1.2E-16	90	26	29	-0.2	8.0E-18	168	6	4.8	0.5
AT1G76820	Eukaryotic translation initiation factor, GTPase activity	45	4	12	50	3.5	2.1	0.0	71	72	38	0.9	5.5E-04	300	23	11	1.1	2.5E-02	155	15	3.4	0.2
AT1G77000	AtSKP2; homolog to human SKP2, may be involved in degradation of a CDK inhibitor	279	64	60	264	2.1	2.1	0.0	150	1602	874	0.9	2.0E-17	1926	551	324	0.8	1.5E-15	728	201	1.9	0.1
AT2G17990	Unknown protein	311	32	45	185	3.3	2.0	0.0	430	584	261	1.2	9.4E-15	2751	232	99	1.2	3.6E-09	456	57	3.0	0.3
AT2G31360	ADS2; delta 9 acyl-lipid desaturases	369	5	28	174	6.2	2.6	0.0	2448	243	107	1.2	5.9E-24	6208	84	79	0.1	2.9E-34	570	41	3.8	0.5
AT3G17250	PP2C; protein phosphatase 2C-related	33	8	8	64	2.0	3.0	0.0	220	147	76	1.0	9.1E-05	1073	57	28	1.0	1.9E-03	90	11	3.0	0.5
AT3G22810	Phosphoinositide binding protein involved in signal transduction	263	48	32	328	2.5	3.4	0.0	138	162	185	-0.2	1.1E-44	208	78	54	0.5	1.1E-13	116	20	2.5	0.4
AT4G01840	AtTKP5; K+ channel family	96	3	5	68	5.0	3.8	0.0	30	74	40	0.9	1.8E-11	46	16	6	1.4	1.8E-03	77	1	6.3	0.3
AT4G16760	ACX1; medium to long-chain acyl-CoA oxidase in jasmonate biosynthesis	898	92	86	645	3.3	2.9	0.0	76	1641	654	1.3	1.6E-47	396	420	172	1.3	3.0E-25	4928	678	2.9	0.0
AT5G02630	CAND6; Lung seven transmembrane receptor family	31	0	0	26	6.0	5.7	0.0	6	91	26	1.8	1.7E-05	51	13	6	1.1	1.2E-04	100	21	2.3	0.0
AT5G03370	acylphosphatase family	168	13	16	83	3.7	2.4	0.0	42	130	46	1.5	1.5E-05	103	43	31	0.5	3.5E-09	119	23	2.4	0.1
AT5G22920	Zinc finger protein	64	7	10	55	3.2	2.5	0.0	62	57	31	0.9	9.6E-05	144	36	24	0.6	3.7E-05	51	7	2.9	1.0
AT5G24460	Unknown protein	342	79	43	278	2.1	2.7	0.0	194	804	307	1.4	1.7E-08	770	173	102	0.8	1.4E-11	1087	192	2.5	0.4
AT5G42235	defensin-like (DEFL) family	2226	160	949	4438	3.8	2.2	0.0	1907	4325	1649	1.4	7.4E-83	5203	10052	4284	1.2	4.7E-156	6478	457	3.8	0.0
AT5G47770	FPS1; farnesyl diphosphate synthase1, involved in isopropenoind biosynthesis	45	6	15	111	2.9	2.9	0.0	66	382	155	1.3	2.9E-06	754	223	86	1.4	3.6E-05	202	70	1.5	0.0
AT5G53250	AGP22; arabinogalatan	58	2	0	28	4.9	5.8	0.0	24	21	53	-1.3	7.4E-23	91	3	4	-0.4	1.6E-04	70	11	2.7	0.0
AT5G53870	ENDODL1; early nodulin-like protein 1; carry electron in plasma	241	21	13	222	3.5	4.1	0.0	38	3781	2611	0.5	9.6E-63	1154	115	153	-0.4	4.2E-53	4519	30	7.2	0.0

Paternally imprinted genes controlled by DNA methylation and PRC2

A paternally expressed imprinted gene that is regulated by PRC2 as described in the above model will not be expressed in a paternal *met1* background and will display biallelic expression in a maternal *fie* background. We identified nine paternally expressed genes (Table 3) that are expressed primarily in the endosperm within the seed. Many of these genes encode potential regulatory proteins, including the transcription factor *AGAMOUS LIKE 92* (*AGL92*) gene, *YUCCA10* (*YUC10*), encoding a homolog of monooxygenase enzymes that synthesize auxin (Zhao et al, 2001), and two F-box genes (*At1g60410* and *At2g21930*). SUVH7 (*At1g17770*) is a SET domain protein related to SUVH4 (KRYPTONITE), SUVH5, and SUVH6 histone H3 lysine 9 (H3K9) methyltransferases required for CHROMOMETHYLASE 3 (CMT3)-mediated non-CG DNA methylation (Johnson et al, 2007), and it is the closest homolog of maternally expressed SUVH8 (Baumbusch et al, 2001). VARIATION IN METHYLATION 5 (VIM5) belongs to a protein family required for maintenance of CG methylation.

The silenced maternal alleles of all nine paternally expressed imprinted genes that we identified and *HDG3* are activated by a maternal *fie* mutation (Figure 1. and Table 3.). Paternal allele expression of seven genes (SUV*H7, AGL92, At1g60410 F-box, At2g21930 F-box, YUC10, At2g36560, and HDG3*) is reduced in endosperm fertilized with *met1* pollen, indicating that DNA methylation is required for WT paternal allele expression and likely prevents the establishment of repressive PRC2 on the paternal allele. For *YUC10*, the *met1* mutation activates expression of the maternal allele (Figure 1. and Table 3.), which was also reported for *PHE1* (Makarevich et al 2008). These results are consistent with the model proposed for regulation of *PHE1* imprinted expression. However, the imprinted status of *VIM5* is unaffected by *met1*, and little DNA methylation is present at or near the *VIM5* gene (Figure1 and Table 3.), suggesting that *VIM5* maternal allele repression may be mediated by PRC2 independent of DNA demethylation.

A *dme* mutation can theoretically lead to activation of the maternal allele, because retention of DNA methylation prevents binding of repressive PRC2 or PRC2 activity is compromised. Among the seven genes that show reduced paternal allele expression in a met1 mutant, we find that the *dme* mutation causes biallelic expression of four genes (*SUVH7, YUC10, At2g36560,* and *HDG3*) (Figure 1.). For two genes (*AGL92* and *At1g60410* F-box), maternal allele expression is activated; however, paternal allele expression still predominates (Table 3.), which may reflect the complex interactions between DNA methylation and PRC2 function that were reported for the regulation of PHE1 (Makarevich et al 2008).

Figure 1. Paternally expressed imprinted genes expression in mutant background.
(A) RT-PCR sequencing chromatographs at selected SNP regions measuring allele-specific expression in reciprocal crosses between Ler and Col ecotypes, in female *fie* Ler crossed to male Col for all genes, and in female Ler crossed to male *met*1-6 Col-*gl*.
(B) CG methylation profiles of genes shown in A and PHE1 are displayed. Genes and transposable elements oriented 5' to 3' and 3' to 5' are shown above and below the line, respectively. Gene models indicated in yellow color represent the imprinted genes shown in A. Arrows indicate 5' and 3' ends of imprinted genes where CG demethylation is detected in WT endosperm.
(C) Expression analysis by semiquantitative RT-PCR in WT reciprocal crosses between Ler and Col ecotypes and in female WT Ler crossed to male *met*1-6 Col-*gl*.
(D) Allele-specific expression of At1g48910 (YUC10) and At1g57800 (VIM5). RT-PCR analysis using F1 endosperm RNA isolated from ColXLer, LerXCol and LerX*met1*. For YUC10 RT-PCR products, HpaII enzyme cuts the Ler allele into a 212- and 77-bp band, whereas the Col allele is cut into 212-, 53-, and 24-bp bands. For VIM5 RT- PCR products, enzyme BsmI cuts the Col allele but not the Ler allele.

Figure 1.

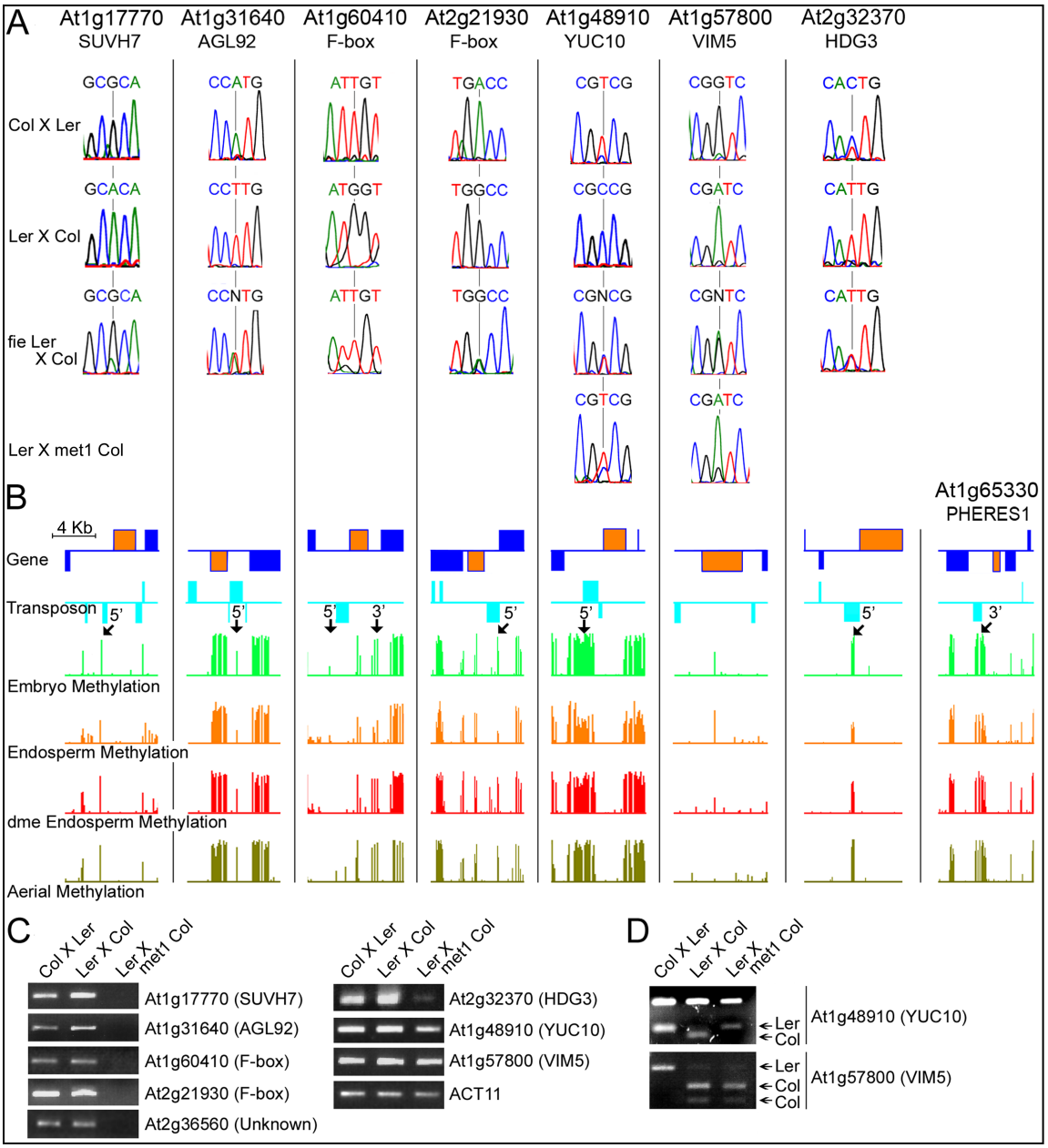

A
At1g17770
SUVH7
At1g31640
AGL92
At1g60410
F-box
At2g21930
F-box
At1g48910
YUC10
At1g57800
VIM5
At2g32370
HDG3
Col X Ler
Ler X Col
fie Ler X Col
Ler X met1 Col
GCGCA
GCACA
GCGCA
CCATG
CCTTG
CCNTG
ATTGT
ATGGT
ATTGT
TGACC
TGGCC
TGGCC
CGTCG
CGCCG
CGNCG
CGTCG
CGGTC
CGATC
CGNTC
CGATC
CACTG
CATTG
CATTG
B
At1g65330
PHERES1
4 Kb
Gene
Transposon
5'
3'
Embryo Methylation
Endosperm Methylation
dme Endosperm Methylation
Aerial Methylation
C
Col X Ler
Ler X Col
Ler X met1 Col
At1g17770 (SUVH7)
At1g31640 (AGL92)
At1g60410 (F-box)
At2g21930 (F-box)
At2g36560 (Unknown)
At2g32370 (HDG3)
At1g48910 (YUC10)
At1g57800 (VIM5)
ACT11
D
Ler
Col
At1g48910 (YUC10)
At1g57800 (VIM5)

Table 3. Paternally expressed imprinted genes imprinted pattern in mutant backgrounds.

CXL_M = sum of CXL 'pure' and 'full' endosperm maternal score; **CXL_P**= sum of CXL 'pure' and 'full' endosperm paternal score; **LXC_M**= sum of LXC 'pure' and 'full' endosperm maternal score; **LXC_P**= sum of CXL 'pure' and 'full' endosperm paternal score; **CXL_M/P** and **LXC_M/P**=log (2) of maternal score divided by paternal score; **P**= Fischer's two tailed p value**; end_exp=** transcriptional scores (number of reads per kb of sequence per 10 million aligned reads) for endosperm; **fie**= result for *fie*XCol-0 endosperm using same statistic calculation; **met1**= result for Col-0X*met1* endosperm using same statistic calculation; **dme**= result for *dme*XLer endosperm using same statistic calculation(please see method)

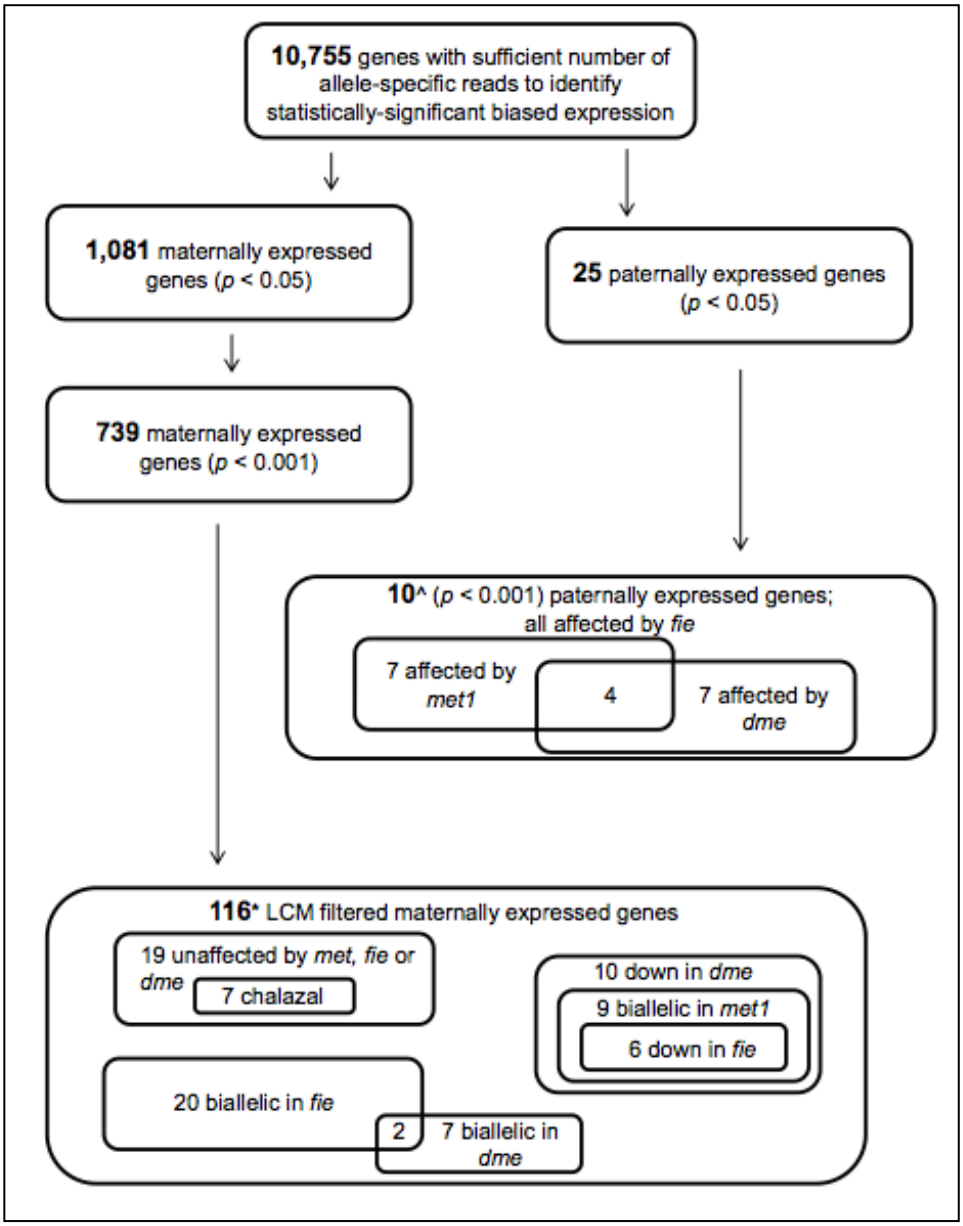
Table 3.

gene	annotation	CxL_M	CxL_P	LxC_P	LxC_M	CxL_M/P	LxC_M/P	P	end_exp	fie_M	fie_p	fie_M/P	P	fie_exp	met1_M	met1_P	met1_M/P	P	met1_exp	dme_M	dme_P	dme_M/P	P	dme_exp
AT1G17770	SUVH7; SU(VAR)3-9 homolog, a SET domain protein	3	5	13	5	-0.7	-1.4	0.0	4	368	213	0.8	0.0	159	0	0	0.0	1.0	0	295	254	0.2	0.0	121
AT1G31640	AGL92;AGAMOUS-LIKE 92, DNA binding	1	6	14	1	-2.6	-3.8	0.0	2	150	131	0.2	0.0	115	0	0	0.0	1.0	0	8	49	-2.6	0.7	14
AT1G48910	YUC10;N-dimethylaniline monooxygenase	36	70	279	30	-1.0	-3.2	0.0	78	124	131	-0.1	0.0	194	6	5	0.3	0.0	12	703	372	0.9	0.0	651
AT1G57800	VIM5; VARIANT IN METHYLATION 5	249	2818	3513	343	-3.5	-3.4	0.0	458	3734	3966	-0.1	0.0	1965	10	246	-4.6	0.0	127	1066	694	0.6	0.0	324
AT1G60410	F-box protein, SKP2 and cyclin-like domain	4	33	32	6	-3.0	-2.4	0.0	8	552	301	0.9	0.0	351	0	0	0.0	1.0	0	17	160	-3.2	0.4	56
AT2G21930	F-box associated domain	2	15	18	0	-2.9	-5.2	0.0	17	16	24	-0.6	0.0	109	0	0	0.0	1.0	0	0	7	-3.8	1.0	17
AT2G36560	Unknown protein	0	55	42	3	-6.8	-3.8	0.0	8	826	227	1.9	0.0	263	0	0	0.0	1.0	0	84	334	-2.0	0.0	78
AT4G11940	Unknown protein	1	6	15	1	-2.6	-3.9	0.0	3	124	141	-0.2	0.0	183	0	0	0.0	1.0	1	67	36	0.9	0.0	54
AT5G63740	Unknown protein	16	34	59	16	-1.1	-1.9	0.0	24	11420	12497	-0.1	0.0	15141	0	0	0.0	1.0	0	1110	1391	-0.3	0.0	1159

Genes imprinting controlled by novel mechanisms

Overall, among the 116 new maternally expressed genes, we found subgroups of new imprinted genes that behave as known imprinted genes in various mutant genetic backgrounds. Nine genes displayed biallelic expression in a paternal *met1* mutant background and were down regulated in a maternal *dme* background. These imprinted genes therefore followed the *FWA* model where DNA methylation in the 5'-flanking region silences transcription. Twenty genes displayed biallelic expression in a maternal *fie* mutant background, and therefore followed the *MEA* model where PRC2 maintains silencing of the paternal allele. However, there were subtle differences between what was expected for these genes and what was observed in different mutant backgrounds (Figure 2.) This may reflect the complexity of the known mechanisms. For example, DME promotes expression of two components of PRC2, *MEA* and *FIS2*, which in turn are needed for repressing activity of imprinted genes. In addition, some of the new imprinted genes discovered in this study may be regulated by novel mechanisms, or combinations of existing mechanisms. For example, *VIM5* partially follows the *PHE1* paradigm as the PRC2 complex affects its expression. However it differs from *PHE1* because DNA methylation is not present and has no impact on its imprinted status (Figure1). Therefore it is highly possible that *VIM5* imprinting mechanism is distinctive.

Figure 2. Flowchart diagram of imprinted gene identification.
The terminal diagrams for male- and female-expressed genes are modified Venn diagrams in which numbers within a box indicate the entire contents of the box to maintain consistency with values stated in the manuscript. For example, there are 20 maternally expressed genes with biallelic expression in *fie* and 7 with biallelic expression in *dme*; the two sets have two genes in common. ^Includes SUVH7 identified by Sanger sequencing and a previously known imprinted gene HDG3. *Includes genes that were close to the LCM cutoff, but were affected by *met1*, *fie*, or *dme*, and SUVH8 and JMJ15 identified by Sanger sequencing



Although we made every effort to filter out artifacts due to maternal RNA contamination, we cannot be absolutely sure that we were completely successful. Indeed, as we compared our maternally imprinted genes not affected by *met1*, *dme* or *fie* mutations to the microarray database of genes expressed in the Arabidopsis seed (http://seedgenenetwork.net/), we found that many of the genes we identified are highly expressed genes in the seed coat (Figure 3). To remove artifacts caused by RNA contamination from this class, we applied stringent criteria for selecting a subgroup of genes that are still imprinted in the three mutation backgrounds. That is, a gene had to be expressed at least 16-fold higher from the maternal than paternal genome in *met1, dme*, and *fie*—to identify 19 genes that were still clearly maternally expressed in all three mutant lines. These genes include *AGO3* (*At1g31290*), a MYB transcription factor gene (*At3g10590*), ARABIDOPSIS SKP1-LIKE E3-ligase component genes *ASK8* (*At3g21830*) and *ASK10* (*At3g21860*), and cytidine deaminase genes (*At4g29570* and *At4g29640*). Interestingly, microarray data was available for 16 of the 19 genes, and about half (7 genes) are highly expressed in the chalazal compartment of the endosperm and the chalazal seed coat (Figure 4A). The chalazal endosperm is adjacent to the chalazal seed coat, and therefore prone to contain seed coat contamination during hand dissection. To investigate this issue, we identified 48 genes from the microarray dataset with expression only in the chalazal seed coat, none of which are on our list (LCM-filtered; $P < 0.001$) of maternally expressed imprinted genes ($P = 0.0004$; Fisher's exact test), indicating that seed coat contamination is unlikely to account for our results. To further rule out contamination, we analyzed expression of four genes (*AGO3*, *MYB* transcription factor, *ASK8*, and *ASK10*) by RT-PCR in CxL LCM-derived endosperm, and all four are clearly maternally expressed. A mechanism of gene imprinting that does not require MET1, FIE, or DME is consistent with our data. Alternatively, there is active nutrient transfer from chalazal seed coat to the chalazal endosperm (Nguyen et al, 2000), suggesting that the mRNA for these genes might be synthesized in the maternal chalazal tissues and transported into the chalazal endosperm. Hence, these results could be explained by RNA transport from the chalazal seed coat to the chalazal endosperm. Recently, it was shown that for one of these seven genes, *At3g10590*, a MYB transcription factor, a transgene consisting of its promoter fused to GUS was expressed in the chalazal endosperm (Le et al, 2010). Hence, maternal expression of the *At3g10590* gene may not be controlled by RNA transport.

Figure 3. Unsupervised hierarchical clustering of expression patterns from LCM microarray data for maternally expressed genes not regulated by MET1, DME, or FIE (GEO accession no. GSE12404).
Hierarchical clustering was carried out using dChip 2008 software. Heat maps are based on signal intensities from the Arabidopsis LCM dataset using the average signal value of biological replicates.
Signals for preglobular, globular, heart, linear cotyledon, and mature green stages were averaged for each seed compartment. Coloration is based on the number of SDs away from mean signal of a gene in all compartments (Z score). Color intensity indicates deviation away from the mean signal for a particular gene: red, higher; blue, lower. Primary red and primary blue are +3 and −3 SDs, respectively. Heat maps have had vertical bars added to differentiate compartments from one another.
Genes not represented on the ATH1 microarray are AT1G54575, AT1G61090, AT1G61097, AT3G05700, AT3G54740, AT4G15080, AT4G16180, AT4G18320, AT5G02970, AT5G22200, and AT5G29860

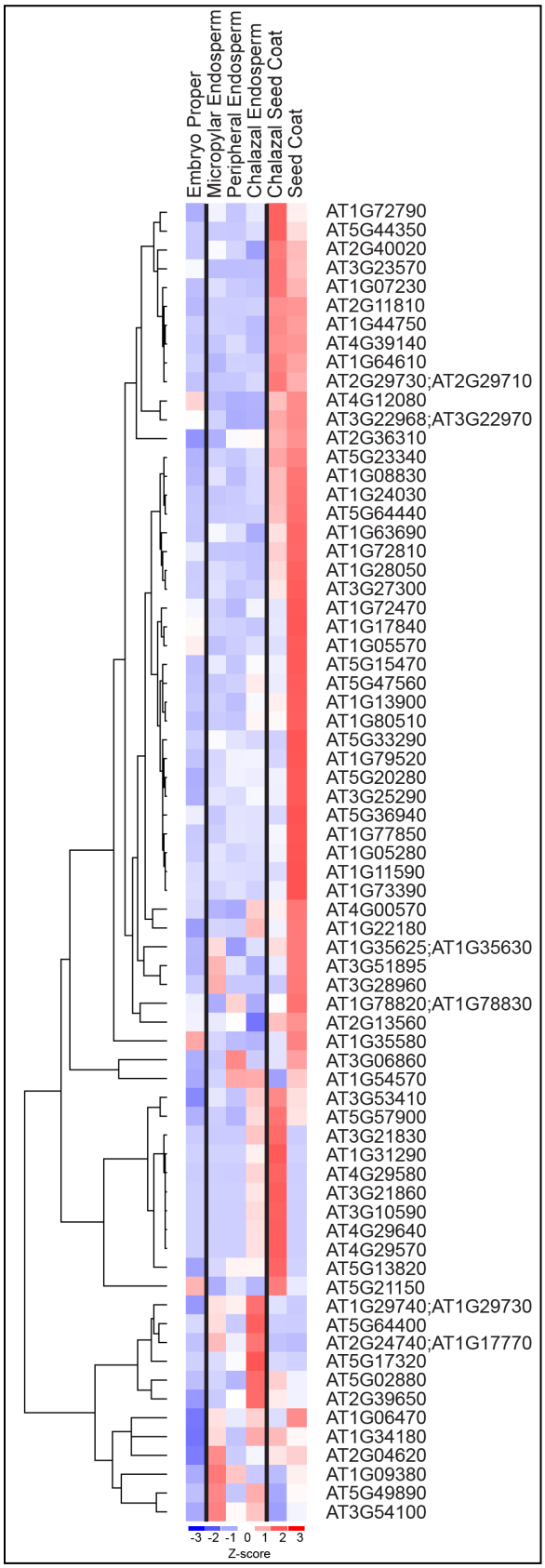
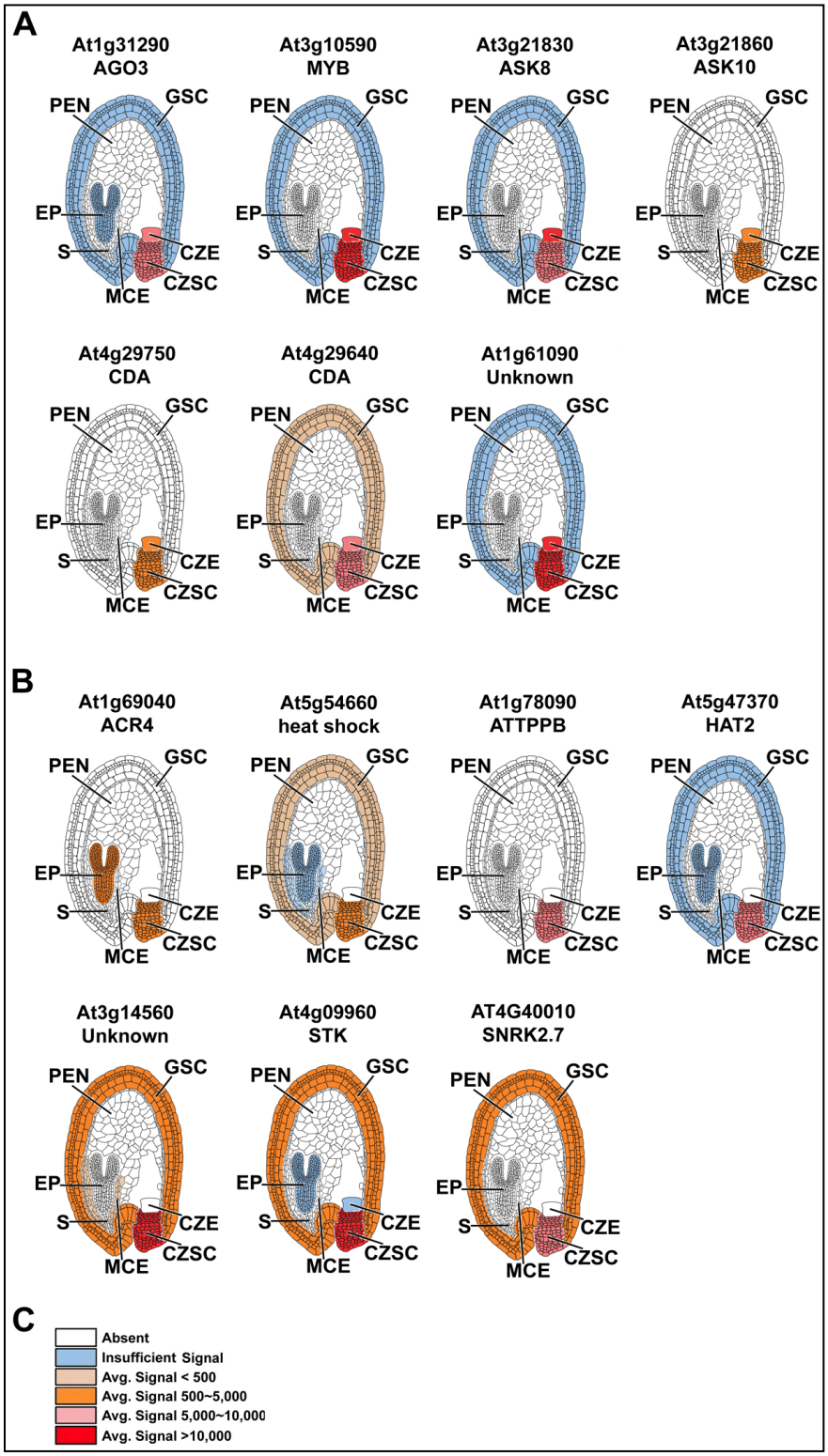

Embryo Proper
Micropylar Endosperm
Peripheral Endosperm
Chalazal Endosperm
Chalazal Seed Coat
Seed Coat
AT1G72790
AT5G44350
AT2G40020
AT3G23570
AT1G07230
AT2G11810
AT1G44750
AT4G39140
AT1G64610
AT2G29730;AT2G29710
AT4G12080
AT3G22968;AT3G22970
AT2G36310
AT5G23340
AT1G08830
AT1G24030
AT5G64440
AT1G63690
AT1G72810
AT1G28050
AT3G27300
AT1G72470
AT1G17840
AT1G05570
AT5G15470
AT5G47560
AT1G13900
AT1G80510
AT5G33290
AT1G79520
AT5G20280
AT3G25290
AT5G36940
AT1G77850
AT1G05280
AT1G11590
AT1G73390
AT4G00570
AT1G22180
AT1G35625;AT1G35630
AT3G51895
AT3G28960
AT1G78820;AT1G78830
AT2G13560
AT1G35580
AT3G06860
AT1G54570
AT3G53410
AT5G57900
AT3G21830
AT1G31290
AT4G29580
AT3G21860
AT3G10590
AT4G29640
AT4G29570
AT5G13820
AT5G21150
AT1G29740;AT1G29730
AT5G64400
AT2G24740;AT1G17770
AT5G17320
AT5G02880
AT2G39650
AT1G06470
AT1G34180
AT2G04620
AT1G09380
AT5G49890
AT3G54100
-3 -2 -1 0 1 2 3
Z-score

Figure 3.

Figure 4. Seed compartment-specific expression patterns from publicly available LCM microarray data (http://seedgenenetwork.net/arabidopsis).
(A) Seven genes maternally expressed in the chalazal seed coat (CZSC) and chalazal endosperm (CZE).
(B) Seven genes expressed in the CZSC.
(C) Color-coded bars represent relative hybridization signals shown in A and B. Additional abbreviations for seed compartments: **EP**, embryo proper; **GSC**, general seed coat; **MCE**, micropylar endosperm; **PEN**, peripheral endosperm; **S**, suspensor. Abbreviations for gene names: **CDA**, cytidine deaminase; **AGO3**, ARGONAUTE 3; **ASK**,
Arabidopsis skip1-like; **ACT**, act repeat 4 amino acid binding; **ATTPPB**, A. thaliana trehalose-6-phosphate phosphatase; **HAT2**, homeodomain A. thaliana transcription factor; **STK**, seed stick transcription factor; **SNRK2.7**, SNF1-related protein kinase 2.7

Figure 4.


A
At1g31290
AGO3
At3g10590
MYB
At3g21830
ASK8
At3g21860
ASK10
At4g29750
CDA
At4g29640
CDA
At1g61090
Unknown
B
At1g69040
ACR4
At5g54660
heat shock
At1g78090
ATTPPB
At5g47370
HAT2
At3g14560
Unknown
At4g09960
STK
AT4G40010
SNRK2.7
PEN
GSC
EP
S
CZE
CZSC
MCE
C
Absent
Insufficient Signal
Avg. Signal < 500
Avg. Signal 500~5,000
Avg. Signal 5,000~10,000
Avg. Signal >10,000

DISCUSSION

Imprinted genes regulated by DNA methylation often encode regulatory proteins.

We have determined the mechanisms that regulate many of the new imprinted genes reported in this study. This was accomplished by analyzing the effects of *met1*, *fie*, and *dme* mutations on imprinted genes expression in the endosperm. Although the function of these imprinted genes greatly varies, it is interesting to observe that imprinted genes controlled by DNA methylation tend to be regulatory, controlling the expression of genes or functioning in hormone signaling (Table1.) By contrast, the more numerous imprinted genes that are regulated by the PRC2 complex tend to control secondary metabolite production (Table 2). Why different epigenetic mechanisms regulate genes of distinct function is unknown. However, Polycomb group proteins function in maintaining rather than establishing the silent state (Schwartz and Pirrotta, 2008). It is possible that DNA demethylation in the central cell initially imprints genes encoding regulatory proteins that, in turn, activate or repress other genes, the transcriptional states of which are cemented by PRC2 activity. If so, mutations in imprinted genes directly regulated by DNA methylation would be predicted to affect the transcriptional status of other imprinted genes, particularly those dependent on PRC2.

Control of genome hypomethylation in the endosperm by *DME*, *VIM*, and *MET* genes.

We previously showed that virtually the entire A. *thaliana* endosperm genome, compared to the embryo genome, is hypomethylated in the CG context, and that this demethylation is largely dependent on DME (Hsieh et al 2009). Here, we show that the *VIM5* gene is primarily expressed from the paternal genome, *MET1* is down-regulated in endosperm, and *VIM* and *MET* genes are up-regulated in *dme*-deficient and *fie*-deficient endosperm (Table 3). These results suggest that DME promotes PRC2 activity, which in turn, suppresses *VIM* and *MET* gene expression. Thus, the global CG hypomethylation in the endosperm might be orchestrated by regulation of *VIM* and *MET* genes in addition to direct DME activity at specific sites. DME is expressed after cell proliferation is completed in the female gametophyte. Hence the suppression of *VIM* and *MET* gene expression would not be predicted to demethylate the entire central cell genome passively by DNA replication. However, after fertilization, when the primary endosperm nucleus rapidly proliferates, the reduced level of VIM and MET would result in passive, global endosperm demethylation. Because both maternal and paternal endosperm genomes would be subject to the same passive DNA demethylation, it is expected that the endosperm hypomethylation would occur equally on maternally- and paternally-derived chromosomes. Thus, one possible model would be that DME directly demethylates a number of discrete loci in the central cell, which results in maternal-specific DNA demethylation that regulates gene imprinting. By contrast, global demethylation is caused, at least in part, by DME-dependent *VIM* and *MET* repression, which results global hypomethylation of both maternal and paternal endosperm genomes compared to the embryo.

METHODS

Plant material and growth condition

Seed growth conditions were the same as in Chapter II. *met*1-*6* (Col-*gl*) heterozygous plants were previously described (Xiao et al, 2003). Heterozygous plants were self-pollinated and homozygous plants for *met*1-6 were selected by genotyping. *dme-2*(Col-*gl*) heterozygous plants and *fie*-1(Ler) heterozygous plants were previously described(Choi et al, 2002; Ohad et al, 1999).

Crosses and RNA isolation from tissue

Homozygous met1-6 plants pollen was used for fertilizing wild-type Ler flowers bud that had been emasculated as described in Chapter II. Silliques were collected at 8 to 10 DAP (day after pollination), when endosperm has sufficient viscosity to be hand-dissected. Heterozygous *dme*-2 and *fie*-1 plants' flowers were emasculated as previously described and pollinated with WT Ler and Col pollen respectively. Seeds bearing maternal *dme*-2 or *fie*-1 were selected based on the phenotype of the seed. These two mutations in the maternal allele result in seed abortion. At 6~8 DAP, endosperm are enlarged and embryo are not developing as WT. Therefore, seeds having a maternal *dme* or *fie* can be selected from WT. Endosperm and embryo from these mutant crosses were isolated and RNA extracted as previously described in Chapter II.

Library construction, and data processing algorithm

Library construction and read processing pipelines were similar to the process for WT crosses as described in Chapter II. Parental and transcriptional scores were obtained, and compared to the wild-type reciprocal cross data described in Chapter II. For a maternally expressed gene, it was considered to become biallelic in a mutant cross background if; i) a gene maternal expression is less than four-fold greater. This was simply the inverse of the cutoff used for selecting maternal expressed gene in Chapter II).

Also, ii) that gene maternal/paternal ratio in WT crosses must be greater compared to the maternal/paternal ratio in the mutant crosses and the difference between mutant and WT p-value must be below 0.001. Therefore, we will be able to take in consideration that each genes transcriptional amount and parental ratio range can differ in a large amount.

Here are few examples with hypothetical numbers to aid understanding of the reader: A gene with a low maternal read counts can be scored as imprinted if its paternal read count is lower by four fold. That is, a gene 'A' that is lowly expressed with a reads score maternal reads: paternal reads ratio of 5:1 can be considered as maternally expressed in WT as its maternal expression is five fold higher. In a mutant cross, gene A's expression can change, for example to a maternal reads: paternal reads ratio of 3: 1. Gene A will only be considered to become biallelic in a specific mutant background if the maternal reads (m): paternal reads (p) ration decreases to below 2.5:1. As another example, gene B is clearly imprinted since it is highly expressed and in WT its read count was m:p = 400:1. For gene B, it is considered to have become biallelic if in a mutant background the expression drops more than three-fold compared to WT, for example m:p = 150:50 since its WT ratio (400/1) is more than twice greater than its mutant ratio (150/50). Also, restricting the p-value difference between WT and mutant will filter any data with too much difference of expression in the two dataset. For example, consider a gene C with expression score in WT is m:p= 400:1 that drops dramatically in mutant cross to m:p = 3:1. This gene C will not be considered as becoming biallelic since the p value of the two data sets will

differ greatly due to the score number difference. Rather, gene C will be considered for further validation by RT-PCR as it might be a gene that becomes silenced in that specific mutant backround, rather than considered as becoming biallelic.

RT-PCR, Sanger sequencing and DNA methylation pattern analysis.

Primers set were designed to amplify cDNA fragments that at least have one SNP between Col and Ler in the amplicon. RNA from mutant cross F1 endosperm were converted to cDNA by random-primed first-strand synthesis (Ambion) or by using Ovation RNA-seq System (NuGen Technologies) according to manufacturer's protocol. RT-PCR products were analyzed for expression by gel-electrophoresis or for sequence by Sanger method as it was done in Chapter II. DNA methylation pattern in Col endosperm, embryo and *dme-2* endosperm were provided from Hsieh et al. (Hsieh et al, 2009). Tair8 Col sequence was used as scaffold for genes and data were processed using SignalMap program (Nimblegen).

Chapter IV
Maternal expression of siRNAs in the Arabidopsis seed

INTRODUCTION

RNA interference, a way of controlling gene expression via small RNAs

In the 1980's, it was stated that antisense RNA molecules effectively inhibit gene expression in several organisms, such as bacteria, amoeba and frog oocytes, and plants (Ecker and Davis, 1986). Consequently, it was discovered that introduction of transgenic antisense RNAs increased the degradation of endogenous messenger RNA (mRNA), suppressing expression post-transcriptionally in Petunia (Van Blokland et al, 1994). In the late 90s, it was shown that this machinery was also present in the worm, and that it involves RNA interference (RNAi), expanding our understanding of RNA as a regulatory molecule (Fire et al, 1998). The RNAi mechanism involves small non-coding RNAs of ~20-30 nucleotides (nt) that are associated with a large protein family present in eukaryotes, the Argonautes (AGO). As eukaryotes can synthesize small RNAs by diverse mechanisms that interact with different Argonautes, these complexes have myriad roles in regulating gene expression. However, all small RNA-mediated regulatory mechanisms have in common four key steps: small RNA biogenesis, loading of the small RNA onto an Argonaute protein, target recognition and function as transcriptional or post-transcriptional regulator.

Small RNAs in Arabidopsis: Biogenesis

Small RNAs are present in all eukaryotes and can be divided in two groups, those involved in RNAi and those that are not. Our understanding of the role of small RNAs is not yet complete and research has focused on further investigation into the biogenesis and role of the small RNAs that are involved in RNAi. Arabidopsis small RNAs that are involved in silencing pathways can be divided by their sources into exogenous or endogenous categories.

Endogenous small RNAs include microRNAs (miRNA, 21-22nt long) that are produced from intergenic regions or non-coding RNA. These miRNAs originate from pre-miRNAs that share similar features with pre-mRNA, like a 5'cap, 3' poly-adenylation (Poly A) tail and sometime introns. Since the first discovered worm's miRNA lin-4, which mutation disrupts the correct timing of cell differentiation, several miRNAs have been found to be important regulators of gene expression. In Arabidopsis, miRNAs play a critical role in development by, for example, controlling expression of PHABULOSA, TCP family or APETALA2 (Lai EC, 2003). Trans-acting siRNAs (tasiRNA) are 21nt long and produced from longer non-coding TASI transcripts (pre-tasiRNAs). TasiRNAs are produced by RNA dependent RNA Polymerase (RdRP), which recognizes a region primed by miRNA. TasiRNAs are specific to plants. Similar to miRNAs, they target non-identical mRNAs and hence are called trans-acting siRNAs (Vasquez 2004; Allen et al, 2005). Another class of 21nt long siRNAs, called natural antisense transcript siRNA (nat-siRNA) are derived from the transcription of overlapping regions in opposite orientation, therefore forming a complimentary dsRNA structure with the sense mRNA molecule. Generally this only occurs in certain physiological conditions, for example, if an organism is under environmental stress (Borsani et al, 2005). Finally, silent endogenous loci like retrotransposons, 5SrDNA and centromeric repeats produce 24-nt small RNAs that are called cis-acting siRNA (casiRNA) as they promote their own silencing via DNA or histone modification (Xie et al, 2004). This latter class also represents most of the endogenous siRNA population in Arabidopsis, and depends predominantly on a specific RNA polymerase for transcription, RNA polymerase IV, and therefore this specific group is also referred as PolIV dependent siRNA (p4-siRNA) (Zhang et al, 2007; Ghildiyal and Zamore, 2009; Mosher et al, 2009).

RNAi can be used as a research tool by scientists or a defense system by the organism. The RNAi pathway can be triggered by introduction of specially designed transgenes resulting in the production of double stranded RNA (dsRNA), long single stranded RNA (ssRNA) that has inverted repeats (therefore forming dsRNA), or transgenic ssRNAs that become dsRNA by RNA DEPENDENT RNA POLYMERASE 6 (RDR6), which have aberrant characteristics, for example the lack of 5' capping in transgenic RNAs (Brodersen and Voinnet, 2006; Gazzani et al., 2004). This mechanism is distinct from the endogenous RNAi pathway, but has the same effect and allows researchers to silence the transcription of genes of interest or initiate the plant defense system to viral infection due to the introduction of aberrant transcripts.

Small RNAs: Dicer to RISC loading

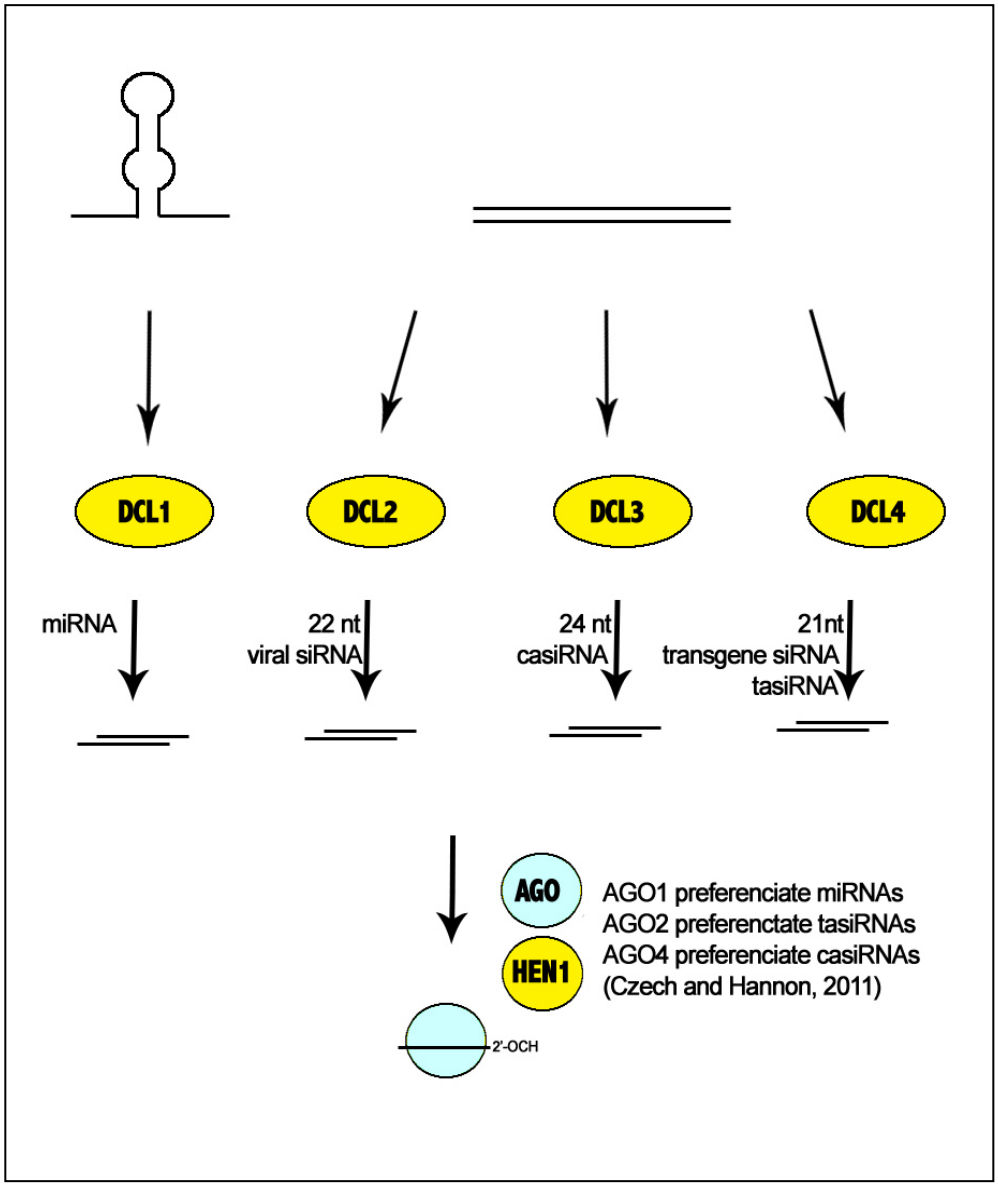


Figure 1. Small RNA processing by DCL and AGO family

When dsRNAs are formed, they are recognized and cleaved by the RNAse III enzyme Dicer family DICER-LIKE (DCL) resulting in small ~20 nt dsRNA with 3' end 2 nt overhangs. There are several members of the DCL family, with different functions. DCL1 is involved in the processing of miRNAs, DCL2 in 22 nt viral resistance related exogenous siRNA, DCL3 in endogenous 24-nt siRNA, and DCL4 in 21-nt exogenous siRNA and tasiRNAs biosynthesis (Dunoyer et al., 2005; Gasciolli et al, 2005; Henderson et al, 2006; Xie et al., 2004). The length of the small RNAs produced by the respective dicer enzymes is important as it seems to influence the role of small RNAs, as 21-nt siRNAs guide mRNA cleavage and 24-nt siRNAs are believed to relate to chromatin modification (Brodersen and Voinnet, 2006)

After the dicer enzyme cuts the dsRNA, HUA ENHANCER1 (HEN1) adds a methyl groups to both 3' ends to protect it from other modifications such as 3' uridylation. These double-stranded small RNAs are loaded with the help of diverse proteins into different ARGONAUTE proteins. One of the strands, the guide strand, is retained, and the other strand, the passenger strand, is discarded in a non-random manner using ATP. This Argonaute/guide strand complex binds with other proteins to form the RNA–induced silencing complex (RISC). By the guidance of the small RNA, RISC will recognize specific sites on genomic DNA or mRNA and function to control gene expression of specific loci by histone modification, DNA methylation or RNA degradation (B. Czech and GJ Hannon, 2011).

P4-siRNA, its role and expression in Arabidopsis

PolIV is needed for the production of the large majority of siRNAs (>90%) present in Arabidopsis (Zhang et al, 2007). PolIV is required for transgenic silencing repression (Herr et al, 2005; Dunoyer et al, 2007) and genotoxic stress resistance (Yao et al, 2010). Mutation of NRPD1A, the largest subunit of PolIV, dramatically reduces the amount of endogenous siRNAs. However, disruption of the P4-siRNA pathway does not seem to have any strong phenotypes involving obvious morphological or fertility defects.

Based on their expression pattern, p4-siRNAs can be divided in two different types, type II being expressed throughout all plant tissues and type I being specifically expressed in the reproductive part of the plant, flower and young silique. Although type II siRNAs are present in all vegetative tissues, specifically in the developing seed, both type I and II are predominantly expressed in the endosperm and seed coat rather than the embryo. Interestingly, in flower tissue, a subset of type I and II siRNAs require only PolIV subunit NRPD1A for expression (Mosher et al, 2009). Moreover, in the seed, by analysis of reciprocal crosses it was discovered that these siRNAs are only of maternal origin, and by genome-wide analysis it was suggested that this could be a trend that a large population of seed siRNAs follow (Mosher et al, 2009). The goal of our investigation was to determine if this maternal contribution of siRNAs in the seed is controlled by same genes involved in the imprinting of known genes in the endosperm.

RESULTS

Mutation of DNA methylation maintenance enzyme MET1 does not alter uniparental p4-siRNA expression in endosperm

It has previously been established that a mutation in MET1 and subsequent loss of most of the Arabidopsis CG methylation does not affect paternal silencing of the type I p4-siRNA at locus 08002 (Mosher et al, 2009). To determine if changes in non-CG methylation affect p4-siRNA paternal silencing, *drm1*/*drm2* (*drm*) and *cmt3* plants were crossed reciprocally with different WT ecotypes (Columbia-0 and Landsberg *erecta*) and the parental source of locus 08002 siRNA analyzed. In general, DRM is required for CHH methylation and CMT3 is required for CHG methylation. However, as DRM and CMT3 functions can overlap, a triple mutant, *drm1*/*drm2*/*cmt3* (*ddc*) was also reciprocally crossed (Figure 1) with different WT ecotypes. Small RNAs were detected specifically from maternal alleles in crosses between the wild-type ecotypes Columbia-0 (Col) and Landsberg erecta (Ler). Thus, demethylation of the paternal genome through the mutations in MET1, CMT3, DRM1/DRM2 was not sufficient to trigger accumulation of paternal p4-siRNAs. Furthermore, loss of all non-CG methylation in the triple mutant drm1 drm2 cmt3 (ddc) was insufficient to trigger paternal p4-siRNA accumulation, indicating that none of CMT3 ore DRM genes affect the uniparental expression of siRNA at 08002 loci.

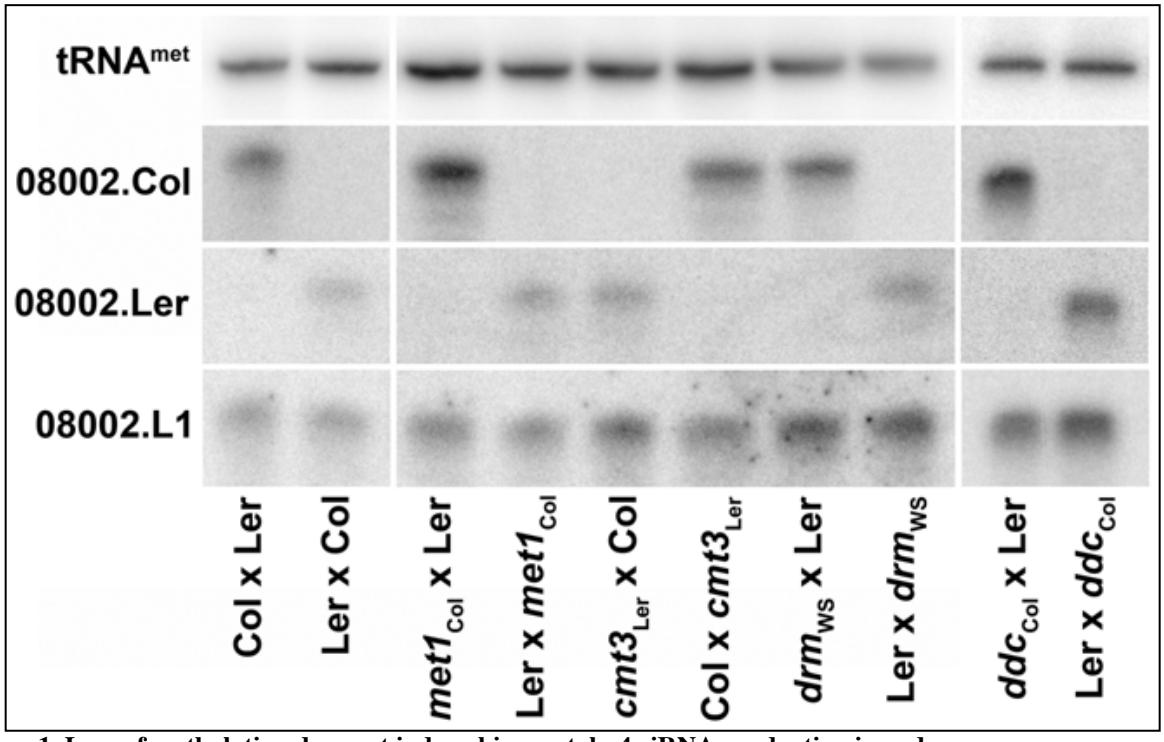


Figure 1. Loss of methylation does not induce biparental p4-siRNA production in endosperm.
Small RNAs were isolated from reciprocal crosses between wild type and DNA methyltransferase mutants at 5 days after fertilization. The parental origin of small RNA was determined with allele-specific small RNA probes (08002.Col and 08002.Ler). 08002.L1 hybridizes to small RNAs from both alleles and is a control for small RNA production at this locus; tRNAmet is a loading control.

As previously described in this thesis (Chapters 1 and 3) maternally expressed genes can be controlled via active demethylation, and the central cell expressed DME is responsible for activation of the maternally expressed MEA, FWA and FIS2. To establish the mechanism of activation of p4-siRNAs in maternal allele, transgenic lines with a constitutive promoter fused to DME protein was used to allow the expression of MEA in leaf tissue (Figure 2). If type I siRNA are expressed in reproductive tissue because of similar mechanism to MEA, type I siRNAs could be present in over-expressing transgenic lines leaf tissue as well.

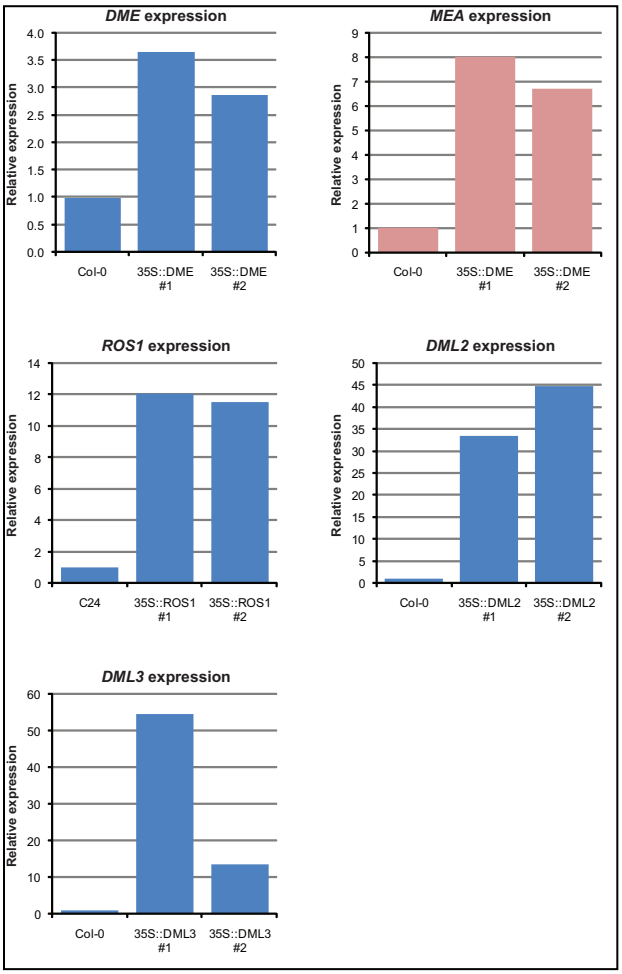


Figure2. Characterization of *DEMETER* family overexpression lines.
Transgenic lines expressing the four members of the *DEMETER* family behind the nearly constitutive *35S* promoter were assayed for transcript accumulation in leaves by quantitative reverse transcription-PCR. Overexpression of *REPRESSOR OF SILENCING* (*ROS1*) is in the C24 ecotype; all other constructs are in Columbia (Col-0). All graphs are mean values for 3 biological replicates and were normalized to GAPDH expression. *35S::DME* and *35S::ROS1* lines are homozygous; *35S::DML2* and *35S::DML3* are pooled samples of homozygous and hemizygous T2 individuals. Overexpression of *DEMETER* (*DME*) is weak, but sufficient to induce expression of *MEDEA* (*MEA*) in leaves (pink bars).

To assay the role of DME family in typeI and typeII siRNA expression, an RNA blot was performed with RNA isolated both floral and leaf tissue from these over-expressing transgenic lines (Figure 3). However, no expression changes were observed in these transgenic lines, showing that DME family ecotopic expression is not sufficient to trigger type I siRNA expression.

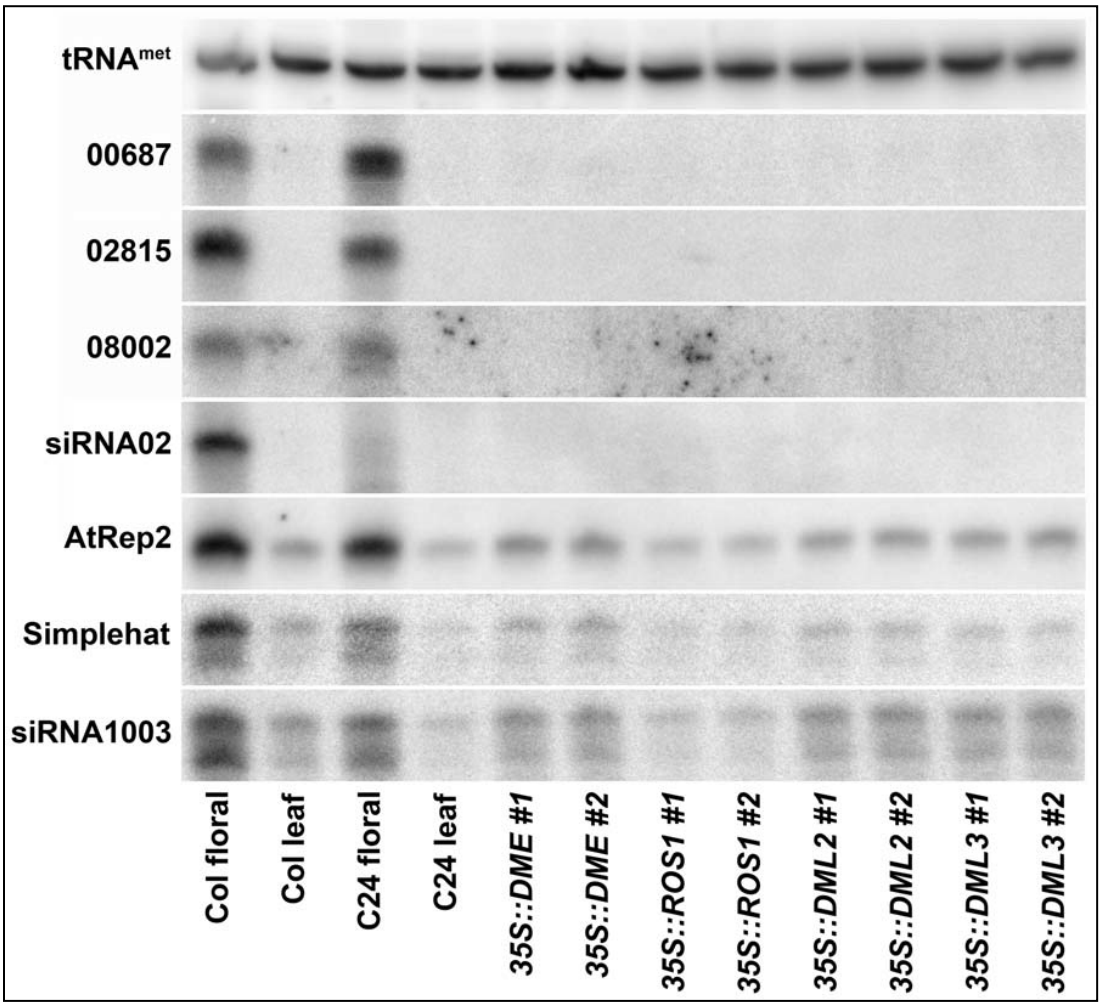


Figure 3. DEMETER family glycosylases are insufficient to induce vegetative expression of p4-siRNAs
typeI siRNA at locus 00687, 02815, 08002 and siRNA02; typeII siRNA AtReP2, Simplehat and siRNA1003

To assess the possibility that DME family proteins trigger maternal expression of locus 08002 siRNA in seeds, DME family overproducing lines were introduced as the paternal donor in crosses and compared to reciprocal crosses (Figure 4).

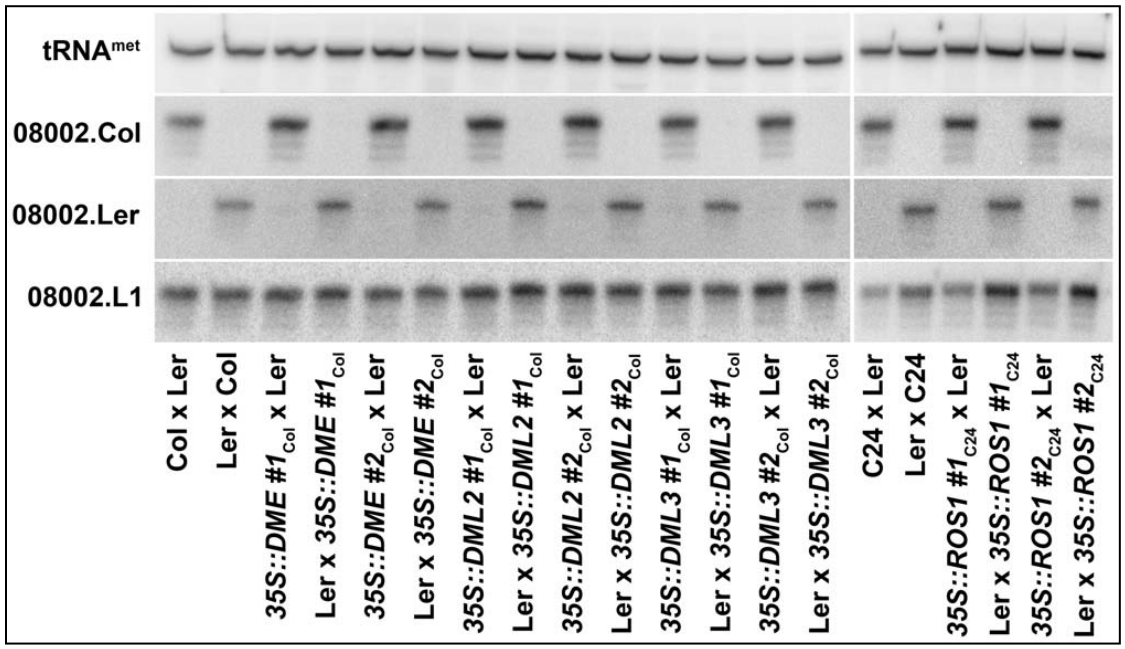


Figure 4. DEMETER family glycosylases do not trigger paternal expression of p4-siRNAs.
Small RNAs were isolated from inter-ecotype crosses between wild type and transgenic lines and parental origin of small RNA was determined as described in figure 1

Expression of DME or any DME family protein (DML2, DML3 or ROS1) in the male gametophyte under the strong 35S promoter did not release the paternal silencing of type I siRNA at locus 08002. In *dme* heterozygote plants crossed to WT, the seeds that inherit the maternal dme mutant gene abort their development. The aborted seeds were dissected and siRNAs accumulation was assayed compared to those that inherit the WT DME gene (Figure 5). As shown below, a maternal mutation of DME did not inhibit p4-siRNAs accumulation.

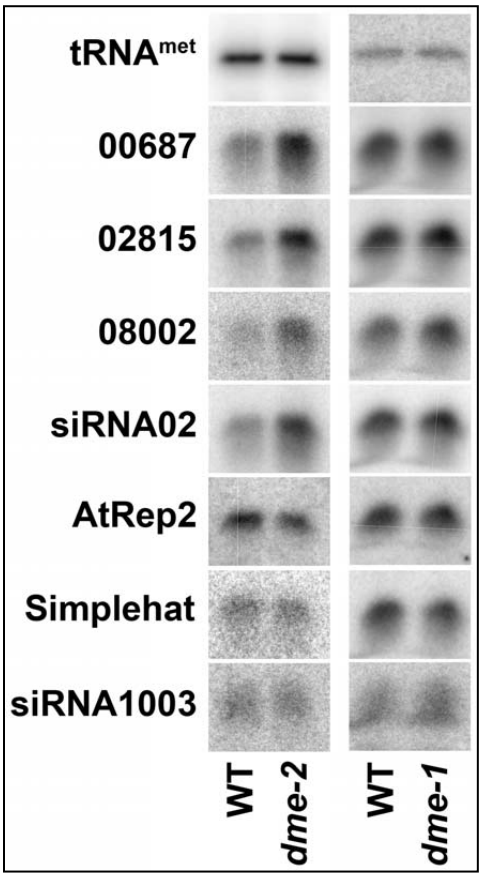


Figure 5. p4-siRNA accumuation in endosperm does not require DEMETER demethylation.
Left side: Developing (WT) or arrested (dme-) seeds were dissected from self-fertilized dme-2 heterozygous fruits 10–12 days after fertilization and small RNAs were extracted.
Right side: RNA was extracted from wild type and dme-1 homozygous fruits at 5 days after anthesis and small RNAs were extracted.

Various chromatin modifications do not affect p4-siRNA expression

As was shown in the previous chapter, loss of PRC2 activity can also affect the expression of imprinted genes. Therefore, to investigate the role of PRC2 or histone modification in the uniparental expression at locus 08002, reciprocal crosses with the FIE mutation, and also mutations in HISTONE DEACETYLASE 6 (HDA6), KRYPTONYTE (KYP), DECREASE IN DNA METHYLATION (DDM1) and MORPHEUS MOLECULE1 (MOM1) were analyzed for their effect on parent-of-origin expression of locus 08002 siRNA (Figure 2).

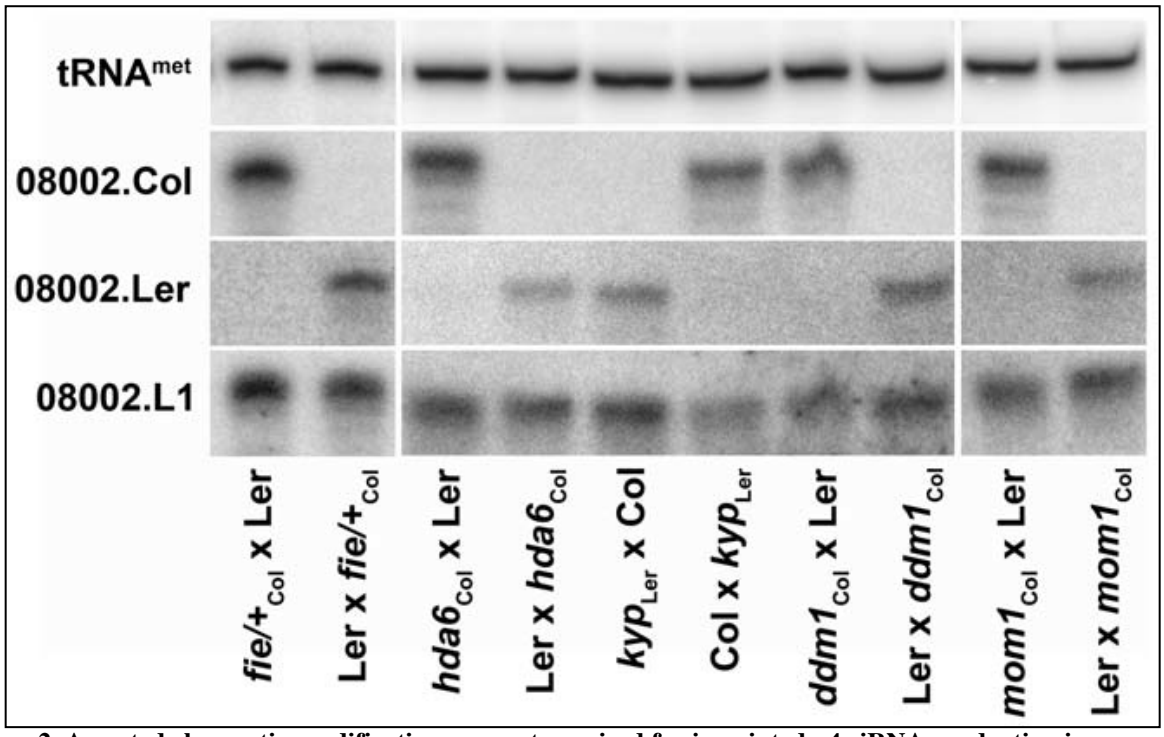


Figure 2. Assorted chromatin modifications are not required for imprinted p4-siRNA production in endosperm.
Small RNAs were isolated from inter-ecotype crosses between wild type and a histone modification mutant; parental origin of small RNA was determined as described in figure 1.

Accumulation of p4-siRNAs from paternal chromosomes was not induced when the PRC2 mutant FIE was transmitted paternally. Likewise, mutations in HDA6, the H3K9 methyltransferase KYP, the nucleosome remodeling proteins DDM1 and MOM1 did not affect uniparental expression of p4-siRNAs.

DISCUSSION

Previously, it has been shown that demethylation occurs in the endosperm but not in the embryo, due to active demethylation in the female gametophyte by DME expression in the central cell prior to fertilization (Gehring et al 2009; Hsieh et al, 2009). This active demethylation results in imprinted expression of genes, as shown in chapter III. It has also been observed that repeat elements flanking transposons are hypomethylated in endosperm compared to embryo (Gehring et al 2009; Hsieh et al, 2009).

Recent studies suggest that reproduction in Arabidopsis is characterized by siRNAs that may move from cell to cell or from tissue to tissue. Demethylation of central and vegetative cells may result in transposon activation and production of siRNAs that can be transported to egg and sperm cells, respectively, where transposon silencing will be reinforced (Schoft et al, 2009; Mosher et al, 2009; Slotikin et al, 2009). Likewise, it has been proposed that maternal demethylation in the central cell might activate siRNA formation in the endosperm that is derived from the maternal genome (Mosher et al, 2009). Finally, in the female gametophyte, siRNA movement from the gamete lineage to somatic companion cells has been suggested by a mechanism involving AGO9: Mutation of AGO9, only expressed in companion cells, resulted in a phenotype that changed the specification of the neighboring gamete cells via a pathway involving the RISC complex (Olmedo-Monfil et al, 2010).

The present model for p4-siRNA directed silencing in endosperm is that maternal specific marks trigger the transcription of these small RNAs via PolIV before fertilization, as maternal flower siRNAs are required for p4-siRNAs expression in seeds (Mosher et al, 2009). In the seed, these p4-siRNAs are loaded into effector complexes and bind transcripts generated by PolV to direct non-CG methylation at TEs (Wierzbicki et al, 2009). However, this is not straightforward, as it has been shown that loss of p4-siRNAs did not reactivate most of the TEs, and p4-siRNAs can also target DNA for demethylation (Mosher et al, 2008; Zheng et al, 2008). Moreover, we have shown that in contrast to the mechanisms known to regulate imprinted genes, the maternal contribution of p4-siRNA is not disrupted by factors that affect DNA methylation or histone modification. One of the possible explanations is that p4-siRNAs expression in the endosperm is controlled by mechanisms involving proteins that are not yet discovered. Accordingly, as outlined in Chapter III, we have found a group of maternally expressed genes whose maternal contribution is not disrupted by MET1, DME or FIE. As an alternative explanation, it is possible that p4-siRNAs are produced in the maternal seed coat and are transported in the endosperm.

METHODS

Plant growth conditions and genotyping

All plants were grown under standard conditions as described in above Chapters. Mutant alleles used were for Columbia-0 ecotype: *met1-1* (Kanke et al, 2003), *drm1-2, drm2-2, cmt3-1* (Chan et al, 2006), *hda6-9* (Havecker et al, unpublished), *ddm1-2* (Vongs et al, 1993), *mom1-2* (Amedeo et al, 2000) and fie (GABI362D08). For Ler (Landsberg erecta): *cmt3-1* (Lindroth et al, 2001) and *kyp-2* (Jackson et al, 2002). For Wassilewskijia: *drm1-1* and *drm2-2* (Cao and Jacobsen, 2002). The *dme-1* and *dme-2* of Columbia *glaburosa* ecotype were backcrossed to Ler (Choi et al, 2002).

Transgenic line generation

Total RNA from wild-type Columbia-0 leaf tissue was used to reverse transcribe and amplify full-length cDNAs of DML2 and DML3 with the following primers:

Gene	Forward primer	Reverse primer
DML2	5'-CACCATGGAAGTGGAAGGTGAAGTG-3'	5'-TCATTCCTCTGTCTTCTCTTTAGTTCTG-3'
DML3	5'-CACCATGTTGACAGATGGTTCACAACAC-3'	5'-CTATATATCATCATCACTCATAAACTTTGGCC-3'

PCR products were introduced into pENTR D-TOPO (Invitrogen) and the resulting entry vectors were recombined into pEARLEYGATE 202 (Earley et al, 2006). 35S::DML2 and 35S::DML3 constructs were stably transformed into wild-type Columbia-0 using standard protocols. Generation of 35S::DME and 35S::ROS1 are described elsewhere (Jullien et al, 2006; Agius et al, 2006).

Overexpression of DME-family glycosylases was verified with quantitative reverse transcription-PCR using QuantiFast SYBR Green One-Step RT-PCR Kit (Qiagen) and the following primers:

Gene	Forward primer	Reverse primer
DME	5'-ATTAAGGATTTCCTAGAACG-3'	5'-ATCCTAACTGCTATCCTTCC-3'
MEA	5'-GCTAATCGTGAATGCGATCC-3'	5'-AGAGAGTCCCATGTAAATGC-3'
ROS1	5'-GGGATGAACCATAAACTTGC-3'	5'-CAACTGGAAAGGCAAGATGG-3'
DML2	5'-GCTTGCCGAAAGAATCAAGG-3'	5'-CCGACATTCGTGTCAACAGG-3'
DML3	5'-GAATGGCTTCGAAATGCTCC-3'	5'-GGTACTCGAATAGTTGATGC-3'
GAPDH	5'-CTCCCTTGGAAGGAGCTAGG-3'	5'-GATGCATTGCTGATGATAGG-3'

RNA extraction and northern hybridizations

RNA was extracted from leaves using TRI® Reagent (Sigma-Aldrich) according to the manufacturer's protocol. RNA from crossed siliques or dissected seeds were extracted as follows: 5–6 siliques were frozen in liquid nitrogen and ground to a fine powder. 500 μL of room

temperature extraction buffer (100 mM glycine pH 9.5, 10 mM EDTA, 100 mM NaCl, 2% SDS) was added and once thawed, samples were further homogenized and placed on ice. Lysates were extracted once with cold Tris-saturated phenol (pH 8.0), twice with cold 25:24:1 Tris-saturated phenol:chloroform:isoamyl alcohol, and once with cold 24:1 chloroform:isoamyl alcohol before precipitation with sodium acetate and ethanol.

Small RNA was enriched from 30–50 μg total RNA with mirVana miRNA isolation columns (Ambion) according to the manufacturer's protocol. Small RNAs were resolved on a 7M urea/1X TBE/15% acrylamide gel (19:1 acrylamide:bisacrylamide) and transferred to Hybond N+ membrane (GE/Amersham). Membranes were UV-crosslinked before pre-hybridization in UltraHyb Oligo buffer (Ambion). Oligonucleotides were labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase and purified over an illustra MicroSpin G-25 column (GE/Amersham). After overnight hybridization with labeled oligonucleotides in UltraHyb Oligo buffer membranes were washed twice in 2X SSC, 0.1% SDS. Hybridization and washing was at 35°C. Membranes were exposed to phosphor-storage screens for detection of siRNAs.

Probe sequences are as follows (underlined bases are LNA):

$tRNA^{met}$ 5'-TCGAACTCTCGACCTCAGGAT-3';
08002.L1 5'-CCCATGGTCTCAAACACATCCTCG-3';
08002.Ler 5'-TCAAGTGAATCTTTAGCGTATGCT-3';
08002.Col 5'-AGTGAATCTAGAGATTTAGCGTAT-3';
00687 5'-GTTCCTCGTTCTACCCTCATACCT-3';
02815 5'-CCATGTCATTCCACCCATCAAAAG-3';
siRNA02 5'-GTTGACCAGTCCGCCAGCCGAT-3';
AtRep2 5'-GCGGGACGGGTTTGGCAGGACGTTACTTAAT-3';
Simplehat 5'-TGGGTTACCCATTTTGACACCCCTA-3';
siRNA1003 5'-ATGCCAAGTTTGGCCTCACGGTCT-3'.

All experiments were replicated with independent biological samples.

Chapter V
Summary

Double fertilization is a process that is unique to angiosperms. One of the products of this process is the endosperm. This tissue provides nutrients to the embryo in Arabidopsis, maize and rice and is viewed as the main tissue where parental specific expression, or imprinting, occurs. Before this study, only a handful of imprinted genes in the Arabidopsis endosperm had been identified and the mechanism of their allele specific expression had been studied in detail. It had been speculated previously that the amount of imprinted genes was higher, as more than 90 imprinted genes have been found in mammals (Bartolomei and Ferguson-Smith, 2011).

In Arabidopsis, imprinting of specific genes have been found individually based on several processes, for example, screening for mutations that have a parental specific effect, RT-PCR in the offspring of reciprocal crosses from different ecotypes, and reporter fusion protein expression of imprinted genes (e.g. *FWA, FIS2*, *MEA*) in mutant backgrounds (Grossniklaus et al., 1998; Soppe et al, 2000; Luo et al, 2000; Kohler et al., 2005). Based on the few known imprinted genes, it was suggested that the mechanism of imprinting involves MET1, DME and/or PRC2 proteins.

The recent advances in next generation DNA sequencing technology has allowed many facets of biological research to expand to the genome-wide scale. It has already been proven that this technique can be used in Arabidopsis for detecting SNPs in the entire genome or transcriptome (Ossowki et al, 2008, Lister et al, 2008). In this study, we used this next generation sequencing to assay the parental allele contribution to the RNA pool in the endosperm of reciprocal crosses between Col and Ler ecotypes as a way to understand seed formation and the imprinting phenomenon in the seeds. Our study has expanded the number of genes known to demonstrate parent of origin-specific expression. Based on SNP availability and sequencing depth we estimate to have been able to analyze roughly half (~10,000 genes) of the total endosperm transcriptome. However, taking into account our rather stringent statistical cutoff ($P < 0.001$) and filtering using LCM data, this estimate should be revised closer to 20-30% of the endosperm transcriptome. Consistent with this fraction, only 3 of the 10 previously described imprinted genes passed our filter. In our study, we found 114 genic RNAs in the endosperm that are preferentially from the maternal allele, and within these 114, 29 genes whose imprinting is controlled by DNA methylation or PcG activity. We also found 9 genes that are paternally biased in the endosperm. Thus, we expect that there may be about 30-50 paternally expressed endosperm genes, about 200 maternally expressed genes that are imprinted and controlled by DNA methylation or PcG activity, and potentially over 500 maternally biased candidates if imprinted genes are regulated by unknown mechanisms or possibly deposited from maternal tissue. Allele-specific gene expression is clearly a major phenomenon in plant endosperm, comparable to the extensive imprinting recently reported in mouse brain (Gregg et al, 2010).

The parental conflict theory (Feil and Berger, 2007) proposes that nutrient allocation is the driving force for the evolution of gene imprinting in mammals and plants. Although the effect on nutrient allocation of the imprinted genes described here is not yet known, the potential lines of conflict between maternal and paternal parents have significantly expanded. At the chromatin level, in addition to the previously discovered maternally expressed PRC2 component, paternally expressed proteins potentially silence target genes by promoting maintenance DNA methylation (*VIM5*) and H3K9 methylation (*SUVH7*), and maternally expressed genes potentially silence targets by regulating the small RNA pathway (*DRB2*), H3K9 methylation (*SUVH8*), H3K4 demethylation, and DRM2-mediated DNA methylation (*JMJ15*). Parental conflict may occur at the posttranslational level, mediated by degradation of specific proteins through the ubiquitin-26S proteasome system, which rivals transcription as a dominant

regulatory mechanism in A. *thaliana* (Viestra, 2009). Parental conflict may also take place through protein–protein interactions. *At1g59930*, a maternally expressed imprinted gene, encodes a truncated MADS box transcription factor that lacks the MADS box domain. Although it is unlikely to bind DNA, this protein may inhibit other MADS box transcription factors through dimerization (de Folter et al, 2005), including the activity of a close full-length relative, the paternally expressed PHE1. Imprinting of hormone synthesis (*YUC10* and *ACX1*) and response (*JLO* and *EIN2*) genes suggests that hormone action may also be involved in parental conflict.

Imprinted expression of genes with regulatory potential is frequently itself regulated by DNA methylation, whereas PRC2 regulates imprinting of genes that participate in cellular metabolism and signaling (Chapter III, Table 1. and 2.). Polycomb group proteins function in maintaining rather than establishing the silent state (Schwartz and Pirrota, 2008). It is possible that DNA demethylation in the central cell initially imprints genes encoding regulatory proteins that, in turn, activate or repress other genes, the transcriptional states of which are cemented by PRC2 activity. If so, mutations in imprinted genes directly regulated by DNA methylation would be predicted to affect the transcriptional status of other imprinted genes, particularly those dependent on PRC2.

In this study, we found a subset of genes and siRNAs that are maternally expressed and do not seem to be regulated by known imprinting regulatory factors (e.g. DME, FIE, MET1). It is possible that their expression is regulated by an unknown mechanism, or/and some of them might be transported from the maternally-derived chalazal seed coat to the offspring-derived chalazal endosperm. At least for one gene that is not regulated by DME, FIE, MET1 or MEA, *At3g10590*, a MYB translation factor, its promoter fused to a reporter was shown to be sufficient to trigger expression in the endosperm. (Le et al, 2010). This gene and gene promoter have been shown to be specific to the chalazal endosperm, a unique tissue that develops specifically from even the unicellular stage and differs cytologicaly from the other compartments of the endosperm (Boisnard-Lorig et al, 2001, Stangeland et al, 2003, Berger et al, 2007). Hence, there is a possibility that an imprinting mechanism using alternative proteins other than DME, MET1 or FIE might exist in discrete parts of the endosperm. Also, another possibility is the active movement of maternal mRNAs and siRNAs from the seed coat to the endosperm. Possible instances of siRNA movement from the companion cells to the female gamete has been already observed (Olmedo-Monfil et al, 2010). Also, there is evidence of intercellular movement of plant RNAs through plasmodesmata (Kehr and Buhtz, 2008), although it remains to be experimentally tested whether RNAs can navigate the apoplastic pathway that connects the chalazal seed coat and endosperm (Stadler et al, 2005).

DME was already known to be responsible for the imprinted expression of some imprinted genes (e.g. *FWA*, *MEA*). In previous studies, it was found that DME is also responsible for genome-wide demethylation in the endosperm, compared to the embryo, as well as hypomethylation at specific sites including repeat elements flanking transposons (Gehring et al 2009; Hsieh et al, 2009). It was suggested that these hypomethylation sites are a direct target of DME, which initiates the base excision repair (BER) pathway, excising 5-methylcytosine, and replacing it with unmethylated cytosine (Gehring et al, 2006). In this study, we found that the *VIM* family members that are usually highly expressed in vegetative tissue (*VIM 1, 2* and *3*), and have overlapping functions in maintaining global CG methylation (Woo et al, 2008), are not highly expressed in the endosperm along with MET1, which is the main CG methyltransferase known in Arabidopsis (Finnegan et a, 1996 Kankel et al, 2003) (Chapter II, Table 4.). Also, we found that although not controlled by MET1, *VIM5* is a paternally expressed gene whose mono-

allelic expression is controlled by DME and FIE (Chapter 3, Table 3. and Figure 1.). Therefore one possible explanation for the genome-wide demethylation resulting from DME activation in the endosperm is that it could be an indirect effect. Not only does the maternal endosperm have lower activity of VIMs compared to the embryo (Chapter II, Table 4.) but also *VIM5* is only expressed paternally, which might also have an effect on its activity. This is in contrast with the embryo, where DME is not active and therefore maternal *VIM5* is not silenced. As mitosis occurs after fertilization, the endosperm might have a less efficient mechanism for maintaining CG methylation compared to the embryo, therefore resulting in the global demethylation that was previously observed.

References:

Agius F, Kapoor A, Zhu JK (2006) Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation, Procceedings In National Academy of Science USA 103: 11796–11801

Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants, Cell 121:207-221

Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J (2000) Disruption of the plant gene MOM releases transcriptional silencing of methylated genes, Nature 405:203-206

Appleby A, Edwards D, Batley J (2009) New technologies of Ultra-High Throuhput Genotyping in Plants, Methods in Molecular Biology, Plant genomics 513:19-39

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 455:818–821

Aufsatz W, Mette MF, van der Winden J, Matzke AJ, Matzke M (2002) RNA-directed DNA methylation in Arabidopsis, Procceedings In National Academy of Science USA 99:16499-16506

Barlow DP (1993) Methylation and imprinting: from host defense to gene regulation? Science 260:309-310

Baroux C Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis, Genes & Development 20:1081–1086

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

Bartolomei MS and Ferguson-Smith AC (2011) Mammalian genomic imprinting, Cold Spring Harbor Perspective in Biology 1-17

Baumbusch LO, Thorstensen T, Krauss V, Fischer A, Naumann K, Assalkhou R, Schulz I, Reuter G, Aalen RB (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes, Nucleic Acids Research 29:4319-4333

Berger F, Fitz Gerald JN, Ingouff M (2007) Arabidopsis, as a model for understanding the basics of endosperm development 8:91-110

References:

Birve A, Sengupta AK, Beuchle D, Larsson J, Kennsion, JA, Rasmuson-Lestander A,, Müller J (2001) *Su(z)12*, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. Development 128: 3371-3379

Boisnard-Lorig C, Colon-Carmora A, Bauch M, Hodge S, Doerner P, Bancharel E, Dumas C, Haseloff J, Berger F (2001) Dynamic analyses of the expression of the HISTONE::YFP fusion protein in Arabidopsis show that syncytial endosperm is divided in mitotic domains, Plant Cell 13: 495-509

Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu JK (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis, Cell 123:1279-1291

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH (2001) Dnmt3L and the establishment of maternal genomic imprints, Science 294: 2536-2539

Brodersen P and Voinnet O (2006) The diversity of RNA silencing pathway in plants, Trends in Genetics 22:268-280

Brown R, Lemmon BE, Nguyen H, Olsen OA (1999) Development of endosperm in Arabidopsis thaliana, Sexual Plant Reproduction 12:32-42

Bureau M, Rast MI, Illmer J, Simon R (2010) JAGGED LATERAL ORGAN (JLO) controls auxin dependent patterning during development of the Arabidopsis embryo and root, Plant Molecular Biology 74:479–491

Cao X, Jacobsen SE (2002) Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing, Current Biology 12:1138-1144

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation, Current Biology, 13:2212-2217

Clark C, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu T, Fu G, Hinds G et al. (2007) Common sequence polymorphisms shaping genetic genetic diversity in Arabidopsis thaliana. Science 317:338:342

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

Chan SW, Henderson IR, Zhang X, Shah G, Ghien JS, Jacobsen SE (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in Arabidopsis, PLoS Genetics 2:e83

Chandra HS, Nanjundiah V (1990) The evolution of genomic imprinting, Development Supplement:47-53

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed development in Arabidopsis thaliana, Procceedings In National Academy of Science USA 94:4223-4228.

Choi Y, Gehring M , Johnson L , Hannon M , Harada JJ , Goldberg RB , Jacobsen SE , Fischer RL (2002) DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis Cell 110:33–42

Chen X (2009) Small RNAs and their roles in plant development, Annual Review of Cell and Developmental Biology 25:21–44

Cooper DW, VanderBerg JL, Sharman GB, Poole WE (1971) Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. Nature New Biology 230: 155-157

Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes, Nature Review Genetics 12:19-31

Day RC, McNoe L, Macknight RC (2007) Evaluation of global RNA amplification and its use for high-throughput transcript analysis of laser-microdissected endosperm, International Journal of Plant Genomics 2007: 61028

de Folter S, Immink RG, Kieffer M, Parenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, Davies B, Angenent GC (2005) Comprehensive interaction map of the Arabidopsis MADS Box transcription factors, Plant Cell 17:1424-1433

de la Casa-Esperón E, Sapienza C (2003) Natural selection and the evolution of genome imprinting, Annual reviews Genetics 37: 349-370

Delaval K, Feil R (2004) Epigenetic regulation of mammalian genomic imprinting, Current Opinion in Genetic Development 14:188-195

Doolittle W F, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution, Nature284:601-603

Dumas C, Rgowsky P (2008) Fertilization and early seed formation, C.R. Biologies 331: 715-725

Dunoyer P, Himber C, Voinnet O (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. Nature Genetics 37:1356-1360

Dunoyer P, Himber C, Ruiz-Ferrer V, Alioua A, Voinnet O (2007) Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways, Nature Genetics 39:848-856

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS (2006) Gateway-compatible vectors for plant functional genomics and proteomics. Plant Journal 45: 616–629

Ecker JR, Davis RW (1986) "Inhibition of gene expression in plant cells by expression of antisense RNA", Proceedings of the National Academy of Science USA 83 (15): 5372–5376.

Elgin SC, Grewal SI (2003) Heterochromatin: silence is golden, Current Biology, 23:R895-898

Feil R, Berger F (2007) Convergent evolution of genomic imprinting in plants and mammals, Trends in Genetics 23:192-199

Finnegan EJ, Peacock WJ, Dennis E (1996) Reduced DNA methylation in A*rabidopsis thaliana* results in abnormal plant development, Developmental Biology 93:8449-8454

Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans, Nature 391 (6669): 806–11

Fitz Gerald JN, Hui PS, Berger F (2009) Polycomb group-dependent imprinting of the actin regulator AtFH5 regulates morphogenesis in Arabidopsis thaliana. Development 136:3399–3404

Fiume E, Fletcher JC (2012) Regulation of Arabidopsis Embryo and Endosperm Development by the Polypeptide Signaling Molecule CLE8, Plant Cell 24:1000-1012

Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs, Current Biology 15:1494-1500

Gazzani S, Lawrenson T, Woodward C, Headon D, Sablowski R (2004) A link between mRNA turnover and RNA interference in Arabidopsis, Science 306:1046-1048

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases, Annual Reviews Biochemistry 74:481-514

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

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

Gehring M, Missirian V, Henkoff S (2011) Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds, PLos ONE 6:e23687

Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ (1999) Multiple DNA methyltransferase genes in Arabidopsis thaliana, Plant Molecular Biology 41:269-278

Ghildiyal M and Zamore PD (2009) Small silencing RNAs: an expanding universe, Nature Reviews Genetics 10 94:108

Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain, Science 329: 643-648

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal Control of Embryogenesis by *MEDEA*, a *Polycomb* Group Gene in *Arabidopsis*, Science 280:446-450.

Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F (2004) Identification of new members of Fertilization Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Development 131:2971-2981

Haig D, Trivers R (1995) The evolution of parental imprinting: a review of hypotheses, Genomic Imprinting: Cause and Consequence pp. 17-28

Haig D (1997) Evolutionary theories of genomic imprinting, Frontier in Molecular Biology 18:211-238

Hall JG (1990) Genomic imprinting: review and relevance to human diseases, American Journal of Human Genetics 46:857-873

Henckel A, Nakabayashi K, Sanz LA, Feil R, Hata K, Arnaud P (2009) Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals, Human Molecular Genetics 18:3375-3383

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning, Nature Genetics 38:721:725

Henderson IR, Jacobsen SE (2008) Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading, Genes Development 22:1597–1606

Hennig L, Derkacheva M (2009) Diversity of Polycomb group complexes in plants: Same rules, different players?, Trends in Genetics 25:414–423

Henikoff S, Comai L (1998) A DNA Methyltransferase Homolog With a Chromodomain Exists in Multiple Polymorphic Forms in Arabidopsis Genetics 149: 307-318

Herr AJ , Jensen MB , Dalmay T , Baulcombe D (2005) RNA Polymerase IV Directs Silencing of Endogenous DNA, Science 308:118–120

Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation, Plant Cell 12:357–369

Holliday, R (1990) Genomic imprinting and allelic exclusion, Development, Supplement:125-129

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell 104:829-838

Hsieh TF, Ibarra CA, Silva P, Zemach A, Esched-Williams L, Fisher RL, Zilberman D (2009) Genome-wide demethylation of the Arabidopsis endosperm. Science 324:1451-1454

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in Arabidopsis endosperm, Proc Natl Acad Sci USA 108:1755-1762

Hurst LD, Atlan A, Bengtsson BO (1996) Genetic conflicts, The Qarterly Review of Biology 71:317-364

Ingram R, Charrier B, Scollan C, Meyer P (1999) Transgenic tobacco plants expressing the Drosophila Polycomb (Pc) chro- modomain show developmental alterations: possible role of Pc chromodomain proteins in chromatin-mediated gene regulation in plants, Plant Cell 11: 1047–1060

Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. Current Biology 19: 1677-1681

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase, Nature 416: 556-560

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation, Current Biology 17:379-384

Jullien PE , Kinoshita T , Ohad N , Berger F (2006) Maintenance of DNA Methylation during the Arabidopsis Life Cycle Is Essential for Parental Imprinting, Plant Cell 18:1360–1372

Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, Berger F (2008) Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis, PLos Biology, 6:e194

Kacem S, Feil R (2009) Chromatin mechanisms in genomic imprinting, Mammallian Genome 20:544-556

Kafri T. Ariel M, Brandes M, Shemer R, Urven L, McCarrey J (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line, Genes Development 6:705-714

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature 429:900-903

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) Arabidopsis *MET1* cytosine methyltransferase mutants, Genetics 163:1109-1122

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T (2003) Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis, Current Biology 13:421-426

Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y, Okano M, Li E, Nozaki M, Sasaki H (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse, Human Molecular Genetics 16: 2272-2280

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

Kehr J, Buhtz A (2008) Long distance transport and movement of RNA through the phloem, Journal of Experimental Botany 59:85-92

Kerk NM, Ceserani T, Tausta SL, Sussex IM, Nelson TM (2003) Laser capture microdissection of cells from plant tissues, Plant Physiology 132:27-35.

Kessler SA, Grossniklaus U (2011) She's the boss: signaling in pollen tube reception, Current Opinion in Plant Biology 14:622-627

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation, Science 303:521-523

Kishimoto N, Sakai H, Jackson J, Jacobsen SE, Meyerowitz EM, Dennis ES, Finnegan EJ (2001) Site specificity of the Arabidopsis METI DNA methyltransferase demonstrated through hypermethylation of the superman locus, Plant Molecular Biology 46:171-183

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1, Genes & Development 17: 1540-1553

Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting, Nature Genetics 37:28:30

Kouzarides T (2007) Chromatin modifications and their function, Cell 128: 693-705

Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB, Fischer RL (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis, Procceedings In National Academy of Science USA 96:4186-4191

Kurihara Y, Matsui A, Kawashima M, Kaminuma E, Ishida J, Morosawa T, Mochizuki Y, Kobayashi N, Toyoda T, Shinozaki K, Seki M (2008) Identification of the candidate genes regulated by RNA directed DNA methylation in Arabidopsis, Biochemical and Biophysical Research Communications 376:553–557

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins, Nature 410:116-120

Lai EC (2003) microRNAs: Runts of the genome assert themselves, Current Biology 13:R925-936

Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamuro JK, Harada JJ, Goldberg RB (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors, Procceedings In National Academy of Science USA 107:8063-8070

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation, Science 292: 2077-2080

Lindroth AM, Shultis D, Jasencakova Z, Fuchs J, Johnson L, Schubert D, Patnaik D, Pradhan S, Goodrich J, Schubert I, Jenuwein T, Khorasanizadeh S, Jacobsen SE (2004)

Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3, EMBO Journal 23: 4286-4296.

Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. Plos Biology 1:420-428

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

Lucifero D, Mann MR, Bartolomei MS, Trasler JM (2004) Gene-specific timing and epigenetic memory in oocyte imprinting, Human Molecular Genetics 13:839-849

Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J. and Chaudhury, A.M. (1999) Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of National Academy of Science USA 96: 296-301

Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds, Proceedings of National Academy of Science USA 97:10637-10642

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm, Plos Genetic, 7:e 1002125

Mansfield SG, Briarty LG, Erny S (1991) Early embryogenesis in Arabidopsis thaliana.I. The mature embryo sac, Canadian Journal of Botany 69:447-460

Ma H, Sundaresan V (2010) Development of flowering gametophytes, Current Topics in Developmental Biology, 91:380-404

Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Köhler C (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis, EMBO Reports 7:947-952

Malagnac F, Bartee L, Bender J (2002) An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation, EMBO Journal 21:6842-6852

McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37: 179-183

Moores T, Mills W (2008) Evolutionary theories of imprinting-Enough already! Genomic Imprinting, 9:116-122

Morgan HD, Santos F, Green K, Dean W, Reik W (2005) Epigenetic reprogramming in mammals, Human Molecular Genetics 14: R47-58

Mosher RA, Schwach F, Studholme D, Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis, Procceedings In National Academy of Science USA 105:3145-3150

Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis, Nature 460:283-286

Murayama A, Sakura K, Nakama M, Yasuzawa-Tanaka K, Fujita E, Tateishi Y, Wang Y, Ushijima T, Baba T, Shibuya K, Shibuya A, Kawabe Y Yanagisawa J (2006) A specific CpG site demethylation in the human interleukin 2 gene promoter is an epigenetic memory, Embo Journal 25, 1081-1092

Nguyen H and Brown RC (2000) The specialized chalazal endosperm in Arabidopsis thaliana and Lepidium virginicum (Brassicaceae), Protoplasma 212:99-110

Nodine MD and Bartel (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos, Nature 482: 94-97

Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL (1999) Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization, Plant Cell 11:407-416

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, Cell 99: 247-257

Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP. (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis, Nature 464:628-632

Orr HA (1995) Somatic mutation favors the evolution of diploidy, Genetics, 139:1441-1447

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads, Genome Research 18: 2024–2033.

Parenicová L, de Folter S, Kieffer M, Homer DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell 15: 1538-1551

Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005) TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat, Plant Cell 17:2966-2980

Pirrota V (1998) Polycombing the genome: PcG, trxG, and chromatin silencing, Cell 93:333-336

Reik W, Constancia M, Fowden A, Anderson N, Dean W, Ferguson-Smith A, Tycko B, Sibley C (2003) Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. Journal of Physiology 547: 35-44.

Renfree MB, Papenfuss AT, Shaw G, Pask AJ (2009) Eggs, embryos and the evolution of imprinting: insights from the platypus genome. Reproduction, Fertility and Development 21:935-942

Schilmiller AL, Koo AJK, Howe GA (2007) Functional diversification of acyl-coenzyme A oxidases in jasmonic acid biosynthesis and action, Plant Physiology 143:812–824

Schwartz YB, Pirrotta V (2008) Polycomb complexes and epigenetic states, Current Opinion in Cell Biology 20:266-273

Scott RJ, Spielman M, Bailey J, Dickinson HG (1998) Parent-of-origin effects on seed development in Arabidopsis *thaliana*, Development 125:3329-3341

Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen, Cell 136: 461-472

Solter, D (1988) Differential imprinting and expression of maternal and paternal genomes, Annual Review of Genetics 22:127-146

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

Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang MS, Jacobsen SE, Schubert I, Fransz PF (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis, EMBO Journal 21:6549-6559

Sørensen MB, Chaudhury AM, Robert H, Bancharel E, Berger F (2001) Polycomb group genes control pattern formation in plant seed, Current Biology 11:277-281

Stadler R, Lauterbach C, Sauer N (2005) Cell-to-cell movement of green fluorescent protein reveals post-phloem transport in the outer integument and identifies symplastic domains in Arabidopsis seeds and embryos, Plant Physiology 139:701-712

Stageland B, Salehian Z, Aalen R, Mandal, AO, Olsen (2003) Isolation of GUS marker lines for genes expressed in Arabidopsis endosperm, embryo and maternal tissue, Journal of Experimental Botanics 54: 279-290

Stepanova AN, Alonso JM (2009) Ethylene signaling and response: Where different regulatory modules meet, Current Opinions in Plant Biology 12:548–555

Sun X, Shantharaj D, Kang X, Ni M (2010) Transcriptional and hormonal signaling control of Arabidopsis seed development, Current Opinion in Plant Biology 13:611-620

Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308: 548-550

Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J (2003) Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin, Procceedings In National Academy of Science USA 100: 8823-8827

Tariq MA, Kim HJ, Jejelowo O, Pourmand N (2011) Whole-transcriptome RNAseq analysis from minute amount of total RNA, Nucleic Acid Research, 39:e120

Thomas, JH (1995) Genomic imprinting proposed as a surveillance mechanism for chromosome loss, Procceedings In National Academy of Science USA, 92:480-482

Tompa R, McCallum CM, Delrow J, Henikoff JG, van Steensel B, Henikoff S (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3, Current Biology 12:65-8

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, Nature 408:796:815

Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, Feil R (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes, Nature Genetics 36:1296-1300

Van Blokland R, Van der Geest N, Mol JNM, Kooter JM (1994). "Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover". Plant Journal 6: 861–77

Varmusa S, Mann M (1994) Genomic imprinting—defusing the ovarian time bomb, Trends in Genetics 10:118-123

Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, Mallory AC, Hilbert JL, Bartel DP, Crété P (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs, Molecular Cell 16:69-79.

Vierstra RD (2009) The ubiquitin-26S proteasome system at the nexus of plant biology, Nature Reviews Molecular Cell Biology 10:385–397

Villar CB, Erilova A, Makarevich G, Trösch R, Köhler C (2009) Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats, Molecular Plants 2: 654-660

Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) Arabidopsis thaliana DNA methylation mutants. Science 260:1926-1928

Wagschal A, Feil R (2006) Genomic imprinting in the placenta, Cytogenentic Genome Research 113:90-98

Wang D, Tyson MD, Jackson SS, Yadegari R (2006) Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis, Procceedings In National Academy of Science USA 103:13244–13249

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

Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C (2010) H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation, PLos Genetics 6:e1001152

Weiss A, Keshet I, Razin A, Cedar H (1996) DNA demethylation in vitro: Involvement of RNA. Cell 86,:709-718

Werner A, Mette MF, Matzke AJM, Matzke M (2004) The role of MET1 in RNA-directed *de novo* and maintenance methylation of CG dinucleotides, Plant Molecular Biology, 54:793:804

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin, Nature Genetics 41:630-634

Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MT, Spillane C, Nordborg M, Rehmsmeier M, Köhler C (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm, PLos Genetic 7: e1002126

Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis, PLoS Genetic 4: e1000156

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

Xiao W, Brown RC, Lemmon BE, Harada JJ, Goldberg RB, Fischer RL (2006) Regulation of seed size by hypomethylation of maternal and paternal genomes, Plant Physiology 142:1160-1168

Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC (2004) Genetic and Functional Diversification of Small RNA Pathways in Plants, Plos Biolgy 2:E104

Yao Y, Bilichak A, Golubov A, Blevins T, Kovalchuk I (2010) Differential sensitivity of Arabidopsis siRNA biogenesis mutants to genotoxic stress, Plant Cell Reports 29:1401-1410

Zemach A, Li Y, Wayburn B, Ben-Meir H, Kiss V, Avivi Y, Kalchenko V, Jacobsen SE, Grafi G (2005) DDM1 binds Arabidopsis methyl-CpG binding domain proteins and affects their subnuclear localization, Plant Cell 17:1549-1558

Zeller G, Clark RM, Schneeberger K, Bohlen A, Weigel D, Ratsch G (2008) Detecting polymorphic regions in Arabidopsis thaliana with resequencing microarrays. Genome Res 18:918-929

Zhao Y, Christensen SK, Fankhauser C, Cashman JR, Cohen JD, Weigel D, Chory J (2001) A role for flavin monooxygenase-like enzymes in auxin biosynthesis, Science 291:306-309

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE (2007) Role of RNA polymerase IV in plant small RNA metabolism, Procceedings In National Academy of Science USA 104:4536-4541

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis, Nature 455:1259-1262

Zhu J, Kapoor A, Sridhar VV, Agius F, Zhu JK (2007) The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in Arabidopsis, Current Biology 17:54–59