A System to Study Chromatin Dynamics Through Pollen Development

By

Joseph Paul Calarco

Under the Supervision of Dr. Robert A. Martienssen
Watson School of Biological Science, Cold Spring Harbor Laboratory

Chapter 2

FACS-based Purification of Arabidopsis Microspores, Sperm Cells and Vegetative Nuclei

Chapter 2 Appendix

Detailed Protocol of Purification Method

Chapter 3

Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA.

Introduction

Results

Discussion

Chapter 4

DDM1 purification, antibody design and attempted crystallizations

Chapter 5

Future Implications and Ongoing Work

Chapter 1
Gametogenesis and Chromatin Dynamics in *Arabidopsis thaliana*

** Sections of this chapter are taken from **Calarco JP** and Martienssen RA, Genome reprogramming and small interfering RNA in the *Arabidopsis* genome. Curr Opin Genet Dev. 2011 Apr;21(2):134-9.

Gametogenesis in Arabidopsis thaliana

Plants alternate between a haploid gametophytic and diploid (or sometimes polyploid) sporophytic stage. Unlike mammals, whose germ cells are derived and set aside early in development, angiosperms maintain a population of undifferentiated diploid stem cells which differentiate only late in development to form functional germ cells. The female and male diploid meiotic precursor cells, respectively called the Megaspore Mother Cell (MMC) and Microspore Mother Cell (PMC) only undergo meiosis upon completion of flower development and are found in the carpel and anther of a mature flower respectively (Boavida et al., 2005). Figure 1 provides a schematic of their development, which is outlined in detail in the following sections.

Macrogametogenesis – Formation of the Female Germline

Female gamete development begins in Arabidopsis with the differentiation of a diploid megaspore mother cell that undergoes meiosis to form four haploid megaspores. Three of these megaspores die, and the surviving megaspore undergoes three rounds of mitosis without cell division resulting in a multinucleate structure called a syncytium, which subsequently cellularizes (Gehring and Henikoff, 2007). This structure represents the mature female gametophyte and is composed of a seven-celled structure consisting of three antipodal cells, one central cell, two synergid cells, and one egg cell (Figure 1). Each cell in this structure is haploid, with the notable exception of the diploid central cell, which is formed by the fusion of two haploid nuclei just before fertilzaition (Boavida et al., 2005).

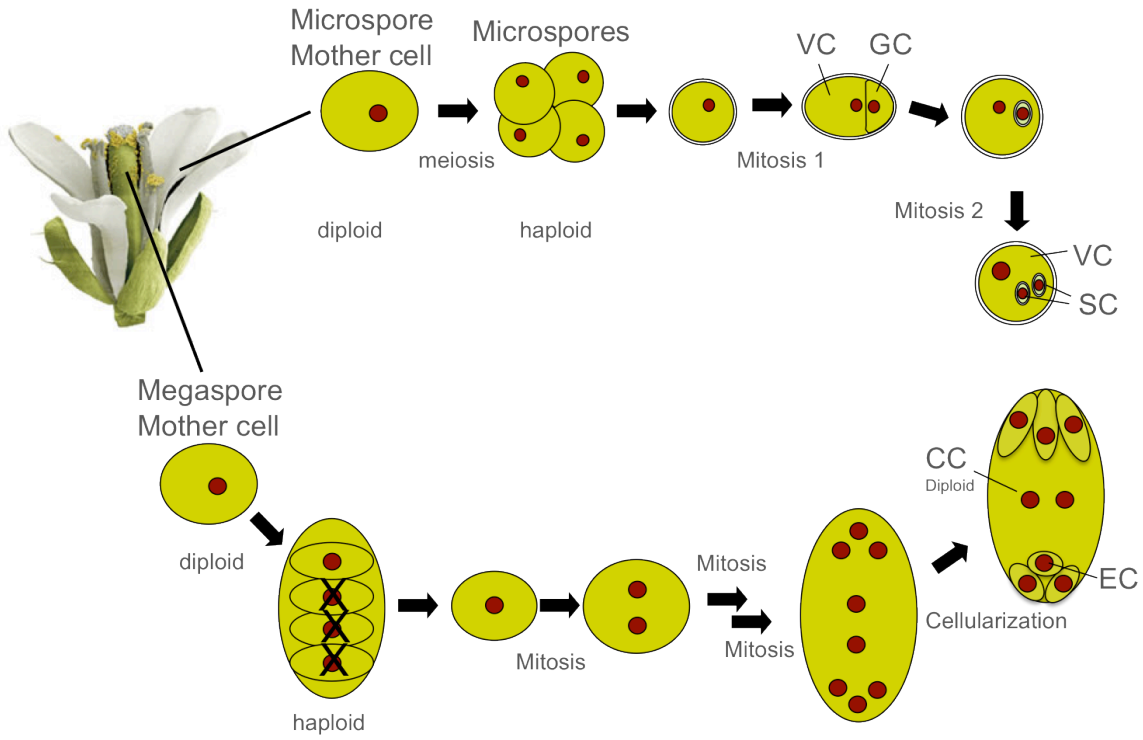


Figure 1 – Development of plant germ cells. Male germ cell development (top) proceeds from the diploid cell called the Microspore Mother Cell (MMC) which undergoes meiosis to form four haploid microspores. Each microspore undergoes two rounds of mitosis, first making a bicellular pollen grain consisting of a vegetative cell (VC) and a generative cell (GC). The generative cell undergoes a second round of mitosis that results in two mature sperm cells embedded in the cytoplasm of the vegetative cell, which arrests after mitosis I. Female gametophyte development begins with the differentiation of a diploid megaspore mother cell that undergoes meiosis to form four haploid megaspores. Three of those cells die and the remaining one undergoes three rounds of mitosis without cellularlization. After mitosis is completed cellularization occurs and the mature female germunit (containing the diploid central cell and haploid egg cell) is formed,.

Microgametogenesis – Formation of the Male Germline

The male gametophyte is formed in the anther of a developing flower: a microspore mother cell undergoes meiosis, resulting in the formation of four haploid microspores, each with their own cellular membrane contained within a developing pollen cell wall. As gametogenesis progresses further, each microspore undergoes a round of mitosis, called Pollen Mitosis I (PM I), which gives rise to bicellular pollen. At this stage the pollen grain is composed of the larger Vegetative Cell (VC) and the smaller Generative Cell (GC) which is embedded within the cytoplasm of the vegetative cell (Calarco and Martienssen, 2011). The generative cell then undergoes a second round of mitosis, called Pollen Mitosis II (PM II) which produces two haploid sperm cells (SC) (McCue et al., 2011). DAPI staining of the cells that comprise the germline through development illustrates that the vegetative nucleus (VN) of the pollen grain has a completely decondensed chromatin structure, including regions that were previously heterochromatic, in contrast to the tightly condensed chromatin found in sperm cell (SC) nuclei (Figure 2). The microspore is less compacted than the sperm cells, but more structured than the VN, as noted through the still visible chromocenters.

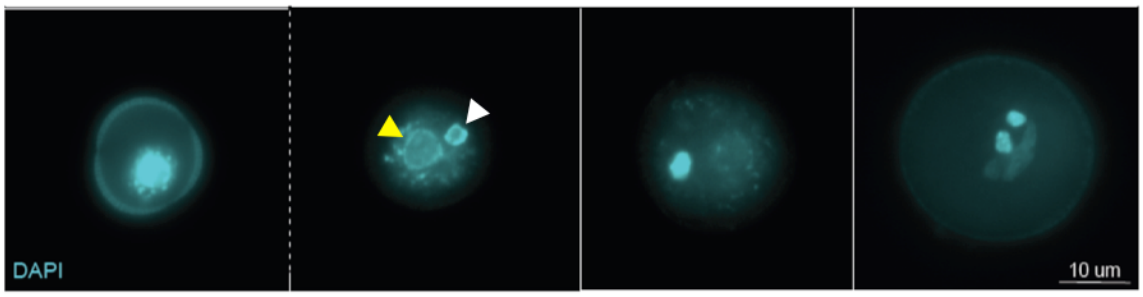


Figure 2 – Chromatin dynamics through Pollen development. DAPI staining of the pollen grain though development underscores the gross changes in chromatin structure observed through its maturation. Compared to the microspore (left most panel) the structure of the nucleus of both the generative cell (white arrow) and the vegetative cell (yellow arrow) begin to condense and decondense respectively (middle panels). By the time they are fully mature they proceed further to a condensed (sperm cells) or decondensed (vegetative cell) state.

Fertilization and Embryo Development

Flowering plants have evolved a unique reproductive mechanism, with the most curious and distinctive trademark being that of double fertilization. Essential to this are the two male sperm cells, and the two female gamete cells –the egg cell and central cell. One sperm cell fertilizes the egg cell, resulting in a diploid embryo while the other sperm cell fuses with the diploid central cell, which gives rise to the triploid endosperm. In Arabidopsis, sperm cells can fertilize either the central cell or egg cell (Hamamura et al., 2011). The role of the endosperm is to support the developing embryo in the seed during the plants early development. Many of the most drastic changes in gene expression take place through germ cell differentiation and embryo development (Borges et al., 2008) and these changes are reflected in the vast alterations in nuclear architecture that occur. These dramatic changes follow from a biological perspective, as these cells have to prepare their genetic information and potentially reset any chromatin modifications they may have accumulated over their development for transmission to the next generation.

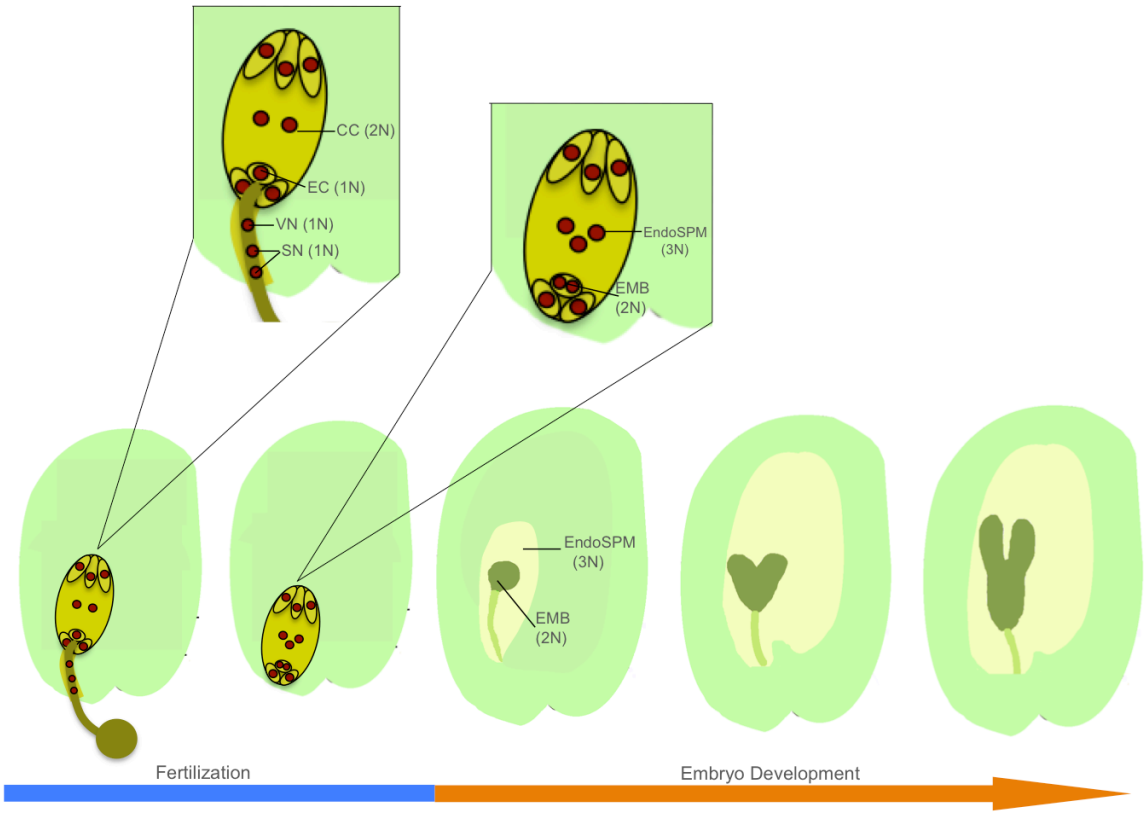


Figure 3 – Schematic Representation of Embryo development. The first portion of the arrow (blue) shows the growing pollen tube which delivers the sperm cells to both the egg and central cell. After fertilization, the embryo (2N, dark green) continues to divide and grow. The endosperm (3N, yellow) continues to grow as well.

Chromatin Modification

What is Chromatin and what is a Nucleosome?

In order to properly and functionally package an organism's DNA into its nucleus, an amalgamation of nucleic acid and protein called chromatin is formed. The manner in which DNA is packaged is important to a number of biological processes and must be properly maintained as chromatin structure influences gene expression, effects how an organism sustains DNA damage and acts as a structural element for mitosis/meiosis (Dinant et al., 2008), among other purposes. The organization of chromatin can be stratified into different levels, with the simplest being the wrapping of DNA around a collection of histone proteins which forms a complex called a nucleosome (Luger et al., 1997b) (Luger et al., 1997a). Depending on how compacted these nucleosomes are, chromatin can be defined as 'heterochromatic', meaning it is densely compacted in interphase, or it can be defined as 'euchromatic' meaning it is less densely packaged. How a particular piece of chromatin becomes heterochromatic or euchromatic depends on covalent chemical modifications made to both the histone tails and/or the DNA bases (Wako and Fukui, 2010).

A single nucleosome is comprised of an octamer of specific histone proteins. Each octamer is composed of two tetramers. Each tetramer is made up of four histone subunits, H2A, H2B, H3 and H4 (Figure 4).

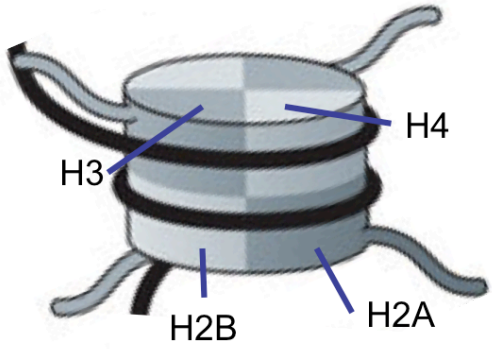


Figure 4 – The Nucleosome. A single nucleosome is comprised of an octamer of specific histone proteins. Each octamer is composed of two tetramers. Each tetramer is made up of four histone subunits, H2A, H2B, H3 and H4. (Modified from Castel and Martienssen 2013)

In Arabidopsis, the study of nucleosome composition is further complicated through the presence of variants of the aforementioned histone subunits (Ingouff and Berger, 2010). A histone octamer then forms a structure reminiscent of a spool around which DNA can be wrapped in a stereotypical manner. Usually, a piece of DNA will make 1.75 turns around a histone octamer, comprising 146 bp of DNA (Luger et al., 1997b). This combination of histone octamer and DNA is called a nucleosome (Figure 4).

Chromatin Remodelling

Chromatin remodeling refers rather broadly to any number of changes that effect chromatin architecture. Usually, this process directs the access of regulatory proteins and the transcriptional machinery to specific regions of the genome, thereby influencing gene expression. The consequences of chromatin remodeling are typically represented in the form of covalent modifications made to histone tails, which are performed by specific families of proteins (Allis et al., 2007). Closely intertwined with enzymes that modify histones are the so called chromatin remodeling proteins, which act as 'motor' proteins that utilize ATP and facilitate the dynamism of chromatin fibers (Ryan and Owen-Hughes, 2011). These chromatin-remodeling complexes can remove, replace or shift histone subunits/ histone octamers /nucleosomes as a means of facilitating access to particular portions of the genome. The proper function of these nucleosome modifiers and remodelers play key roles in any number of essential processes from transcription, replication, mitosis, meiosis, chromosomal segregation and DNA damage and repair.

Essential to chromatin dynamics are histone tail modifications. These comprise covalent modifications that are made to the specific residues on histone subunits, known as the histone tails. These tails are a stretch of peptide on the N-terminus of histone subunits (most commonly H3) which are outside of the nucleosome core. Many covalent chemical modifications mediated

by a host of different enzymes have been reported on histone tails through mass spectroscopy with the function of many of them still unknown (Tan et al., 2011). These enzymatic modifications include, but are certainly not limited to acetylation and methylation. These particular modifications affect the manner in which histone octamers will interact with surrounding DNA and hence 'loosen' or 'tighten' the DNA surrounding the histones. Lysine and arginine residues, present at stereotypical positions in the tails, are some of the best studied with regards to methylation. Significant correlations can be made between the methylation status of particular lysine residues and transcriptional activation. For example mono-, di- or trimethylation of H3K4 (where H3 refers to the particular histone subunit in the octamer being modified, and K4 refers to the specific residue, lysine at position 4, and the prefix mono, di and tri refer to the number of a methyl groups attached) and H3K36 are typically correlated with transcriptional activation, whereas methylation at di- or trimethylation of H3K9, and di- or trimethylation of H3K27 are correlated with transcriptional repression. H3K9 methylation specifically, is a hallmark of constitutive heterochromatin (Rea et al., 2000; Strahl and Allis, 2000)

These correlations have led to the idea of a universal histone code, where these specific, stereotypical modifications made to histone tails recruit effector proteins that recognize the specific marks on the tails via particular domains specialized for binding a specific moiety. It is these proteins which 'read' the histone code and are thus recruited to act upon or alter chromatin structure as a means to activate or repress a particular sequence of DNA and influence transcription (Jenuwein and Allis, 2001).

	Lysine			
Modification	**H3K4**	**H3K9**	**H3K27**	**H3K79**
Dimethylation	Activating	Repressing	Repressing	Activating
Trimethylation	Activating	Repressing	Repressing	Activating/Repressing

Table 1 – Summary of some Histone tail methylation modifications. Represented here are some of the most prevalent Histone H3 tail modifications, specifically Lysine 4, 9, 27 and 79. H3K4 tri methylation is found at actively transcribed promoters, typically just after the transcriptional start site. H3K9me3 is typically found at constitutively repressed loci, while H3K27me2/3 are found at facultatively repressed genes.

Chromatin Remodelers

Much work has been done to study the 'engines' of chromatin remodeling. To this end, it has been observed that most chromatin remodelers exist as part of large, multisubunit complexes that utilize the energy from the hydrolysis of ATP to restructure chromatin by shifting or even ejecting nucleosomes. One important family is the SWI/SNF remodelers, of which DDM1 (Decrease in DNA Methylation 1) is an example. The first alleles of *ddm1* were isolated in a forward genetic screen looking for aberration in DNA methylation (Vongs et al., 1993). Mutations in DDM1 showed a 70% reduction in DNA methylation, mostly occurring in repeat sequences and regions rich in transposable elements (Miura et al., 2001) (Singer et al., 2001) (Brzeski and Jerzmanowski, 2003). In addition to defects in DNA methylation, *ddm1* plants also show defects in H3K9me2 (Lippman et al., 2004, Gendrel et al., 2002), suggesting multiple roles for DDM1 in directing chromatin structure.

DNA methylation

Cytosine methylation is a stable covalent modification made to carbon 5 of cytosine, which influences gene expression, imprinting (monoallelic expression) and X chromosome inactivation (Law and Jacobsen, 2010). In plants, cytosine methylation occurs in various sequence contexts, with the most common being the symmetric CG methylation. Other possible methylation contexts include symmetric CHG methylation (where H=A,T or C) and asymmetric CHH (Henderson and Jacobsen, 2007). Averaging methylation genome wide, CG is the most abundant with 24% of CG sequences being methylated, while CHG stands at 6.7% and CHH at 1.7% (Cokus et al., 2008).

In mammalian systems, cytosine methylation patterns are controlled by a family of proteins called DNMT (DNA methyltransferase), with DMNT3 performing de novo methylation

and DNMT1 performing maintenance CG methylation in a replication dependent manner, recognizing a hemimethylated daughter strand. In Arabidopsis, DNA methylation is dependent principally on three specific enzymes, each the main component in establishing/maintaining methylation in the three sequence contexts introduced above. MET1, was first identified in the same screen for DNA methylation defects that uncovered DDM1 (Vongs et al., 1993), maintains CG methylation in a replication dependent manner. CMT3 is a plant specific methyl transferase whose mechanism of action is tightly linked to H3K9 methylation (Lindroth et al., 2001) (Jackson et al., 2002). Finally, DRM2 (Johnson et al., 2008) establishes methylation *de novo* during each cell division and is dependent on an intricate pathway called RdDM (RNA directed DNA Methylation) further components of which are being uncovered through genetic screens (Law and Jacobsen, 2010).

Maintenance Methylation

CG methylation is maintained in a replication dependent manner, by recruiting DNA methyltransferases directly to foci of replication (Iida et al., 2002). This replication dependent mechanism means that during DNA replication the parental strands remain methylated whereas the newly synthesized daugher strands are devoid of methyl marks. Both DNMT1 and MET1 have been shown to be involved in the targeting of hemi methylated DNA, (Law and Jacobsen, 2010) through the interaction of proteins which possess an SRA domain (Woo et al., 2007) and hence directly recognize methylated cytosines. Though CG methylation sites are predominantly located in transposable elements and surrounding regions (Slotkin and Martienssen, 2007), CG methylation can also be found in gene bodies; however, this observation is poorly understood and to date no clear link has been found with either repression or activation (Vaughn et al., 2007, Zilberman et al., 2007).

CHG methylation is maintained through a mechanism involving histone tail methylation at lysine 9 (Jackson et al., 2002) as shown by profiling studies comparing both H3K9 methylation and CHG methylation (Zhang et al., 2009) (Bernatavichute et al., 2008) in wild type and K9 methyltransferases mutants, with the latter showing a significant decrease in CHG methylation. This interdependence follows, as KYP, the main K9 methyltransferase involved in this process possesses an SRA domain, while CMT3 possesses a chromodomain, which binds methylated K9. This was confirmed in recent work by showing that CMT3 associates with H3K9 methylated nucleosomes, through both a BAH domain and a chromodomain, and by showing that CHG methylation is downstream of H3K9 methylation (Du et al., 2012).

de novo DNA methylation

Asymmetric CHH methylation is not retained through the recognition of a hemi methylated parental strand; instead, it must be re-established *de novo* with each cell cycle. Similar to CMT3, DMR2 also requires, to a certain extent, proteins with SRA binding domains, specifically SUVH9 and SUVH2 (Johnson et al., 2008). First hypothesized in the early nineties, RNA targeted *de novo* methylation was first shown to act through transgenic viroid cDNA (Wassenegger et al., 1994). This was shown experimentally by infecting tobacco plants with oligomeric copies of cDNA from Potato Spindle Tuber Viroid (PSTVd) whose viroid RNAs accumulate in the nucleus (Wassenegger et al., 1994). Much work has been done since the initial observation, mostly influenced by the discovery of RNA interference (RNAi) leading to the discovery of new components important to this process. These include Argonaute genes, Dicer genes, RNA Dependent RNA polymerase (RDRP) genes and sRNA specific RNA polymerase subunits. The further use of screens has served to uncover more components, including plant specific RNA polymerase subunits and a host of other co-factors (Law and Jacobsen, 2010).

Mechanistically, RdDM is dependent on a class of 24nt siRNAs whose transcription relies on RNA Polymerase IV (PolIV) (Herr AJ Baulcombe 2005) (Onodera Pikaard 2005) which is thought to produce a single stranded RNA transcript. Its processing is further dependent on RDR2, which generates double stranded RNA that is cleaved by DCL3, loading the processed 24nt RNAs in a specific Argonaute protein such as AGO4 (Law and Jacobsen, 2010). Once loaded with the siRNA, AGO4 is recruited to the site of a nascent PolV transcription, either through a direct interaction with PolV or potentially through complementary binding of the siRNA loaded into the argonaute. Finally, AGO4 associates with a complex containing DRM2 (He et al., 2009) (Wierzbicki et al., 2009) To this end, *de novo* methylation requires a siRNA signal to direct it. An interesting study extended this to newly inserted or unmethylated genomic sequences that produce siRNAs (Teixeira et al., 2009): the authors were able to show that loci

producing siRNAs can be remethylated in all sequence contexts over multiple generations, and that this methylation is stably maintained once re-established. Figure 5 provides a general overview of the different types of DNA methylation and how they are interrelated.

DNA methylation through development and mechanisms of DNA demethylation

Regardless of mechanism or context, DNA methylation is maintained through most normal somatic cell divisions. DNA methylation levels are modulated in very specific cell types, underscoring the importance in obtaining tissue and even cell type specific data, especially when addressing mechanistic questions. This is especially illustrated in germ cells, where it has been established in mammalian systems that DNA undergoes rapid and extensive DNA demethylation through paternal germ cell development and subsequently during early embryo development (Popp et al., 2010).

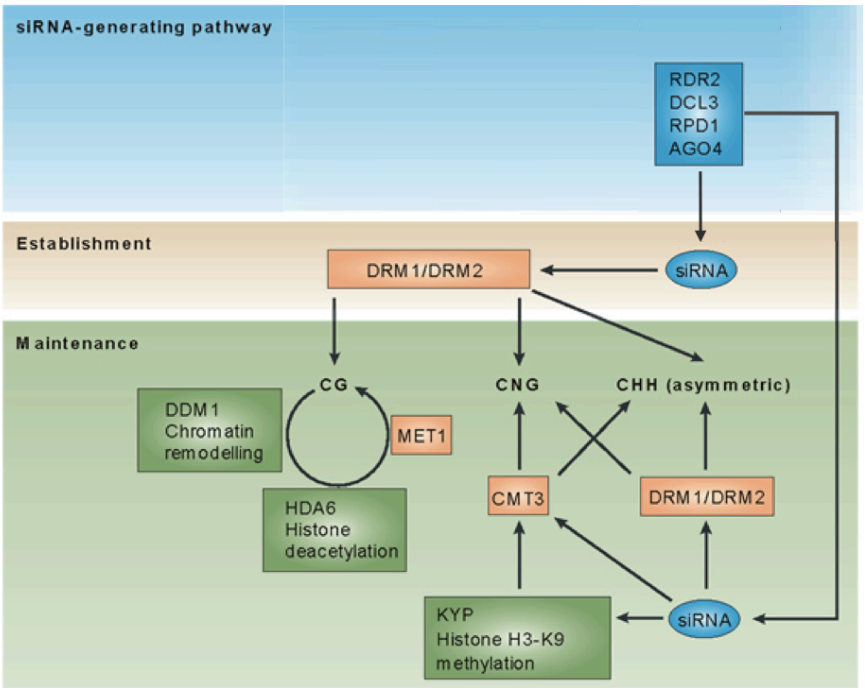


Figure 5 - Different pathways control the maintenance and establishment of cytosine methylation in different sequence contexts. The maintenance of CG methylation requires DNA methyltransferase MET1, DDM1 and HDA6 histone-deacetylase activity. Maintenance of CHG methylation can be directed by histone H3K9 methylation catalyzed by KRYPTONITE (KYP). DRM2 can be targeted to establish *de novo* methylation from siRNA generated from the RDR2/DCL3/RPD1/AGO4 pathway.

**Figure modified from Chan, SW, Nature Reviews Genetics, 2005

With regards to male germline development it was known that DNA methylation levels are affected at a number of transposable element loci in the vegetative cell, and through the use of bisulfite sequencing in purified sperm cells and total pollen (which contains both the sperm cell and the vegetative cell) it was shown that methylation was lost exclusively in the VC (Schoft et al., 2009) (Slotkin et al., 2009). Importantly, chromosomes from the vegetative cell are not passed on to the next generation, consequently these active elements are not transmitted to the next generation where they could have deleterious effects. It has been proposed that the reactivated transposons in the vegetative cell possess heritable genetic information in the form of small RNAs which are able to move between the surrounding vegetative cell into the sperm (Slotkin et al., 2009). This observation regarding DNA hypomethylation at a few loci led to extrapolation to the entire VN, or at least to TE sequences. This mass hypomethylation is understandable from a cytological standpoint, as the vegetative nucleus has a massively decondensed structure compared to the sperm cells (figure 2), which correlates with a lack of repressive chromatin marks.

What mediates this transposon de methylation in the vegetative cell of pollen? Because of the limited number of cell divisions during male germline development, passive loss of methylation cannot account for the observed hypomethlyation in the vegetative cell. Furthermore, at the time this study was initiated none of the known active DNA demethylases were detected via RT-PCR in mature pollen or purified SN or VN. It was not until 2011 when DEMETER (DME), a DNA glycosolase was shown to be expressed in the VN (Schoft et al, 2011). Nonetheless, DME expression in the VN is unlikely to be sufficient for the observed loss of methylation in the VN. Alternatively, a novel group of proteins could mediate this process in the vegetative cell. Two papers published in 2009 highlighted the existence of a modified methylcytosine with a hydroxyl group added to the methyl (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). These studies also indicate that the protein that mediates this process, called TET1, is part of a broad family of oxidoreducatase proteins. This family of enzymes catalyzes the transfer of electrons from one molecule to another through a REDOX reaction. The

oxidoreducatese protein uncovered is part of a specific family which uses α-Ketoglutarate as a source of OH and iron (Fe^{+2}) to perform redox chemistry. TET1 then is involved in putting a hydroxy group on the methyl on carbon 5 of cytosine in tissues where it is present.

There has been much speculation as to the purpose of the hydroxymethylcytosine (hmC) species and the pervading ideas in the field suggest that this species exists either as a targeting mechanism for base excision pathways or as an intermediate in direct cytosine demethylation. This idea is founded on work done in bacteria where homologous proteins to TET1 –called AlkB- have been shown to mediate demethylation of nitrogen 3 of cytosine (Trewick et al., 2002). In bacteria these methyl groups are attached parasitically via alkylating agents in the environment and the AlkB protein is involved in removing the methyl group actively. It has been shown that this protein can hydroxylate the methyl group on the nitrogen and that this is sufficient to lower the activation energy of the new hydroxymethyl moiety so much so that it is a sufficiently active leaving group. This would result in the spontaneous breaking of a carbon nitrogen bond, releasing the hydroxymethyl species and leaving unmodified cytosine. It is proposed that a similar mechanism can occur in eukaryotes with 5-methyl Cytosine. Though the hydroxylation of methyl cytosine by a TET1 style oxidoreductase is a necessary step, it will not sufficiently lower the activation energy to spontaneously break a carbon carbon bond in the same manner as bacterial cytosine demethylation –which is a weaker carbon nitrogen bond. Consequently, other proteins are likely to be involved in a concerted demethylation pathway, by chemically modifying the hydroxymethyl group to make it more reactive.

Since the initial observation of hmC in mammals, it has indeed been shown to be an intermediate in active demethylation (Seisenberger et al., 2013). Currently, the mechanism is thought to proceed as follows; firstly, mC is oxidized to hmC via a TET enzyme. hmC may be deaminated to hydroxymethyluracil, or further oxidized by a TET family protein to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5caC can then be excised by a glycosolyase and then replaced by an unmodified cytosine. At the time of writing, hmC has not

been discovered in Arabidopsis, despite the efforts of a number of groups (Martienssen, Jacobsen, Gehring, Zhu, personal communication).

In Arabidopsis, various mechanisms that mediate cytosine demethylation have been characterized. Active demethylation is mediated via a suite of DNA glycosolyases, which likely work in concert with base excision repair proteins (Zhu, et al. 2009). DME (DeMEter) (Choi et al., 2002) and ROS1 (Repressor Of Silencing 1) (Gong et al., 2002) (Morales-Ruiz et al., 2006) were the first examples uncovered, while related genes like DML2 and DML3 (DeMeter Like 2/3) were discovered afterwards (Ortega-Galisteo et al., 2008) (Ortega-Galisteo AP 2003). Mutations in these genes result in very specific and targeted gains in DNA methylation across the genome at stereotypical locations (Gehring et al., 2009) (Calarco et al., 2012) (Hsieh et al., 2009; Ibarra et al., 2012), in accordance with a role in DNA demethlyation. DNA glycoslyases work mechanistically by flipping the base out of the helix and positioning it into the active site of the enzyme, cleaving the N-glycosidic bond between the sugar-phosphate backbone and the deoxyribose, leaving an abasic site and nicking the DNA backbone in the process. The site then undergoes base excision repair and the resulting site is replaced by an unmodified cytosine.

Genome Reprogramming, siRNA Production and Movement in the Germline

Transposable Elements

First discovered by Barbara McClintock in maize (McClintock, 1948), transposable elements (TEs) are usually thought of as 'parasitic' stretches of DNA that can duplicate and move from one location in the genome to another. Due to their rapid amplification they comprise a large percentage of many eukaryotic genomes. McClintock was the first to suggest that TEs have a major role in influencing gene expression, a concept that is garnering a wider acceptance in plants. Though typically heterochromatic and transcriptionally silent, these elements play a crucial role in the structure and evolution of eukaryotic genomes (Biemont, 2009). When activated, transposons can have deleterious effects on the host, inserting into genic and regulatory regions or promoting chromosomal re-arrangements, causing genome instability (Slotkin et al., 2007).

Mechanisms of transposition have been widely studied and important distinctions must be made between different types of TEs. This form of classification is made on transposition mechanism. Both McClintock's Ac Elements and P elements found in Drosophila are examples of so-called 'DNA Transposons' which transpose through a DNA intermediate. These are also known as 'class II' TEs. In order to transpose, these elements utilize an enzyme that excises its DNA sequence from a particular location and allow the excised fragment to insert in a different part of the genome. The excision event produces nucleotide overhangs which are then repaired by DNA polymerases and DNA ligases. DNA transposons can be further separated into 'cut and paste' style DNA transposons (eg. MuDR elements) or 'rolling circle' DNA transposons (Helitrons).

A further class of TEs, called 'class I' or retrotransposons transpose via an RNA intermediate. These elements are first transcribed from DNA to RNA, and then are reverse transcribed from RNA to DNA. This new DNA copy is then inserted into a different location in the genome. Retrotransposons can be further classified as LTR or nonLTR retrotransposons, meaning they do or do not possess Long Terminal Repeats (LTRs). Either class of TE can lose its self-sufficiency in transposition, accumulating mutations in the genes needed to mediate the transposition. These elements may still transpose though, but only by using the protein products of these genes produced by other transposons present in the genome. These elements have the further designation as 'non-autonomous' while TEs that possess functional copies of all of the genes necessary for transposition are called 'autonomous'. In order to suppress their mutagenic potential, many eukaryotic species have evolved surveillance systems to target transposons to ensure they are inactivated.

siRNAs, Transposon defense and the germline

Double stranded RNA is used as a substrate for the production of small interfering RNAs (siRNAs), which mediate the silencing by acting not only at the locus where they were derived but also at other homologous copies of the transposon in the genome (Martienssen et al., 2008). This transcriptional repression either involves post-transcriptional silencing or the deposition of repressive chromatin modifications, resulting in transcriptional inactivation. As introduced above, in plants transcriptional silencing is sometimes coupled with DNA methylation in a process known as RNA-directed DNA methylation and is mediated by a specific class of siRNAs (Kanno et al. 2010) (Daxinger et al., 2009). However, DNA methylation can be maintained through cell divisions, independently of RNAi, by replication-dependent DNA methyl transferases that are also responsible for the trans-generational inheritance of silencing (Mathieu et al., 2007). This heritable repression of gene expression is known as epigenetic regulation:

which can be defined as a form of heritable variation that does not rely on changes in the primary DNA sequence. The advantage of this type of regulation is that such changes are readily reversible, making them amenable to short-term environmental modulations (Rando and Verstrepen, 2007). Mobile siRNA signals can be grouped into different classes stemming from their size or specific biosynthesis pathways, but they share the same general purpose –to regulate gene expression, either post-transcriptionally or by guiding epigenetic modifications. Most important to the studies presented herein are 21nt and 24nt siRNA and their role in transposable element silencing in the germline is reviewed below.

In plants, much of our current understanding of epigenetics comes from the analysis of sporophytic tissues; however, recent findings point to a prominent role for epigenetic regulation through siRNAs during gametogenesis and seed development (Olmedo-Monfil et al., 2010) (Mosher et al., 2009; Slotkin et al., 2009). As described earlier, plant meiocytes do not differentiate during embryogenesis but instead are formed much later in development. Male gametophyte development in Arabidopsis starts with a diploid mother cell in the anther which undergoes meiosis, forming a tetrad of haploid microspores. This is followed by the first round of pollen mitosis resulting in a bicellular grain which consists of a vegetative cell and a generative cell. The latter undergoes a subsequent mitotic division resulting in two sperm cells. The two haploid sperm cells (SCs) and the vegetative cell make up the tricellular mature pollen grain. During fertilization, one sperm cell fuses with the egg cell while the other fuses with the central cell, thus initiating the development of the embryo and endosperm, respectively.

Transposons are de-repressed specifically in the pollen vegetative nucleus (VN), resulting in the reactivation and mobilization of transposable elements (Slotkin et al., 2009). Though the activated transposons themselves are not passed to the sperm cell, they contribute information through pollen specific, epigenetically activated 21nt siRNAs (easiRNAs). It has been shown using reporter genes that these small RNAs—or their precursors—move from the vegetative cell to the sperm cells, where they can influence silencing (Slotkin et al., 2009).Why is epigenetic

reactivation of transposons needed in the VN? One idea is that transient transposon activation occurs in the VN, where they are 'unmasked' and made apparent to the sperm cell via the production of mobile small RNA, allowing for the re-enforcement of their silencing in the sperm cell itself and/or the developing embryo after fertilization. What mediates this de-repression still remains an interesting question, as a combination of DNA demethylation and chromatin remodeling is likely involved. The mechanism by which this occurs is the main focus of my thesis. Examining sporophytic tissue, transposition occurs only in mutants that lose both heterochromatic histone modification and DNA methylation, such as the noted chromatin remodeler DDM1 (Miura et al., 2001) (Tsukahara et al., 2009) (Singer et al., 2001). Importantly, DDM1 is not expressed in the VN so it follows that transposon de-repression occurs here.

MET1 also maintains repressive epigenetic states in the male germ unit as mutants have been shown to affect imprinted loci transmitted from the pollen parent (Jullien et al., 2006) and to be involved in transposon silencing as both transposons and centromeric repeats are unregulated in *met1* plants (Lippman et al., 2003) . MET1 is known to be expressed in sperm cells and is expressed early in pollen development (Honys and Twell, 2004). How DNA methylation is regulated in the VN remains an interesting question and will be discussed in further detail in a following section. A similar reprogramming mechanism can be proposed for germline development in animals, as epigenetic marks (both DNA methylation and histone tail marks) are lost and reset each generation, and transient expression of transposons is induced in the germline (Calvi and Gelbart, 1994; Dupressoir and Heidmann, 1996; Ostertag et al., 2002; Pasyukova et al., 1997). These transcripts are subsequently processed into piRNAs (Brennecke et al., 2007) that direct remethylation in sperm (Carmell et al., 2007), and into endogenous siRNA that may have a similar function in oocytes (Malone et al., 2009). Mammals have orthologs of both DDM1 (Lsh1) and MET1 (Dnmt1), but a link between reprogramming and small RNA biogenesis has yet to be established.

siRNA biogenesis and mobility in the female germline

As described earlier, female gametophyte development begins with the differentiation of a diploid megaspore mother cell (MMC) that undergoes meiosis to form four haploid megaspores. Three of these megaspores die, and the surviving megaspore undergoes several rounds of mitosis without cell division, resulting in a multinucleate syncytium that subsequently cellularizes. The mature female gametophyte is a seven-celled structure consisting of three antipodal cells, one (diploid) central cell, two synergid cells, and one egg cell. As introduced above, Arabidopsis has two unique RNA polymerases, PolIV and PolV, which have evolved an exclusive function in siRNA biogenesis and transcriptional silencing respectively (Haag et al., 2009; Onodera et al., 2005; Pontes et al., 2009). Profiling of siRNA production in seeds from reciprocal crosses of PolIV mutants with wild-type plants showed that the siRNAs are 24nt in length and that they are maternal in origin, as no change in siRNA profiles was observed in libraries made from developing seeds when the PolIV mutation was paternally inherited in inter-ecotype crosses with polymorphic siRNAs (Mosher et al., 2009). This indicates that PolIV-dependent siRNAs originate from female gametophytic cells and are likely derived specifically from maternal chromosomes. The surrounding maternal tissue also produces siRNA, as the developing integuments (seed coat) express the siRNA binding protein ARGONAUTE9 (AGO9) (Olmedo-Monfil et al., 2010).

AGO9 is expressed in the ovule and in the anther, and preferentially interacts with 24-nucleotide siRNAs derived from transposable elements, as the profiling of siRNAs bound to it shows the majority of reads map to transposons. These AGO9 bound 24nt siRNA may then travel into the female gamete as crosses of *ago9* mutants to enhancer trap lines indicated that its activity is necessary to silence TEs in the female gametophyte (Olmedo-Monfil et al., 2010). These

results further demonstrate the importance of siRNA production and transport to the germ cells and underline the role of epigenetic reprogramming in companion cells to produce these mobile siRNA signals.

DNA demethylation and TE derepression in germline companion cells

DNA methylation plays a major role in silencing transposons, as mutations in genes causing transposon derepression tend to be coupled with loss of DNA methylation (Lippman et al., 2003) (Teixeira et al., 2009). The production of transposon-specific siRNAs in both the male and female germline is accompanied by changes in DNA methylation of transposons in companion cells. During female gametogenesis, the development of the maternal endosperm is accompanied by genome-wide DNA demethylation, mediated by a combination of active demethylation by the DNA repair glycosolase DME (Hsieh et al., 2009) and replication dependent passive demethylation mediated by the Retinoblastoma (Rb) pathway (Jullien and Berger, 2010; Jullien et al., 2008). DME, as detailed above, is an active DNA glycosolyase and works through the base excision repair pathway (Choi et al., 2002), while the effects of Rb are replication coupled, with a complex involving RB-RELATED (RBR) that blocks MET1 expression. To this end DNA, demethylation in the endosperm could be accompanied by DNA remethylation during early embryo development (Choi et al., 2002) (Gehring et al., 2006). These observations can be summarized into a model whereby DNA demethylation mediated by DME and RBR in the central cell leads to transposon reactivation. Transposon mRNA is then processed into 24nt siRNAs, which presumably move to the egg cell — either bound to AGO9 or not — and are used in the developing embryo to reinforce and/or re-establish DNA methylation in the next generation. The existence of 21nt siRNAs in the male germline suggests a distinct biogenesis pathway and possibly a separate model of transposon derepression in the pollen companion cell. With regards to DNA demethylation, both DME, ROS1 and its RNA binding partner ROS3

(Zheng et al., 2008) are lowly expressed in the VN after the first mitosis (Schoft et al 2011), though they may not account for all of the loss of methylation thought to be observed. RBR is expressed in the VN, and MET1 and DDM1 proteins are absent, but both genes are expressed at the earlier microspore stage (Slotkin et al., 2009) (Jullien and Berger, 2010). As there is only one cell division between the generative and vegetative lineages, passive demethylation can thus be ruled out as a mechanism for loss of methylation in the VN. Indeed, methylation profiling of specific loci in pollen and purified sperm, as well as in purified VN, has indicated that CG dinucleotides remain methylated, and CHH and CHG trinucleotides lose methylation only in specific contexts (Slotkin et al, 2009). Without DNA glycosolyases, some other active process must exist. In animals, deamination by the activation- induced cytosine deaminase (Bhutani et al., 2010) or the conversion of methylcytosine to hydroxymethylcytosine (Tahiliani et al., 2009) (Kriaucionis and Heintz, 2009) in combination with DNA repair mechanisms has been proposed to contribute to active DNA demethylation. Similar mechanisms could potentially exist in pollen, but have not yet been uncovered.

Parental siRNAs in the developing embryo

Transposon defense is not the only role for germline-specific siRNAs. One hypothesis is that this maternal and paternal cache of 'transposon information' is created to be readily accessible and equipped to encounter the other genome in the fertilized egg. In Arabidopsis, hybrids between species can have problems developing seed, thus giving rise to infertile progeny (Martienssen, 2010). Both classes of siRNA — 21nt and 24nt — silence transposons with complementary sequences (Slotkin and Martienssen, 2007), and it has been suggested that the 21nt post-transcriptional pathway interacts with the 24nt transcriptional silencing pathway (Eamens et al., 2008). This raises the interesting possibility that both classes of parental siRNA are required in the zygote for permanent silencing (Martienssen, 2010). In that case the load of

foreign transposons could act as a hybridization barrier if they were activated in the zygote. Consistent with this idea, perturbations on the female side of the pathway through the mutation of PolIV resulted in a decrease of DNA methylation (Gehring et al., 2009). It would be interesting to see if this demethylation is coupled to the reactivation of transposons in developing embryos of offspring from mutant mothers in crosses between different accessions of *A. thaliana* that have different TE content.

Fortuitously, the influence of transposon derived siRNAs on interspecies crosses can be studied in Arabidopsis, as transposon sequences in *A. arenosa* and *A. lyrata* are notably different from *A. thaliana* (Beaulieu et al., 2009; Chen et al., 2008; Josefsson et al., 2006) and these species are inter-fertile. In interspecific hybrids, the 21nt and 24nt siRNAs derived from the parental transposons would differ in each species germline, and would not match transposon sequences in the other parent's genome. This could result in problems in transposon silencing in the developing embryo, especially if one class is absolutely required. Blocking these siRNA biogenesis pathways selectively in either germline through mutation could indicate whether these maternally or paternally derived siRNAs do indeed exert any influence in hybrid development (Martienssen et al. 2010). Further, these experiments will also indicated if one of these pathways takes precedent over the other, providing some understanding as to how both these maternal and paternal caches of siRNA ensure the integrity of the parental genomes in out crossing.

Genomic imprinting

Genomic imprinting is a deviation from classical biallelic gene expression normally seen in most diploid organisms. When a locus is imprinted, only one of the two copies of a particular locus is expressed, while the other remains transcriptionally silent. The establishment of mono allelic expression is thought to occur in the stages of gametogenesis and embryogenesis, where the imprint is established and then maintained in somatic cells afterwards. Imprinting was experimentally proven by the use of nuclear transfer experiments, where aberrations in development were observed following the transfer of nuclei to enucleated egg cells in mouse from cells of the same sex (Surani et al., 1984, Solter et al., 1984).

The main role for imprinting seems to lie in embryo development, as many of the mutations that effect the proper establishment of monoallelic expression cause developmental defects in early embryos. A further interesting role for imprinting has been uncovered in mammalian systems, where certain tissues in the brain show mono allelic expression at approximately 200 loci. The expression of these loci in a mono allelic manner has been implicated in brain development in mice (Wilkinson et al., 2007).

New imprinted loci are being constantly uncovered and having a complete catalogue of them will help in understanding this phenomenon. This is mostly being done through the analysis of high throughput sequencing data looking at the RNA profiles of F1 hybrids between related species with single nucleotide polymorphisms (SNPs) between them, allowing for the mapping of transcripts back to one parent or the other (Gehring et al., 2011; McKeown et al., 2011; Wolff et al., 2011). This analysis is far from simple as seen through the very small overlap of imprinted regions uncovered in the *Arabidopsis* studies cited above, as well as the controversy in the uncovering neural imprinted loci in mouse (Gregg et al., 2010a; Gregg et al., 2010b) (DeVeale et al., 2012).

In plants, the most important site for imprinting is in the endosperm, the tissue that is derived from the fusion of the female central cell and one of the male sperm cells. Work trying to understand the mechanisms controlling the establishment and maintenance of mono-allelic expression has implicated the involvement of small RNAs, DNA methylation and histone tail modifications (Gehring and Henikoff, 2007). While, as stated above, mammalian imprinting can occur in both embryonic and extra-embryonic cells, plant imprinting tends to occur only in endosperm (Kohler and Weinhofer-Molisch, 2010). Single parent of origin expression was also thought to occur in the developing embryo in Arabidopsis, which uncovered a rather striking observation, that expression of the paternal genome is delayed in early embryogenesis via a mechanism hypothesized to be different from imprinting observed in the endosperm (Autran et al., 2011). This work has also been recently challenged in the field as Nodine and Bartel (Nodine and Bartel, 2012) have shown that in both early and late embryo stages the contribution of both the maternal and paternal genomes are essentially equal.

Mechanisms of imprinting

In plants one of the main ways to guide imprinting is through DNA methylation. Through work in *met1* plants, the importance of DNA methylation to the maintenance and inheritance of imprints during gametogenesis and inheritance has been established (Mathieu et al., 2007; Saze et al., 2003). Further studies have provided some insight into the mechanism as imprinted regions are methylated by default and selectively demethylated in the endosperm (Choi et al., 2002). This work has been done on a number of imprinted regions on a locus by locus manner with a few notable examples: MEDEA (MEA), FLOWERING WAGENINGEN (FWA) and FERTILIZATION INDEPENDENT SEED 2 (FIS2) are expressed from the maternal allele (Kinoshita et al., 2004) (Jullien et al., 2006), and PHERES1 (PHE1) is preferentially expressed from the paternal allele (Makarevich et al., 2008). This work has since been extended to a number

of other loci in genome wide studies. Further, DEMETER is expressed in the central cell and endosperm and is involved in large scale DNA demethylation during endosperm development (Hsieh et al., 2009) (Gehring et al., 2009).

The mechanisms by which FIS2 and FWA are imprinted are rather well studied and involve DNA methylation (Jullien et al., 2006; Kinoshita et al., 2004). MEA however seems to be imprinted independently of DNA methylation and instead is reliant on the methylation of a residue on the histone H3 tail at lysine 27 (H3K27me) which is mediated by the Polycomb Complex (PcG). Expression of MEA is controlled by a few regions that surround the gene itself and a recent study has shown that one of the main regions that controls MEA's expression has no signs of DNA methylation, nor is the expression of MEA effected in mutations of DME and MET1 (Wohrmann et al., 2012)

Examining both the classically studied imprinted genes and the new list of imprinted loci uncovered via crosses between related species it should be noted that a substantial number of them are surrounded by transposable elements (Gehring et al., 2009), a curious fact that will become important when looking at epialleles and DNA methylation.

Epialleles

As outlined above, the establishment and maintenance of DNA methylation is important to gene expression. If this chromatin modification is not faithfully transmitted, loci that are normally repressed and silent become activated. Consequently, the modulation of methylation status from one cell to the next, or from one generation to another represents an efficient and reversible way to control the expression of a locus. As such, variation among individuals within a population with respect to the degree of DNA methylation can produce novel phenotypes that are potentially heritable across generations. These loci that show occasional differential methylation are termed epialleles. A number of such loci have been implicated in floral symmetry and disease resistance, amongst other phenotypes (Kakutani, 2002). Two classic examples of epialleles in *Arabidopsis* include SUPERMAN (SUP) and FWA. SUP encodes a gene important in flower development. A screen done in the Meyerowitz lab uncovered a number of independently isolated alleles with various levels of phenotypic intensity (Jacobsen and Meyerowitz, 1997). Profiling of the nucleotide sequence of each allele surprisingly revealed no mutations. Instead, these alleles were shown to be differentially methylated and thereby affecting the expression of SUP, each of these alleles were therefore designated epialleles. A similar situation is true of the flowering time gene FWA, where stable demethylation of the promoter region of FWA was stably demethylated. This hypomethylated promoter region arose in both DDM1 and MET1 backgrounds, underscoring the importance of DNA methylation in this process (Lippman et al., 2004). The promoter in the case of FWA is provided by a transposable element inserted 5' to the FWA locus, which illustrates an interesting observation, that transposable elements can directly affect the expression of nearby genes. This observation is important to keep in mind in light of the observation that many imprinted loci (like FWA) are surrounded by transposable elements. Both SUP and FWA are examples of stable DNA methylation-associated epiallelic inheritance.

Tracking the accumulation of Epialleles

Though these and a number of other epialleles have been identified in plants, very little is known about how they arise in the population. Two recent studies report the accumulation of DNA methylation dependent epialleles over multiple generations of self crosses (Schmitz et al., 2011) (Becker et al., 2011). By generating genome-wide methylation profiles of both ancestral and descendant lines the authors were able to uncover a number of sites that showed variable methylation patterns. Through this study more than 100 loci were found to gain/lose DNA methylation sporadically in young leaf tissue after 30 generations of inbreeding by single seed descent (Schmitz et al., 2011) (Becker et al., 2011). Methylation gains were recurrent, occurring at the same loci in multiple independent lines, leading to the proposal that methylation gains and losses might be pre-programmed in the germline (Schmitz et al., 2011). By comparing the methylation data of a few of these epiallelic loci to gene expression data, the authors were able to show a correlation with the gain/loss of methylation and the expression of transcripts from these loci. The existence of epialleles could be in part responsible for some of the phenotypic diversity observed in natural populations. In order to address how these DNA methylation profiles are transmitted from one generation to the next the analysis of the cells which constitute the germline is crucial.

Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R.*, et al.* (2007). New nomenclature for chromatin-modifying enzymes. Cell *131*, 633-636.

Autran, D., Baroux, C., Raissig, M. T., Lenormand, T., Wittig, M., Grob, S., Steimer, A., Barann, M., Klostermeier, U. C., Leblanc, O.*, et al.* (2011). Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. Cell *145*, 707-719.

Beaulieu, J., Jean, M., and Belzile, F. (2009). The allotetraploid Arabidopsis thaliana-Arabidopsis lyrata subsp. petraea as an alternative model system for the study of polyploidy in plants. Mol Genet Genomics *281*, 421-435.

Becker, C., Hagmann, J., Muller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245-249.

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

Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., and Blau, H. M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Biemont, C. (2009). Are transposable elements simply silenced or are they under house arrest? Trends Genet *25*, 333-334.

Boavida, L. C., Becker, J. D., and Feijo, J. A. (2005). The making of gametes in higher plants. Int J Dev Biol *49*, 595-614.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J. A., and Becker, J. D. (2008). Comparative transcriptomics of Arabidopsis sperm cells. Plant Physiol *148*, 1168-1181.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J Biol Chem *278*, 823-828.

Calarco, J. P., Borges, F., Donoghue, M. T., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijo, J. A., Becker, J. D., and Martienssen, R. A. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194-205.

Calarco, J. P., and Martienssen, R. A. (2011). Genome reprogramming and small interfering RNA in the Arabidopsis germline. Curr Opin Genet Dev *21*, 134-139.

Calvi, B. R., and Gelbart, W. M. (1994). The basis for germline specificity of the hobo transposable element in Drosophila melanogaster. Embo J *13*, 1636-1644.

Carmell, M. A., Girard, A., van de Kant, H. J., Bourc'his, D., Bestor, T. H., de Rooij, D. G., and Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Chen, M., Ha, M., Lackey, E., Wang, J., and Chen, Z. J. (2008). RNAi of met1 reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in Arabidopsis allopolyploids. Genetics *178*, 1845-1858.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E., and Fischer, R. L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell *110*, 33-42.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A. J., and Matzke, M. (2009). A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. Embo J *28*, 48-57.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet *8*, e1002600.

Dinant, C., Houtsmuller, A. B., and Vermeulen, W. (2008). Chromatin structure and DNA damage repair. Epigenetics Chromatin *1*, 9.

Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, A.*, et al.* (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167-180.

Dupressoir, A., and Heidmann, T. (1996). Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice. Mol Cell Biol *16*, 4495-4503.

Eamens, A., Vaistij, F. E., and Jones, L. (2008). NRPD1a and NRPD1b are required to maintain post-transcriptional RNA silencing and RNA-directed DNA methylation in Arabidopsis. Plant J *55*, 596-606.

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

Gehring, M., and Henikoff, S. (2007). DNA methylation dynamics in plant genomes. Biochim Biophys Acta *1769*, 276-286.

Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PLoS One *6*, e23687.

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T., David, L., and Zhu, J. K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

Gregg, C., Zhang, J., Butler, J. E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. Science *329*, 682-685.

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

Haag, J. R., Pontes, O., and Pikaard, C. S. (2009). Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. PLoS One *4*, e4110.

Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M. M., Sasaki, N., Nakano, A., Berger, F., and Higashiyama, T. (2011). Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in Arabidopsis thaliana. Curr Biol *21*, 497-502.

He, X. J., Hsu, Y. F., Zhu, S., Wierzbicki, A. T., Pontes, O., Pikaard, C. S., Liu, H. L., Wang, C. S., Jin, H., and Zhu, J. K. (2009). An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell *137*, 498-508.

Henderson, I. R., and Jacobsen, S. E. (2007). Epigenetic inheritance in plants. Nature *447*, 418-424.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol *5*, R85.

Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science *324*, 1451-1454.

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T. F., Uzawa, R., Rodrigues, J. A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T.*, et al.* (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science *337*, 1360-1364.

Iida, T., Suetake, I., Tajima, S., Morioka, H., Ohta, S., Obuse, C., and Tsurimoto, T. (2002). PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. Genes Cells *7*, 997-1007.
Ingouff, M., and Berger, F. (2010). Histone3 variants in plants. Chromosoma *119*, 27-33.

Jackson, J. P., Lindroth, A. M., Cao, X., and Jacobsen, S. E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jacobsen, S. E., and Meyerowitz, E. M. (1997). Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science *277*, 1100-1103.

Jeddeloh, J. A., Stokes, T. L., and Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet *22*, 94-97.

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. Science *293*, 1074-1080.

Johnson, L. M., Law, J. A., Khattar, A., Henderson, I. R., and Jacobsen, S. E. (2008). SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet *4*, e1000280.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol *16*, 1322-1328.

Jullien, P. E., and Berger, F. (2010). DNA methylation reprogramming during plant sexual reproduction? Trends Genet *26*, 394-399.

Jullien, P. E., Kinoshita, T., Ohad, N., and Berger, F. (2006). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting.
Plant Cell *18*, 1360-1372.

Jullien, P. E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., and Berger, F. (2008). Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol *6*, e194.

Kakutani, T. (2002). Epi-alleles in plants: inheritance of epigenetic information over generations. Plant Cell Physiol *43*, 1106-1111.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Kreil, D. P., Breinig, F., Lind, M., Schmitt, M. J., Simon, S. A., Gurazada, S. G.*, et al.* RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. EMBO Rep *11*, 65-71.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S. E., Fischer, R. L., and Kakutani, T. (2004). One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science *303*, 521-523.

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

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S., and Jacobsen, S. E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science *292*, 2077-2080.

Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D.*, et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471-476.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol *1*, E67.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., and Richmond, T. J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol *272*, 301-311.

Makarevich, G., Villar, C. B., Erilova, A., and Kohler, C. (2008). Mechanism of PHERES1 imprinting in Arabidopsis. J Cell Sci *121*, 906-912.

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., and Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Martienssen, R. A. (2010). Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis. New Phytol *186*, 46-53.

Martienssen, R. A., Kloc, A., Slotkin, R. K., and Tanurdzic, M. (2008). Epigenetic inheritance and reprogramming in plants and fission yeast. Cold Spring Harb Symp Quant Biol *73*, 265-271.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell *130*, 851-862.

McClintock, B. (1956). Controlling elements and the gene. Cold Spring Harb Symp Quant Biol *21*, 197-216.

McCue, A. D., Cresti, M., Feijo, J. A., and Slotkin, R. K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. J Exp Bot *62*, 1621-1631.

McKeown, P. C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M. W., Donoghue, M. T., Fort, A., Duszynska, D., Comte, A., Lao, N. T.*, et al.* (2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. BMC Plant Biol *11*, 113.

Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H., and Kakutani, T. (2001). Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. Nature *411*, 212-214.

Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martinez-Macias, M. I., Ariza, R. R., and Roldan-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A *103*, 6853-6858.

Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J., and Baulcombe, D. C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature *460*, 283-286.

Nodine, M. D., and Bartel, D. P. (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. Nature *482*, 94-97.

Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., Slotkin, R. K., Martienssen, R. A., and Vielle-Calzada, J. P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628-632.

Onodera, Y., Haag, J. R., Ream, T., Nunes, P. C., Pontes, O., and Pikaard, C. S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613-622.

Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R., and Roldan-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol *67*, 671-681.

Ostertag, E. M., DeBerardinis, R. J., Goodier, J. L., Zhang, Y., Yang, N., Gerton, G. L., and Kazazian, H. H., Jr. (2002). A mouse model of human L1 retrotransposition. Nat Genet *32*, 655-660.

Pasyukova, E., Nuzhdin, S., Li, W., and Flavell, A. J. (1997). Germ line transposition of the copia retrotransposon in Drosophila melanogaster is restricted to males by tissue-specific control of copia RNA levels. Mol Gen Genet *255*, 115-124.

Pontes, O., Costa-Nunes, P., Vithayathil, P., and Pikaard, C. S. (2009). RNA polymerase V functions in Arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. Mol Plant *2*, 700-710.

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Rando, O. J., and Verstrepen, K. J. (2007). Timescales of genetic and epigenetic inheritance. Cell *128*, 655-668.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593-599.

Ryan, D. P., and Owen-Hughes, T. (2011). Snf2-family proteins: chromatin remodellers for any occasion. Curr Opin Chem Biol *15*, 649-656.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nature Genetics *34*, 65-69.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., and Ecker, J. R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science *334*, 369-373.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

Seisenberger, S., Peat, J. R., Hore, T. A., Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos Trans R Soc Lond B Biol Sci *368*, 20110330.

Singer, T., Yordan, C., and Martienssen, R. A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev *15*, 591-602.

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet *8*, 272-285.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., and Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461-472.

Soppe, W. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I., and Fransz, P. F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J *21*, 6549-6559.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Surani, M. A., Barton, S. C., and Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548-550.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N.*, et al.* (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. Cell *146*, 1016-1028.

Teixeira, F. K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M. F.*, et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature *419*, 174-178.

Tsukahara, S., Kobayashi, A., Kawabe, A., Mathieu, O., Miura, A., and Kakutani, T. (2009). Bursts of retrotransposition reproduced in Arabidopsis. Nature *461*, 423-426.

Vongs, A., Kakutani, T., Martienssen, R. A., and Richards, E. J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926-1928.

Wako, T., and Fukui, K. (2010). Higher organization and histone modification of the plant nucleus and chromosome. Cytogenet Genome Res *129*, 55-63.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H. L. (1994). RNA-directed de novo methylation of genomic sequences in plants. Cell *76*, 567-576.

Wierzbicki, A. T., Ream, T. S., Haag, J. R., and Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet *41*, 630-634.

Wilkinson, L. S., Davies, W., and Isles, A. R. (2007). Genomic imprinting effects on brain development and function. Nat Rev Neurosci *8*, 832-843.

Wohrmann, H. J., Gagliardini, V., Raissig, M. T., Wehrle, W., Arand, J., Schmidt, A., Tierling, S., Page, D. R., Schob, H., Walter, J., and Grossniklaus, U. (2012). Identification of a DNA methylation-independent imprinting control region at the Arabidopsis MEDEA locus. Genes Dev *26*, 1837-1850.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Kohler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. PLoS Genet *7*, e1002126.

Woo, H. R., Pontes, O., Pikaard, C. S., and Richards, E. J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev *21*, 267-277.

Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol *10*, R62.

Zheng, X., Pontes, O., Zhu, J., Miki, D., Zhang, F., Li, W. X., Iida, K., Kapoor, A., Pikaard, C. S., and Zhu, J. K. (2008). ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature *455*, 1259-1262.

Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet *39*, 61-69.

Chapter 2

FACS-based Purification of Arabidopsis Microspores, Sperm Cells and Vegetative Nuclei

** Some figures and text from this section are taken from the publication Borges F, Gardner R, Lopes T, **Calarco JP**, Bovida LC, Slotkin RK, Martienssen RA, Becker JD, *BMC Plant Methods, 2012*
Male germ cell development and the need for precise a purification strategy

As explained in the introductory chapter, plant germ cells differentiate late in the organism's development, giving rise to the male and female gametophytes. The male gametophyte, which is housed within a mature structure called a pollen grain, is derived from the post-meiotic haploid microspore, which undergoes two mitotic divisions after it is formed. This produces the male germ unit (MGU), which is composed of a vegetative cell nucleus (VN) and two sperm cells (SC). The sperm cells are embedded within the cytoplasm of the vegetative cell. The vegetative cell is formed after pollen mitosis I (PM I) and arrests here, while the generative cell undergoes a second mitosis called pollen mitosis II (PM II) and forms the two sperm cells (Boavida et al., 2005).

The fact that the sperm cells are immersed in the cytoplasm of the pollen vegetative cell provides a unique and interesting system of communication and dependence on one another. For example, it has been well established that the main function of the vegetative cell is to facilitate the guidance and delivery of the sperm cells to the egg cell. A more recently uncovered role for the vegetative cell is one of siRNA dynamics and transport, suggesting a communication between the VN and SN (McCue et al., 2011). Further support for this idea of communication between these two cells comes by way of cytological studies, where careful examination of the MGU suggests the existence of cytoplamic channels connecting the VN to the SC, providing a conduit for communication (McCue et al., 2011). A number of studies report various techniques to isolate generative cells in a number of plant species, though none of these species possess the wealth of genetic tools or relatively quick generation time of *Arabidopsis thaliana.* Examples include *Lilium longiflorum* (lily) (Okada et al., 2006), *Oryza sativa* (rice) (Gou et al., 2009), *Zea mays* (maize) (Engel et al., 2003) (Russell et al., 2012), and more recently adopted *Nicotiana tabacum* (tobacco) (Xin et al., 2011). In *Arabidopsis thaliana*, major technical difficulties impede the potential in adapting these established protocols, as the relatively small size made them difficult to separate from surrounding tissue.

The importance of developing a purification system is underscored by the fact that pollen development has developed into a useful molecular systems to study not only gametogenesis and inheritance, but as a natural system to study cell cycle control and chromatin dynamics (Berger and Twell, 2011), as considerable changes in chromatin structure occur over very few cell division. With regards to the cell cycle, sperm cells are arrested in mid S phase and progress through the cell cycle only upon pollen tube growth, which can be induced on plates (Friedman, WE. 1999).

Advances in the area of tissue isolation have been made over the past decade and have been essential in studies that highlight the differences in both transcriptional profiles and epigenetic states of both the VN and the SC (Borges et al., 2008) (Schoft et al., 2009; Slotkin et al., 2009). As an example, being able to isolate and analyze sperm cells and compare them to total pollen was essential to uncovering the potential role of the VN and its participation in actively in controlling heritable epigenetic modifications in the germline. The observation that active transposable elements (TEs) were expressed in total pollen but not in purified sperm cells lead to the discovery of a specific class of siRNAs which derive from these expressed elements and accumulate in the gametes (Slotkin et al., 2009). At the time this work was being performed, only sperm cells were able to be isolated through the use of florescent markers (Borges et al., 2008). Analysis in the VN then could only be performed via a so-called subtractive inference as shown above, examining a particular locus in purified sperm cells (RNA transcription, DNA methylation etc) and comparing this to total pollen. If a particular chromatin profile or transcript were present in total pollen, but not in purified sperm cells then one could infer something was specific to the vegetative cell. This was how it was shown that transposable element expression and transposition occurred exclusively in the VN and not the SC. Though this SC vs. total pollen method exists as a powerful means of examining the male germ unit, in order to truly understand the relationship between these different cell types especially with regards to epigenetic reprogramming in pollen, a simple and powerful method to co-purify the two differentiated types of nuclei from the same mature pollen was needed. Additionally, to make pollen development an even more powerful system to study epigenetic reprogramming we sought to introduce a temporal component, through isolation of the post meiotic male germ unit precursor cell, the microspore. This would allow us a system where in we would be able to examine the roles of various genes in the erasure, addition and maintenance of chromatin modifications all through pollen development, with the ultimate goal of understanding how transposable element control effects siRNA production and epigenetic inheritance.

In this section I will describe a fast and reliable method to isolate Arabidopsis SC and VN which was developed by collaborators in the groups of Jorg Becker and Jose Feijo, as well as a method to isolate microspores which was developed personally. The methods are based on further developing previously reported techniques to isolate mature pollen using high-speed cell sorting (Borges et al., 2008). The first study described a fluorescent marker line specifically labeling differentiated SC in mature pollen, which allowed their FACS-purification and genome-wide transcriptional profiling (Borges et al., 2008). This method allowed for the purification of pure and viable sperm cell fractions, though it was laborious and inefficient, requiring a large number

of plants in order to obtain sufficient starting material. Since the publication of these original studies the method has been improved significantly by using stronger fluorescent markers and more efficient methodologies for pollen disruption, resulting in larger amounts of highly pure material at very high rates. Importantly, it also allowed co-purification of the vegetative nucleus from the same genetic background. This method developed by Filipe Borges will be explained in detail in this section. Furthermore, I will explain a less efficient though no less powerful method for isolating SC and VN through the use of dyes, a method adapted from the Tamaru lab (Schoft et al., 2011). Though advantageous in that various mutants can be analyzed without the need for generations of crosses to insert fluorescent marker genes, the need for DNA dyes may become problematic for certain down-stream applications such as ChIPs, as it is known to interfere with chromatin condensation and nucleosome positioning (Wojcik and Dobrucki, 2008) (Mari et al., 2010). Lastly, I will explain the method I developed to purify microspores without the use of either dyes or marker genes, resulting in a powerful system for studying germ cell development.

Initial Purification of Sperm Cells, Borges et al 2008

The original system employed a FACS based protocol to purify sperm cells from plants that contained a transgene expressing eGPF under the control of a sperm specific promoter. In this first manifestation this was the AtGEX2 promoter. Arabidopsis sperm cells expressing eGFP were isolated using a high-speed cell sorter from crude fractions of ruptured pollen grains. This method involved a number of mechanical filtration steps (Figure 1) that resulted in substantial tissue loss along the way but ensured pure product. It also involved a specific sperm extraction (SE) buffer which contained sucrose and was important in ensuring the sperm cells would not rupture once outside the pollen grain. The composition of the sperm extraction (SE) buffer is as follows :1.3 mM H3BO3, 3.6 mM CaCl2, 0.74 mM KH2PO4, 438 mM Suc, 7 mM MOPS, 0.83 mM MgSO4 in dH20.

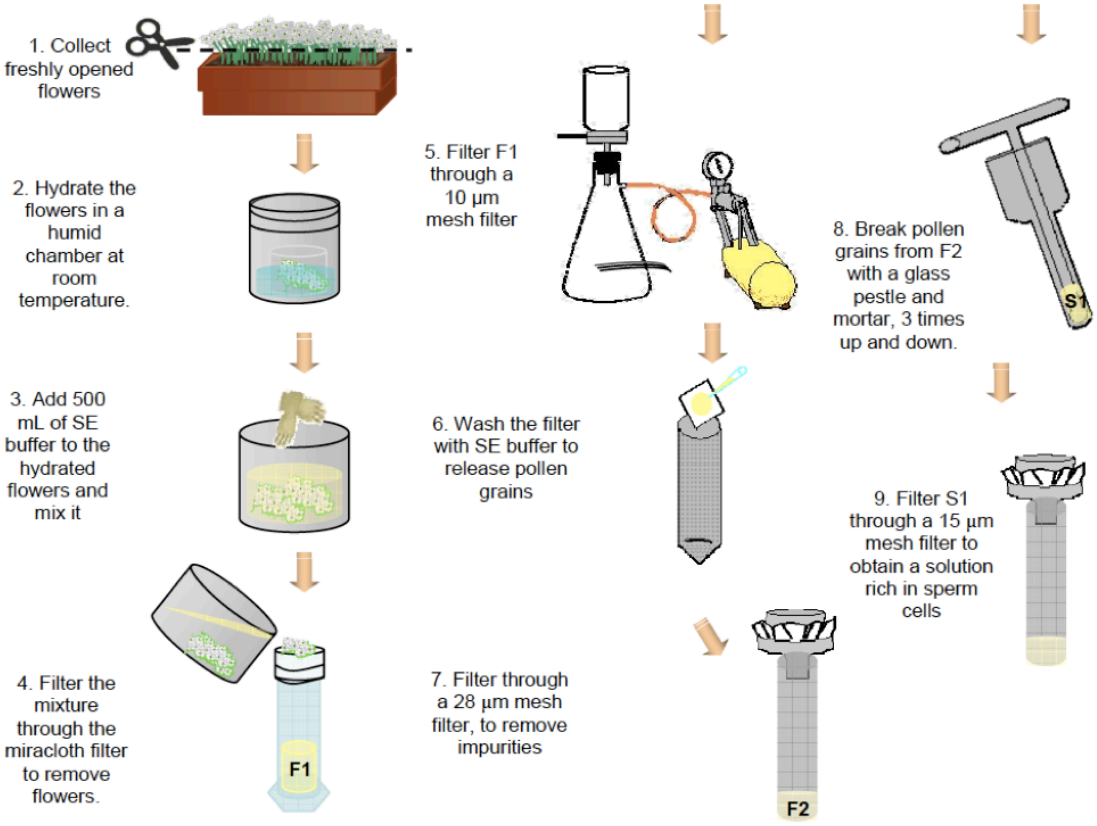


Figure 1 - Schematic representation of initial pollen purification, breakage technique. Open Flowers were cut and hydrated to release pollen grains from flowers. The solution now containing pollen grains is filtered through miracloth to remove flowers then though a 10um filter, keeping the pollen grains above and allowing small debris and liquid though. The filter was then washed to release the pollen grains and then broken with glass mortar and pestle. (Figure taken from Borges *et al.* Plant Physiol. Vol. 148, 2008*)*

Isolated pollen grains were ruptured with glass mortar and pestle and the resulting solution is a liquid mixture containing intact sperm cells, cell debris and some intact pollen grains. This solution was loaded into a flow cytometer and sperm cells were selected based on their size and intracellular complexity (as assessed by the angle of their light scattering properties), their GFP signal, and presence of DNA (stained with DRAQ5). Low 'granulosity' (represented by low side scatter [SSC] signal) and GFP positive signals were used to identify the sperm cells (R1 in Fig. 2) as a unique population from any of the other debris on the plot. A low forward scatter (FSC) signal (small particles) within the GFP positive population further discriminated between sperm cells and other small entities within the sample (R2 in Fig. 2). To assure exclusion of debris within the small sized population containing GFP-positive sperm cells, DRAQ5 which is DNA dye was used (R3 in Fig. 2). Its important to note that each session yielded only around 100,000 cells. Purity of the populations were verified via bright field microscopy (Figure 3), which when counted showed purities of over 98%

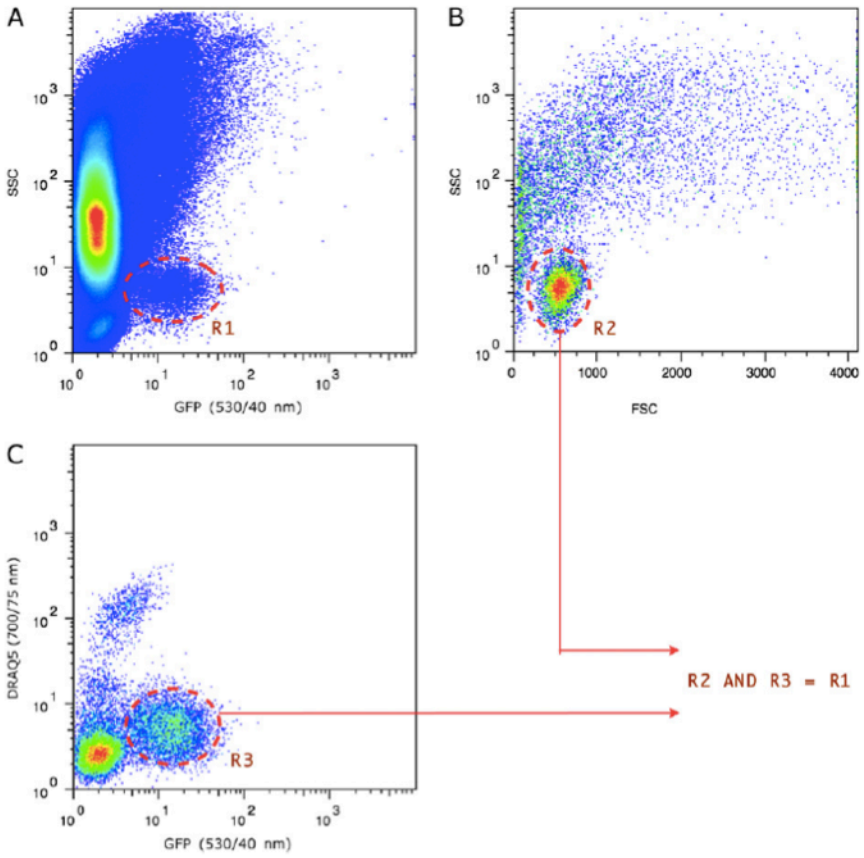


FIGURE 2 – FACS plots illustrating the profiles of the original sperm cell sorting protocol based on cell size (FSC), intracellular complexity (SSC), GFP signal, and presence of intracellular DNA, via DRAQ5 staining. Low granulosity (low SSC) and GFP positive signals were used to identify the sperm cell population (R1) from the total population. To guarantee purity, a low FSC signal (small particles; R2) within the GFP/DRAQ5 double positive population (R3) were used to exclude other small particles. A displays total population, B shows cells within region R2, and C shows cells within region R3. (Figure taken from Borges *et al.* Plant Physiol. Vol. 148, 2008)

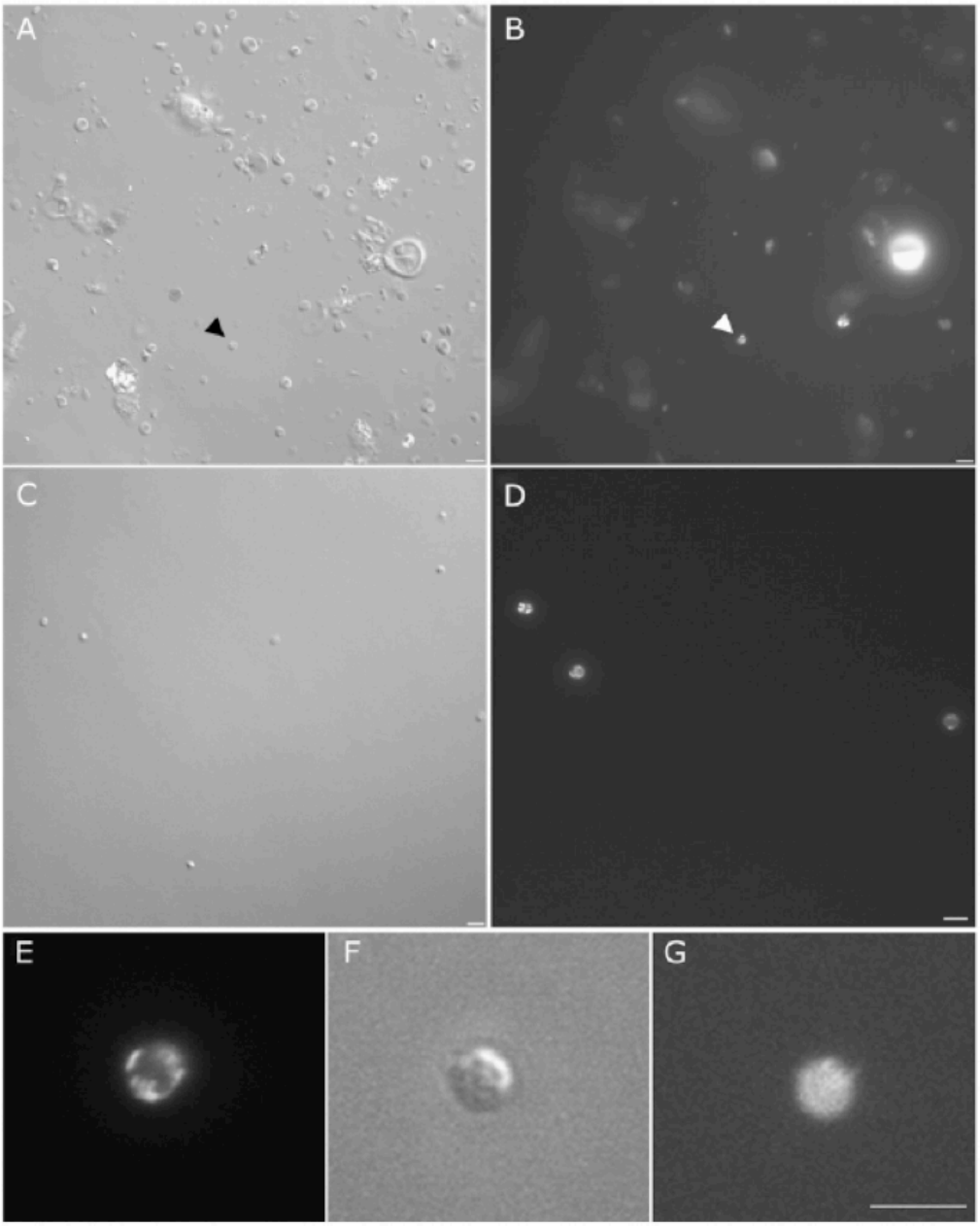


FIGURE 3 – Visualization of FACS-purified sperm cells. Wide-field fluorescence microscopy was used to visualize *Arabidopsis* sperm cells expressing AtGEX2::eGFP before (B) and after FACS purification (D). Differential interference microscopy (DIC) microscopy confirmed that the debris in the filtrate before FACS (A) was removed after sorting (C). For DIC imaging of FACS purified sperm cells (C), we captured and merged several images along the optical axis. A higher magnification of a sorted sperm cell shows GFP fluorescence (E), cell-shape integrity by DIC microscopy (F), and cell viability using fluorescein diacetate staining (G). The bars represent 5 microns and the arrowheads are pointing to sperm cells. (Figure taken from Borges *et al.* Plant Physiol. Vol. 148, 2008*)*

Co-purification of sperm cell and vegetative nuclei

Though the original protocol was functional and provided pure populations of sperm cells, two major limitations still needed to be addressed in order to make the protocol more powerful. Firstly, the procedure was inefficient; a session of sorting (~5-7 hours on a machine) yielded only 100,000 sperm cells. Secondly, we could not obtain any information from the vegetative cell. Naturally it would be impossible to purify the entire vegetative cell without the sperm cell (recall from the introduction, the sperm cell is embedded within the vegetative cell's cytoplasm), but it would be possible to purify the nucleus. In order to improve and simplify the SC-sorting method and to additionally co-purify the VN from the same genetic background, a transgenic line expressing distinct fluorescent proteins in both nuclei was generated. The ACT11 promoter driving histone H2B fused to mRFP was used as a VN marker (Rotman et al., 2005), while the promoter for MGH3, a sperm cell specific histone 3 variant was used to tag sperm cells (Ingouff et al., 2007; Okada et al., 2005). A homozygous plant harboring the ACT11p::H2B::mRFP transgene in Col background was crossed with a stronger sperm-specific marker line containing the MGH3p-MGH3::eGFP construct. MGH3 is expressed as early as bicellular pollen, and is highly abundant in the sperm nuclei of mature pollen (Figure 4). This resulted in pollen grains expressing eGFP or mRFP in either the SN or VN.

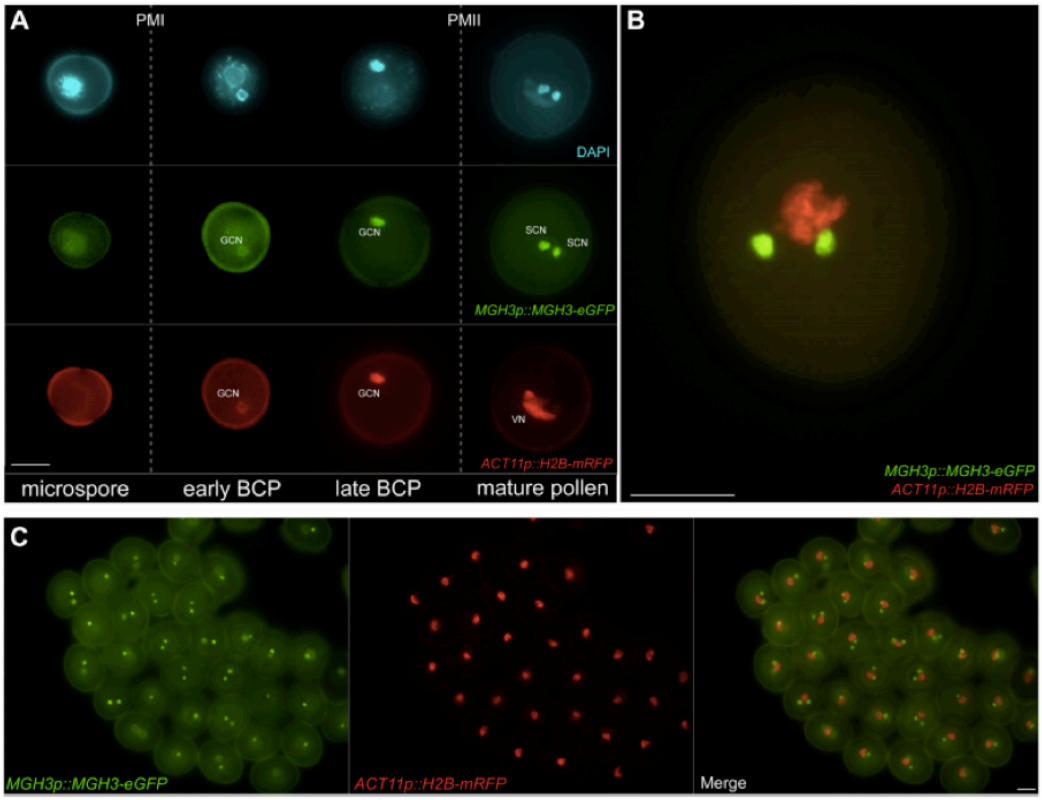


Figure 4 - Expression pattern of GFP- and RFP-fusion proteins during pollen development. (A) MGH3p::MGH3-eGFP localizes in the generative cell nucleus (GCN) after the first pollen mitosis, and is strongly accumulated in the sperm cell nucleus (SCN) of mature pollen. ACT11p::H2B-mRFP is initially expressed in the GCN until late bicellular pollen (BCP). After the second pollen mitosis the expression of this transgene is down-regulated in the germline, and it becomes strongly expressed in the vegetative nucleus (VN). (B) Merged magnification of a mature pollen grain expressing both transgenes. (C) Population of mature pollen grains from double homozygous plants, confirming strong and stable expression of both transgenes. Scale bars: 10 µm. *Figure from Borges F, Gardner R, Lopes T, **Calarco JP**, Boavida LC, Slotkin RK, Martienssen RA, Becker JD, *BMC Plant Methods, 2012*

Since our collaborators created this reporter line, I was able to make a number of advancements to the pollen collection protocol, where now pollen grains were collected from open flowers by vortexing with nuclei extraction buffer (see end of chapter for buffer composition), and released pollen was then disrupted by additional vortexing with glass beads, which proved more efficient than breaking with mortar and pestle.

Enriched filtrates containing SC and VN released from broken pollen were co-purified by FACS based on their distinct fluorescence properties (Figure 5A). Populations were inspected for purity via bright field microscopy as before (Figure 5B) and further via RT-PCR on SC and VC-specific transcripts. MGH3 an SN specific gene is only expressed in the SC and was not detected in the sorted VN fraction, while VEX1, a VC-expressed gene was also not detected in the SC fraction (Figure 5D), indicative of two pure populations. Re-analyzing sorted populations stained with 4',6-diamidino-2-phenylindole (DAPI) confirmed that the SC and VN populations were consistently more than 99% pure (Figure 5C). Furthermore, the ratio SC/VN after sorting was consistently 2:1, demonstrating a good recovery for both types of nuclei after pollen disruption (data not shown). The procedure was optimized to obtain higher yields of VN and SC nuclei/cells, maximizing the number of sorted events per second. Where as before a 5 hour sort would yield 100,000 nuclei using the current line we can purify between 1,000,000/1,500,000 sperm cells, a greater than ten fold increase in collection rate. Pushing the limits of the machine, we can sort at a rate of 900 events per second, compared to ~100 events per second in the previous manifestation of the protocol.

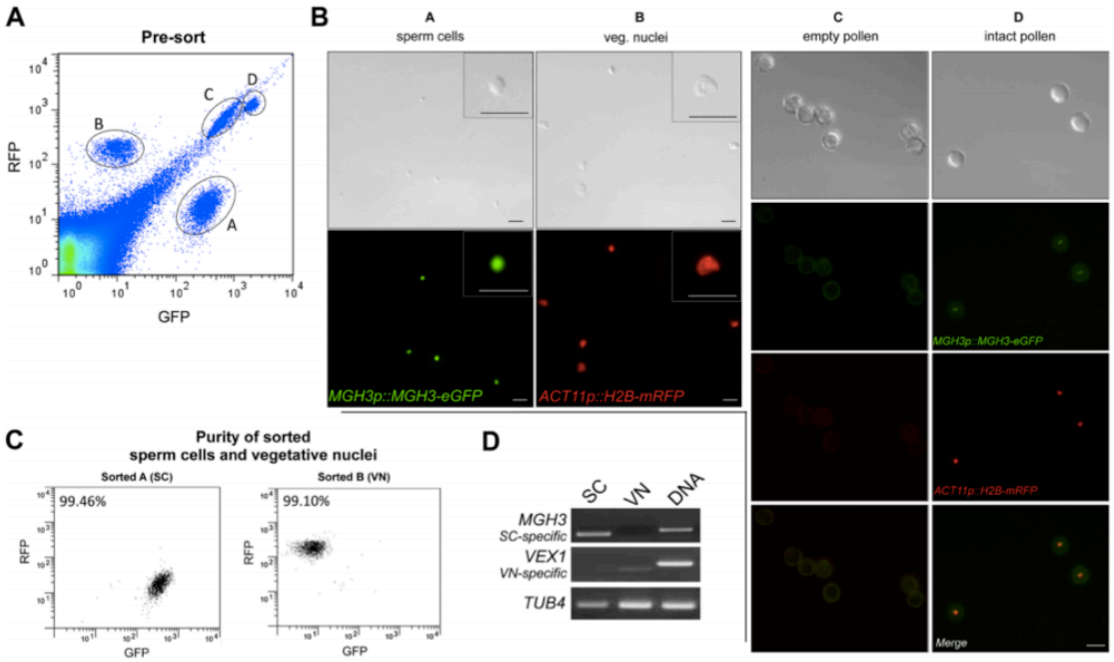


Figure 5 - Purification of sperm and vegetative nuclei by FACS. (A) Four distinct cell populations are highlighted in the filtrate pre-sorting. Sperm cells (SC) nuclei are identified based on their strong GFP signal (Population A), whereas vegetative nuclei (VN) separate towards the opposite axis based on strong RFP signal (Population B). In addition the filtrate contains empty (Population C) and intact pollen grains (Population D). (B) Purity of sorted SC and VN fractions was confirmed by DIC and fluorescence microcopy; scale bar: 10 µm. Populations C and D were confirmed to represent empty and intact pollen, respectively; scale bar: 30 µm (C) Sorted SC and VN samples were stained with DAPI and run through a flow cytometer to check for purity. Purity is determined by measuring the percentage of SC and VN present within the total number of DAPI positive events, corresponding to all DNA-containing particles present in the sorted sample. (D) RT-PCR analyses confirmed that each fraction is enriched for cell-specific transcripts (MGH3 for SC and VEX1 for VN), and devoid of contaminating RNAs. TUB4 was used as control. *Figure from Borges F, Gardner R, Lopes T, **Calarco JP**, Boavida LC, Slotkin RK, Martienssen RA, Becker JD, *BMC Plant Methods, 2012*

In parallel to employment of the marker assisted purification method described above, Schoft et al. 2009 developed a method to sort SC and VN via FACS based on the difference in granularity of the vegetative nucleus and sperm cell combined with DNA staining, in this case with SbyrGreen (Figure 6). We have adapted this sorting protocol with our pollen purification and pollen grain rupturing technique and have been able to obtain highly pure fractions of SC and VN, though not at a rate as quickly as with the reporters. Employing dyes resulted in rates of ~300-400 events/second compared to rates of 700-800 events/second with the reporter line.

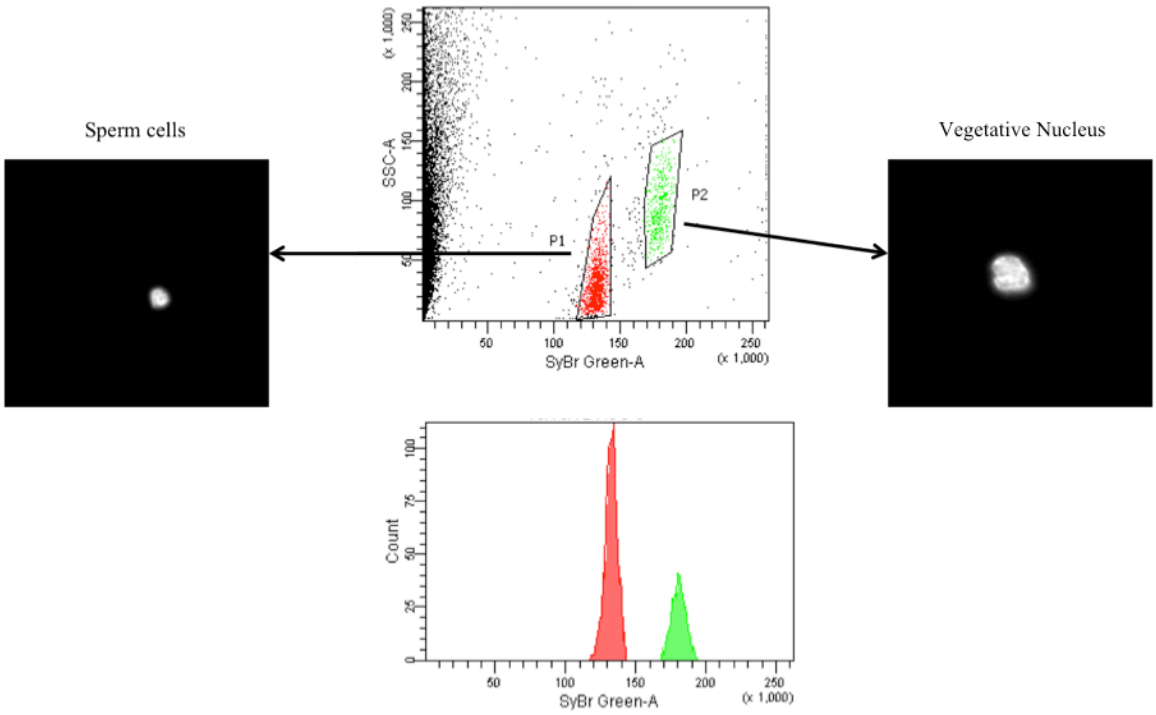


Figure 6 – SybrGreen based sorting of SC and VN without transgenes. Through taking advantage of changes in granularity of the nuclei distinct populations of SC DNA and VN DNA can be separated. The number of events correspond roughly to a 2:1 ration of SC to VN.

Microspore purification by FACS

As mentioned above, to truly understand the developmental implications of the genetic and epigenetic mechanisms that control VN and SC differentiation, analysis of their precursor microspore cell is essential. Available methods to isolate Arabidopsis microspores relied on Percoll density gradients (Honys and Twell, 2004), and would not provide a sufficiently pure fraction suitable for genomic analysis at the DNA level, besides the problem of relatively low yield. Though purities of close to 80% were reported in these studies, in my hands I was unable to obtain populations greater than 70% pure. Additionally, the protocol is inefficient, losing tissue with each centrifugation step. Previous attempts to isolate microspores by flow cytometry were successfully applied in Brassica napus (Pechan et al., 1988), however, to our knowledge these same methods were not tried in *Arabidopsis thaliana*. I collected closed flower buds of wild type plants which contain pollen through all stages of development and by grinding gently in microspore extraction buffer (MEB) (see details at end of the chapter) very gently with ceramic mortar and pestle to dislodge microspores from tetrads, but careful not to rupture the forming pollen grain which is essential to the purification process. The process of microgametogenesis is accompanied by a rapid increase in pollen grain diameter through mitosis 1 and mitosis 2 (Figure 7). Taking advantage of this property we are able to enrich for the population containing only microspores and bicellular pollen via mechanical filtration through a 20um filter before loading the solution on the FACS sorter.

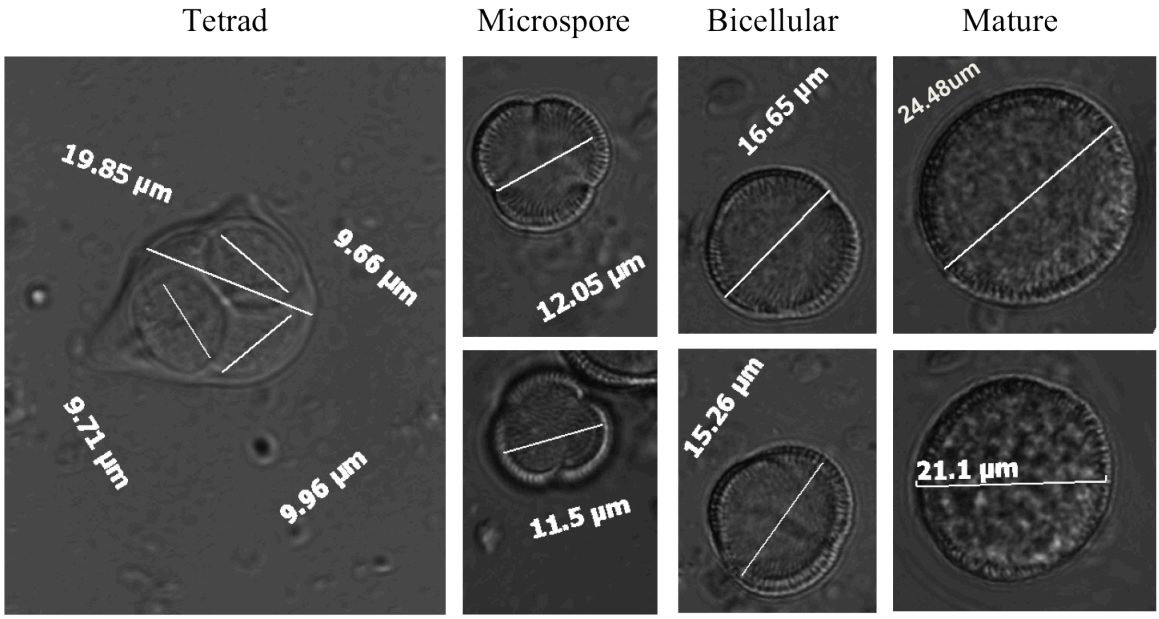


Figure 7 – Pollen increases in diameter throughout pollen development. DIC image of the various stages of male gametogenesis from dissected inflorescence. Microspores still contained in the tetrad are under 10um in diameter and consistently grow throughout pollen development until the ranger between 20-24um when fully mature.

Essential to this system to purify microspores was an observation from the method to purify SC and VN, which revealed that intact pollen grains have a high amount of autofluorescence (population D in Figure 5A). As such, we exploited this property, in addition to the fact that developing pollen at the microspore stage is much smaller in comparison with pollen at later stages of development (Figure 8). The sorted population was analyzed by microscopy and confirmed to contain mostly microspores (97.7%) (Figure 8). A very small fraction (2.3%) of early bicellular pollen was also observed (Figure 8). As these are probably cells that have just gone through the first pollen mitosis, they have identical size and autofluorescent properties as microspores, and are therefore impossible to distinguish with our FACS settings.

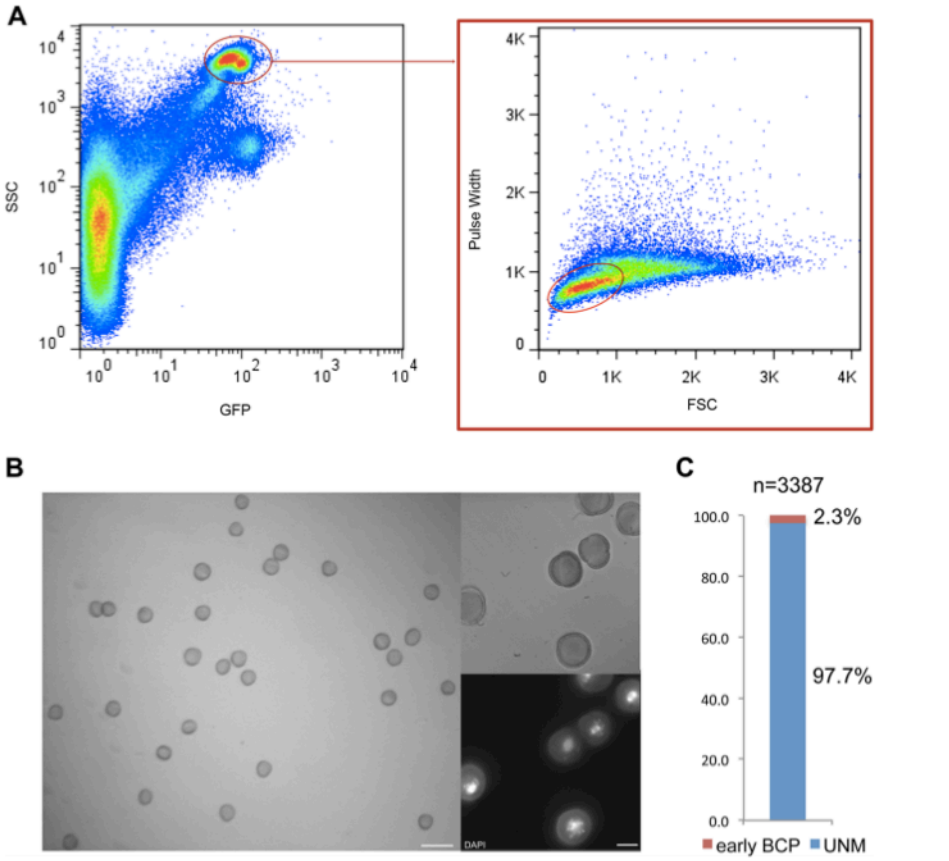


Figure 8 -Microspore sorting. (A) Pollen population is characterized by an elevated high angle scatter (SSC) and autofluorescence (observed in the GFP channel using a 530/40 nm bandpass filter). Within this population, microspores (right panel, circular gate) can be differentiated from bicellular and tricellular pollen by their characteristic smaller size, captured by a diminished low angle scatter (FSC) and time-of-flight (Pulse Width), as compared to other stages of pollen development. (B) Sorted microspores were inspected by microscopy to show purity and integrity, as revealed by DAPI staining. Scale bars: 30 μm (left panel) and 10 μm (right panels). (C) Sorted microspores stained with DAPI were counted on a wide-field fluorescence microscope to confirm that most sorted cells are Uninucleate Microspores (UNM). BCP - Bicellular Pollen. *Figure from Borges F, Gardner R, Lopes T, Calarco JP, Boavida LC, Slotkin RK, Martienssen RA, Becker JD, *BMC Plant Methods, 2012*

It has been well established that gametogenesis and fertilization involve a series of genetic and epigenetic reprogramming events. The organism undergoes an accurate and well-regulated series of events wherein it is decided which chromatin marks are kept and which are reset. This process is given even greater weight in Arabidopsis as the germline is made later in development, meaning any chromatin modifications acquired over the lifetime of the organism must be considered. Because of this, gametogenesis is often characterized by transcriptional activation of a number of specific germline genes and repression of somatic gene expression, as well as dynamic changes of DNA methylation and histone tail modifications on a genome-wide scale. A comprehensive interpretation of such phenomena relies on analyzing isolated cell populations, and as such it demands methods for efficient and robust purification of each cell type involved in the process. We now have such a system to understand these important mechanisms, as FACS coupled to nuclei-specific fluorescent markers provides fast and reliable results to isolate Arabidopsis sperm cell nuclei and vegetative nucleus fractions that are consistently more than 99% pure. Sorted nuclei are suitable for most molecular analysis such as DNA methylation profiling and chromatin immunoprecipitation. Our system using fluorescent proteins under the control of strong cell-specific promoters eliminates the need for DNA dyes, if they prove to be problematic for certain experiments. If they are not, then they too can be used in lieu of florescent reporters and though less efficient than markers, it saves the time needed to cross two markers into a mutant of interest. Besides the isolation of VN and SC, we developed a FACS-based protocol allowing the isolation of microspores from wild-type or mutant plants with unprecedented purity, accomplished again without the use of DNA dyes.

Buffer Compositions

Galbraith buffer

45 mM MgCl2, 30 mM Sodium Citrate, 20 mM MOPS, 1% Triton-100, pH to 7.0

Sperm Extraction Buffer (SEB)

1.3 mM H3BO3, 3.6 mM CaCl2, 0.74 mM KH2PO4, 438 mM sucrose, 5.83 mM MgSO4, 7 mM MOPS at pH 6

Microspore extraction buffer (MEB)

10mM $CaCl_2$, 2mM MES, 1mM KCl, 1% H_3BO_3, 10% Sucrose, pH 7.5

References

Berger, F., and Twell, D. (2011). Germline specification and function in plants. Annu Rev Plant Biol *62*, 461-484.

Boavida, L. C., Becker, J. D., and Feijo, J. A. (2005). The making of gametes in higher plants. Int J Dev Biol *49*, 595-614.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J. A., and Becker, J. D. (2008). Comparative transcriptomics of Arabidopsis sperm cells. Plant physiology *148*, 1168-1181.

Borges, F., Pereira, P. A., Slotkin, R. K., Martienssen, R. A., and Becker, J. D. (2011). MicroRNA activity in the Arabidopsis male germline. J Exp Bot *62*, 1611-1620.

Calarco, J. P., and Martienssen, R. A. (2011). Genome reprogramming and small interfering RNA in the Arabidopsis germline. Curr Opin Genet Dev *21*, 134-139.

Engel, M. L., Chaboud, A., Dumas, C., and McCormick, S. (2003). Sperm cells of Zea mays have a complex complement of mRNAs. Plant J *34*, 697-707.

Gou, X., Yuan, T., Wei, X., and Russell, S. D. (2009). Gene expression in the dimorphic sperm cells of Plumbago zeylanica: transcript profiling, diversity, and relationship to cell type. Plant J *60*, 33-47.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol *5*, R85.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. Curr Biol *17*, 1032-1037.

Mari, P. O., Verbiest, V., Sabbioneda, S., Gourdin, A. M., Wijgers, N., Dinant, C., Lehmann, A. R., Vermeulen, W., and Giglia-Mari, G. (2010). Influence of the live cell DNA marker DRAQ5 on chromatin-associated processes. DNA Repair (Amst) *9*, 848-855.

McCue, A. D., Cresti, M., Feijo, J. A., and Slotkin, R. K. (2011). Cytoplasmic connection of sperm cells to the pollen vegetative cell nucleus: potential roles of the male germ unit revisited. J Exp Bot *62*, 1621-1631.

Okada, T., Bhalla, P. L., and Singh, M. B. (2006). Expressed sequence tag analysis of Lilium longiflorum generative cells. Plant Cell Physiol *47*, 698-705.

Okada, T., Endo, M., Singh, M. B., and Bhalla, P. L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J *44*, 557-568.

Pechan, P. M., Keller, W. A., Mandy, F., and Bergeron, M. (1988). Selection of Brassica napus L. embryogenic microspores by flow sorting. Plant Cell Rep *7*, 396-398.

Rotman, N., Durbarry, A., Wardle, A., Yang, W. C., Chaboud, A., Faure, J. E., Berger, F., and Twell, D. (2005). A novel class of MYB factors controls sperm-cell formation in plants. Curr Biol *15*, 244-248.

Russell, S. D., Gou, X., Wong, C. E., Wang, X., Yuan, T., Wei, X., Bhalla, P. L., and Singh, M. B. (2012). Genomic profiling of rice sperm cell transcripts reveals conserved and distinct elements in the flowering plant male germ lineage. New Phytol *195*, 560-573.

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J. S., Park, G. T.*, et al.* (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A *108*, 8042-8047.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., and Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461-472.

Wojcik, K., and Dobrucki, J. W. (2008). Interaction of a DNA intercalator DRAQ5, and a minor groove binder SYTO17, with chromatin in live cells--influence on chromatin organization and histone-DNA interactions. Cytometry A *73*, 555-562.

Xin, H. P., Peng, X. B., Ning, J., Yan, T. T., Ma, L. G., and Sun, M. X. (2011). Expressed sequence-tag analysis of tobacco sperm cells reveals a unique transcriptional profile and selective persistence of paternal transcripts after fertilization. Sex Plant Reprod *24*, 37-46.

Chapter 3
Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA.

** Text and figures taken from **Calarco, J. P**., Borges, F., Donoghue, M. T., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijo, J. A., Becker, J. D. et al. (2012) 'Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA', *Cell*.

Introduction

Epigenetic reprogramming via DNA methylation Dynamics

Epigenetic inheritance refers to the transmission of modified genetic material from one generation to the next. These "epialleles" are not caused by mutations in the DNA sequence, but instead by covalent modification of chromatin and DNA, guided by developmental and environmental cues. In general, epigenetic modifications that are programmed during development must be reset in the germline, so that the zygote is restored to pluripotency and can once again initiate embryonic development. For example, imprinted genes in the mouse are expressed predominantly from either the paternal allele or from the maternal allele in the diploid embryo, and so must be reprogrammed in the germline depending on its sex (Bartolomei and Ferguson-Smith, 2011). Indeed, the mouse genome undergoes several rounds of DNA methylation, demethylation and repair as germ cells differentiate, as well as in the embryo after fertilization when imprinted genes are largely immune (Bartolomei and Ferguson-Smith, 2011; Feng et al., 2010; Popp et al., 2010). For this reason, epigenetic inheritance is thought to be rare in mammals, and is generally restricted to non-essential genes.

Flowering plants are an important exception to this rule, as epigenetic modification during development can be inherited for hundreds of generations with dramatic developmental consequences (Cubas et al., 1999). The first (and most common) examples of epigenetic inheritance in plants involved transposable elements (TE), which can regulate nearby genes, and undergo epigenetic switches during development, resulting in the inheritance of epialleles (Martienssen et al., 1990; McClintock, 1965). As in mammals, epigenetic inheritance of transposon activity in plants involves DNA methylation (Becker et al., 2011; Cubas et al., 1999; Martienssen and Baron, 1994; Schmitz et al., 2011). Imprinted genes tend to be flanked by transposable elements, whose methylation can influence their expression (Radford et al., 2011). However, imprinting in plants is largely restricted to the extra-embryonic endosperm, a terminally differentiated tissue within the seed, so that imprinted chromatin and DNA modifications need not be removed once they are established (Feng et al., 2010; Jullien and Berger, 2009; Raissig et al., 2011). The extent of reprogramming in the plant germline thus remains an important question.

Unlike mammals, which set aside their germline in early development, flowering plants give rise to germ cells during post-embryonic growth and development, in some cases many years after embryogenesis is complete. The pollen mother cell (PMC) on the paternal side and the

megaspore mother cell (MMC) on the maternal side are specified from somatic cells in developing flowers (Boavida et al., 2005). In the anthers, the PMC undergoes meiosis resulting in four haploid microspores. Each microspore subsequently undergoes an asymmetric division to differentiate a larger vegetative cell and a smaller generative cell, which represents the male germline (Figure 1A). The vegetative cell exits the cell cycle into G_0, while the generative cell undergoes a further symmetric division to produce two identical sperm cells that are surrounded by the vegetative cell (Berger and Twell, 2011).

The most conspicuous evidence of reprogramming in the plant germline is that the vegetative nucleus (VN) of the pollen grain has completely decondensed heterochromatin, in contrast to the tightly condensed chromatin found in sperm cell (SC) nuclei (Figure 1A). Heterochromatin in plants is mostly occupied by TEs and repeats (Lippman et al., 2004). TE repression is important for genome integrity and mutants in *DDM1* (*DECREASE in DNA METHYLATION 1*) and *MET1* (*DNA METHYLTRANSFERASE 1*) have reduced DNA methylation levels resulting in up-regulation of TEs (Lippman et al., 2004). MET1 maintains CG methylation, and its activity in the germline impacts epigenetic inheritance (Jullien et al., 2006; Saze et al., 2003). In plants, CHROMOMETHYLASE3 (CMT3) maintains CHG methylation, guided by histone modification, and cytosines can also be methylated in an asymmetric CHH context guided by RNA interference (RNAi) (Law and Jacobsen, 2010). RNA-directed DNA methylation (RdDM) requires the DNA methyltransferase DOMAINS REARRANGED METHYLASE 2 (DRM2), and the RNA polymerase IV and V subunits NRPD1a, and NRPE1a, which are involved in production and utilization of 24nt siRNA (Haag and Pikaard, 2011). These mechanisms interact, so that RdDM is required to re-methylate TEs in *ddm1* mutants. TEs without matching siRNA cannot be remethylated even when DDM1 function is restored through crosses to wild-type plants (Teixeira et al., 2009).

Loss of heterochromatin in the vegetative nucleus of the pollen grain is accompanied by the loss of DDM1, the activation of TEs, and the production of a novel class of 21nt siRNAs which accumulate in sperm cells (Slotkin et al., 2009). However, while some TEs and repeats were found to be demethylated in the VN, others were hypermethylated so that the role of DNA methylation in pollen reprogramming was unclear (Schoft et al., 2011; Schoft et al., 2009; Slotkin et al., 2009). We set out to determine the dynamics of DNA methylation during pollen development, via bisulfite sequencing of genomic DNA from *Arabidopsis* microspores, and from their derivative sperm and vegetative cells (Figure 1A). We found that symmetric CG and CHG

methylation were largely retained in *Arabidopsis* pollen. However, CHH methylation was lost from at least 1500 TEs, mostly long terminal repeat (LTR) retrotransposons, in microspores and sperm cells. In the VN, more than 100 DNA transposons and non-LTR retrotransposons were targeted for CG demethylation by DNA glycosylases. Many of these transposons, including those that flank imprinted genes, gave rise to 24nt siRNA in sperm cells where DNA glycosylases are not expressed. Recently discovered recurrent epialleles were pre-methylated in sperm cells guided by a similar mechanism. Thus reprogramming of DNA methylation in pollen contributes to transposon silencing, the transgenerational recurrence of epialleles, and imprinting of maternally expressed genes.

Results

Sequencing of the methylome from individual pollen cell types presents a significant challenge, especially in *Arabidopsis* where pollen yields are limiting. Sperm cells and vegetative nuclei were isolated using Fluorescence Activated Cell Sorting (FACS), through the use of cell specific promoters driving the expression of Red and Green Fluorescent Protein (RFP and GFP) (Borges et al., 2012). Microspores were obtained from young flower buds through a combination of mechanical filtration and purification through FACS, taking advantage of their small size and autofluorescent properties (Borges et al., 2012). Genomic DNA was isolated from each nuclear fraction, treated with sodium bisulfite and sequenced at 7-17x coverage (Table S1). To test whether each cytosine was methylated, the proportion of methylated reads to un-methylated reads was compared to the background error rate using a binomial test for each cytosine with sufficient coverage (28)(28). The data was plotted as a heatmap on all five chromosomes, compared with the methylome of somatic cells from leaves (Figure 1B).

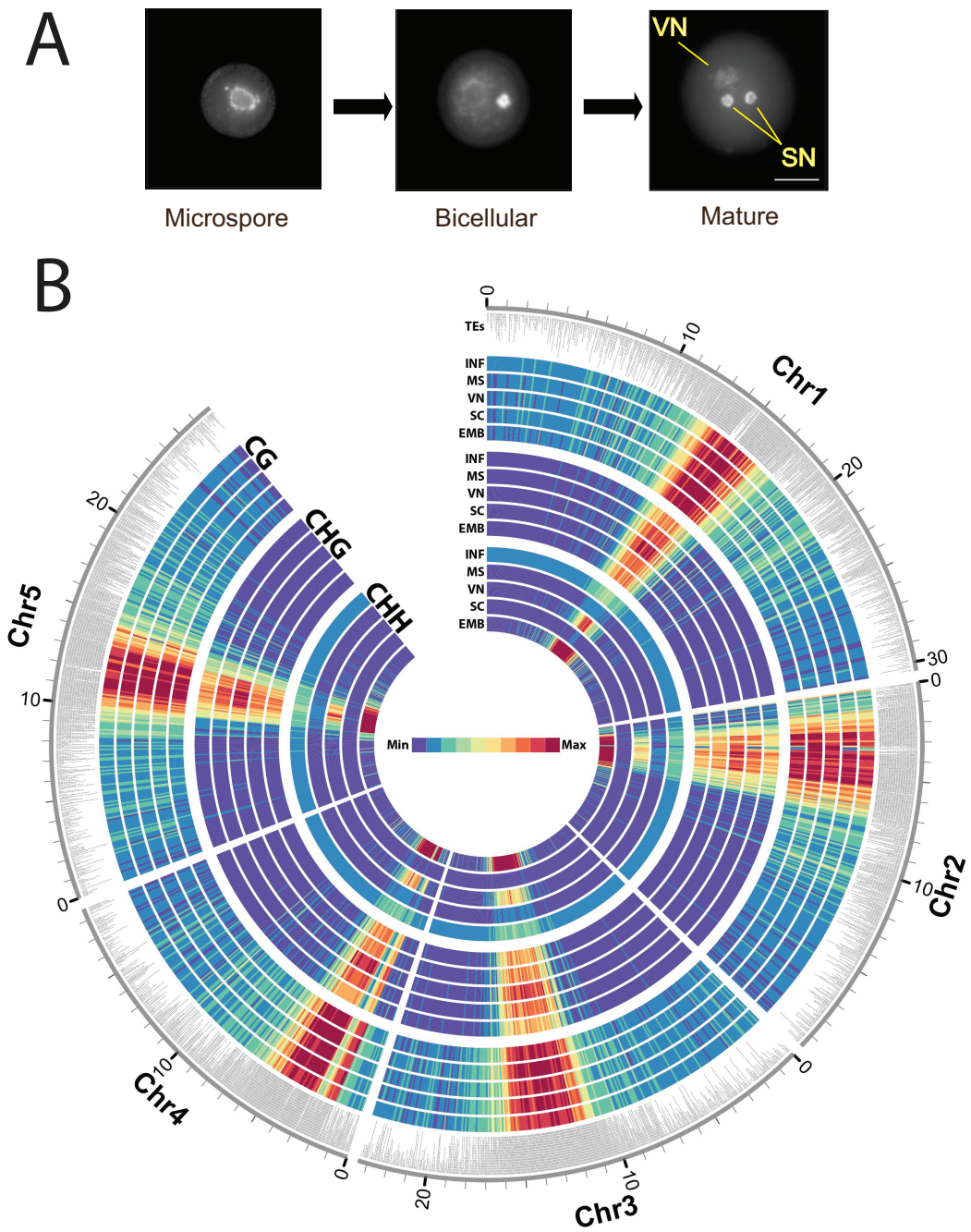


Figure 1. DNA methylation and small RNA accumulation during pollen development. (A) Pollen development: the uninucleate microspore divides asymmetrically giving rise to bicellular pollen, which consists of a larger vegetative cell embedding a smaller generative cell. A second mitotic division of the generative cell originates two sperm cells. The three cell types analyzed in this study were stained with DAPI to highlight heterochromatin, which is lost in the vegetative nucleus (VN) but not in the sperm cell nuclei (SC) (bar = 10μm). **(B)** Heat map representation of DNA methylation. Bisulfite sequencing of genomic DNA from each cell type was performed as described. Methylation density is represented in 10kb blocks, separated by context and cell type. CG (CG methylation), CHG (CHG methylation), CHH (CHH methylation), INF (Inflorescence), MS (Microspore), VN (Vegetative nucleus), SC (Sperm Cell), EMB (Embryo). The maximum value of the heat map is calibrated to the VN. The outer annotation track highlights the position of transposons (TEs).

We observed a strong enrichment of DNA methylation in the pericentromeric heterochromatin in pollen (Figure 1B) resembling methylation profiles obtained previously from somatic cells (Cokus et al., 2008). The observed maintenance of symmetric CG and CHG methylation in pollen is consistent with expression of the maintenance DNA methyltransferases MET1 and CMT3 during microspore and generative cell division (Honys and Twell, 2004). Strikingly, however, CHH methylation in microspores and sperm cells was lost from pericentromeric retrotransposons and satellite repeats, and subsequently restored in the VN (Figure 1B).

Differential methylation of transposons in pollen cell types

To identify regions of the genome subject to differential methylation, we first identified Single Methylation Polymorphisms (SMPs) in a pairwise fashion (VN vs. microspore, SC vs. microspore, and VN vs. SC). Using the SMP information we next identified differentially methylated regions (DMRs). For CHH methylation, DMRs were defined as regions containing at least five SMPs, each < 50bp apart and containing a minimum of ten methylated cytosines. For CG and CHG methylation (which were far less variable), DMRs were defined as regions containing at least three SMPs, each < 50bp apart and containing at least five methylated cytosines. For each putative DMR the methylation calls were pooled across the whole region and then tested using Fisher's exact test.

We found that almost all DMRs corresponded to intergenic regions and transposable elements, and strikingly, that almost all CHH DMRs were hypomethylated in sperm cells while CG DMRs were hypomethylated in the VN (Figure 2A). We found that 2270 CHH DMRs overlapped with 1781 different TEs, including 1483 LTR/Gypsy elements and 139 DNA transposons (Figure 2B, Table S2). Pairwise comparisons of VN vs. microspore and VN vs. SC yielded similar results (Figure 2A,B; Table S2) indicating that these retrotransposons were similarly unmethylated in microspores. An example of an Athila LTR retrotransposon, in which CHH methylation is reduced in microspores and sperm cells, is shown in Figure S1.

We uncovered 221 CG hypomethylated regions (CG DMRs) in the VN relative to the SC (Figure 2, Table S2) that overlapped with 109 different TEs (Table S2), including *AtMu1a* (*At4g08680*), as previously reported (Schoft et al., 2011; Slotkin et al., 2009). 29 of these TEs were RC/helitrons, 34 were DNA/MuDR transposons, and the remainders were mostly non-LTR

retrotransposons (Figure 2B). A similar trend was observed in a pairwise comparison between VN and microspores (Figure 2A,B) and there was a high degree of overlap between CG DMRs in the VN in both pairwise comparisons (Figure 2C). In contrast, CG methylation was very similar in SC and microspores with only a very few loci demethylated in microspores (Figure 2A). These same loci (15/21 DMR) were also demethylated in the CHG context in microspores relative to VN and SC (Figure 2A). CG DMRs in the VN and CHH DMRs in SC did not overlap (Figure 2C), suggesting that differential methylation might be due to differential expression of DNA methyltransferases and DNA demethylases in each pollen cell type.

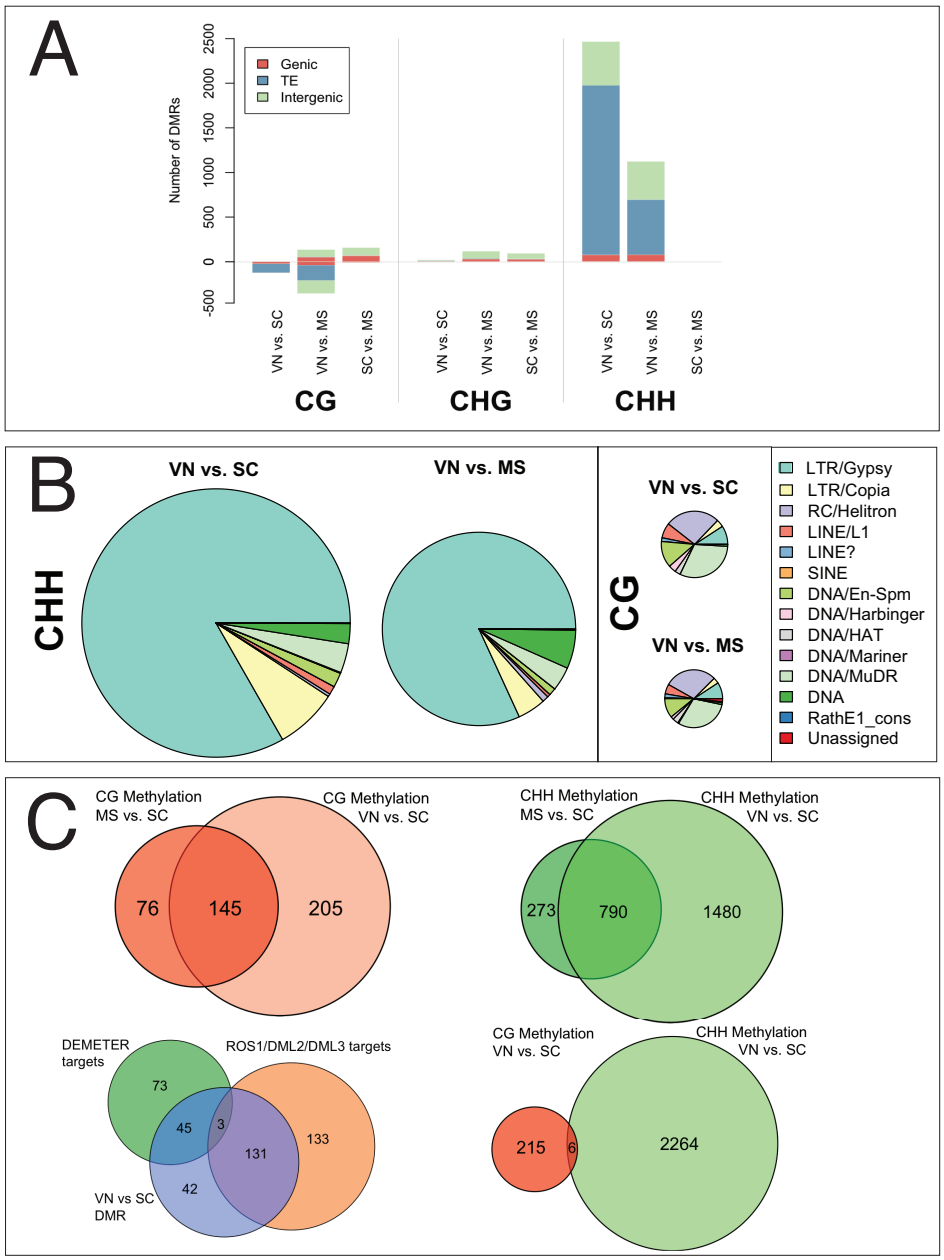


Figure 2. Differentially Methylated Regions (DMRs) during pollen development. (A) DMRs were detected in a pairwise manner by comparing the bisulfite-sequence profiles from each of the three pollen cell types (vegetative nucleus-VN, sperm cell-SC, and microspore-MS) in each methylation context (CG, CHG, CHH). Annotated features (Genic, TE and Intergenic) overlapping one or more DMR in each cell type and methylation context were identified using TAIR10 annotation. Bars represent the number of DMRs overlapping each feature class. Where a DMR overlaps two or more features each feature is counted once. **(B)** Scaled distribution of transposon classes overlapping DMRs in the VN. TEs that matched each DMR were identified. Where a DMR overlaps two or more TE superfamilies each overlap is counted once. DMRs that lost CG methylation in the VN were enriched for class II DNA transposons, while DMRs that lost CHH methylation in sperm cells were enriched for class I LTR/gypsy transposons. There were very few CHG DMRs (data not shown) and these did not overlap transposons. **(C)** CG DMRs (red, upper left) and CHH DMRs (green, upper right) were similar in pairwise comparisons between the VN and the microspore, and the VN and the SC. CG DMRs in the VN (blue, bottom left) overlap with DMRs detected between WT endosperm and *dme* endosperm (green, bottom left), which are targets of DEMETER (Hsieh et al., 2009), and with DMRs between inflorescence and *ros1/dml2/dml3* inflorescence (Lister et al., 2008) which are targets of ROS1 and its homologs (orange, bottom left). In the VN, CG DMRs (pink, bottom right) and CHH DMRs (green, bottom right) do not overlap.

Loss of symmetric CG methylation in vegetative cells

The DNA glycosylase DEMETER (DME) is expressed in the VN, along with its homologs ROS1, DEMETER-LIKE2 (DML2) and DML3 (Schoft et al., 2011). DME is required for the demethylation of transposons and repeats that surround the imprinted Maternally Expressed Genes (MEGs) *MEDEA* (*MEA*) and *FLOWERING OF WAGENINGEN* (*FWA).* These genes are normally expressed from the maternal allele in the endosperm, but are also expressed in the VN of the pollen grain (Schoft et al., 2011). In order to determine whether CG DMRs in the VN were targets of DNA glycosylases, we performed pairwise analysis of CG DMRs between VN and SC, between endosperm and *dme* mutant endosperm (Hsieh et al., 2009), and between WT inflorescence and *ros1/dml2/dml3* mutant inflorescence (Lister et al., 2008). Using the same DMR analytical pipeline, we found 267 targets of ROS1/DML2/DML3 (RDD) in inflorescence, and 121 targets of DME in the endosperm (Hsieh et al., 2009; Lister et al., 2008). Of the 221 DMRs hypomethylated in the VN, 134 DMRs were targets of RDD, and 48 were targeted by DME (Figure 2C). This accounts for 83% of all the DMRs which show decreased CG methylation in the VN compared to SC (Figure 2C). Similar values were obtained for CG DMRs between VN and microspore (Figure S2A). *DME* is only expressed in the VN of pollen and in the central cell of the female gametophyte, while *ROS1*, *DML2* and *DML3* are widely expressed in somatic tissues as well as in the VN. However, none of these genes are expressed in sperm cells (Schoft et al., 2011). Hence, DNA demethylases are responsible for the loss of CG in the VN.

Loss of asymmetric CHH methylation in sperm cells

The overall level of CHH methylation in microspores was approximately half the level found in the inflorescence (Table S1), as if reductional division during meiosis was not accompanied by RNA directed DNA methylation (RdDM). CHH methylation in sperm cells was further reduced, and the remnants were observed on both DNA strands (Table S1), likely reflecting random segregation of unmethylated strands after meiosis (Schoft et al., 2009). We hypothesized that loss of CHH methylation in the SC could be the result of differential expression of proteins required for CHH methylation. The DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2), a homolog of mammalian Dnmt3, is required for CHH methylation, guided by 24nt siRNA (Cao and Jacobsen, 2002). We constructed a DRM2-GFP transgene fusion driven by the DRM2 promoter that was introduced into plants. We found that the DRM2-GFP fusion protein was barely detectable in microspores, but

accumulated prominently in the VN at the bicellular stage (Figure 3). Very low levels were detected in the generative cell and in mature sperm cells (Figure 3), implying that the male germline has only a limited capacity for *de novo* CHH methylation which would account for progressive loss of CHH methylation from microspores to sperm cells.

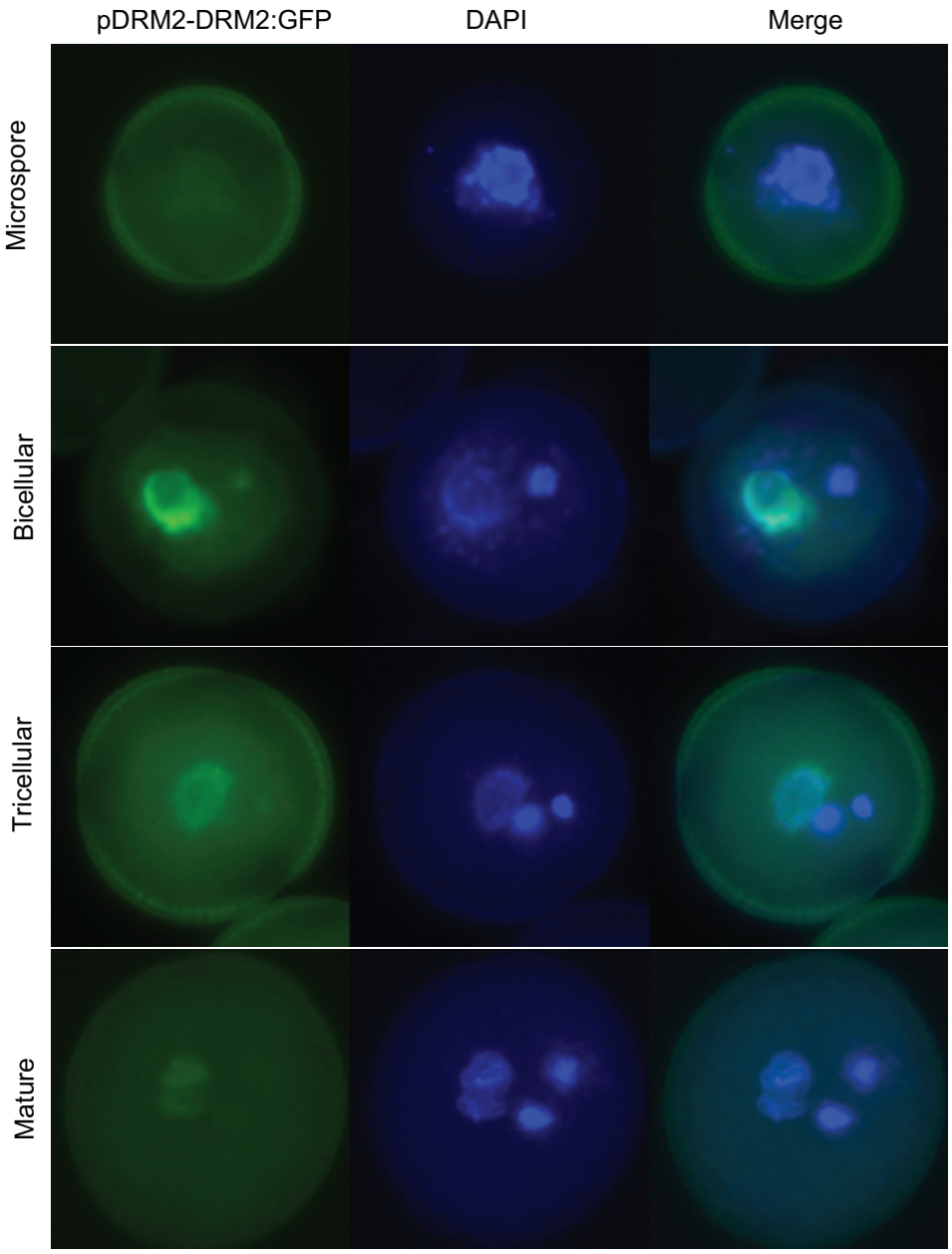


Figure 3. DRM2 expression during pollen development. GFP expression (green) was visualized in pollen from a *pDRM2-DRM2::GFP* transgenic plant, counterstained with DAPI (blue). Microspores and pollen at the bicellular, tricellular and mature stages are shown. DRM2 was expressed at a low level in the microspore and sperm cells, and at a much higher level in the vegetative nucleus at the bicellular and tricellular stage.

Small RNA guide remethylation of transposons and imprinted genes

CHH methylation of retrotransposons is guided by 24nt small RNA (Haag and Pikaard, 2011; Law and Jacobsen, 2010). In sperm cells, CHH methylation is sharply reduced (Table S1; Figure 1B) and several genes required for 24nt siRNA biogenesis are no longer expressed in mature pollen (Grant-Downton et al., 2009; Honys and Twell, 2004; Pina et al., 2005) or sperm (Borges et al., 2008). CHH methylation is restored in the embryo (Hsieh et al., 2009; Jullien and Berger, 2012), and therefore must occur during or after fertilization. 24nt siRNA accumulate to high levels in the seed, and are maternal in origin in the seed coat and the endosperm (Lu et al., 2012; Mosher et al., 2009). Therefore, maternal 24nt siRNA might guide restoration of CHH methylation to incoming retrotransposons from sperm. To test this idea, we examined the size distribution of small RNA in sperm cells (Slotkin et al., 2009) and in seeds (Lu et al., 2012) corresponding to CHH DMRs in pollen (Figure 4). We found that DMRs that had lost CHH methylation in sperm cells matched both 21nt and 24nt siRNA in sperm cells but matched mostly 24nt siRNA in seeds (Figure 4). Thus retrotransposons that lost CHH methylation in sperm cells would be remethylated in seeds, guided at least in part by maternal 24nt siRNA, and high levels of RdDM activity during embryogenesis (Jullien and Berger, 2012).

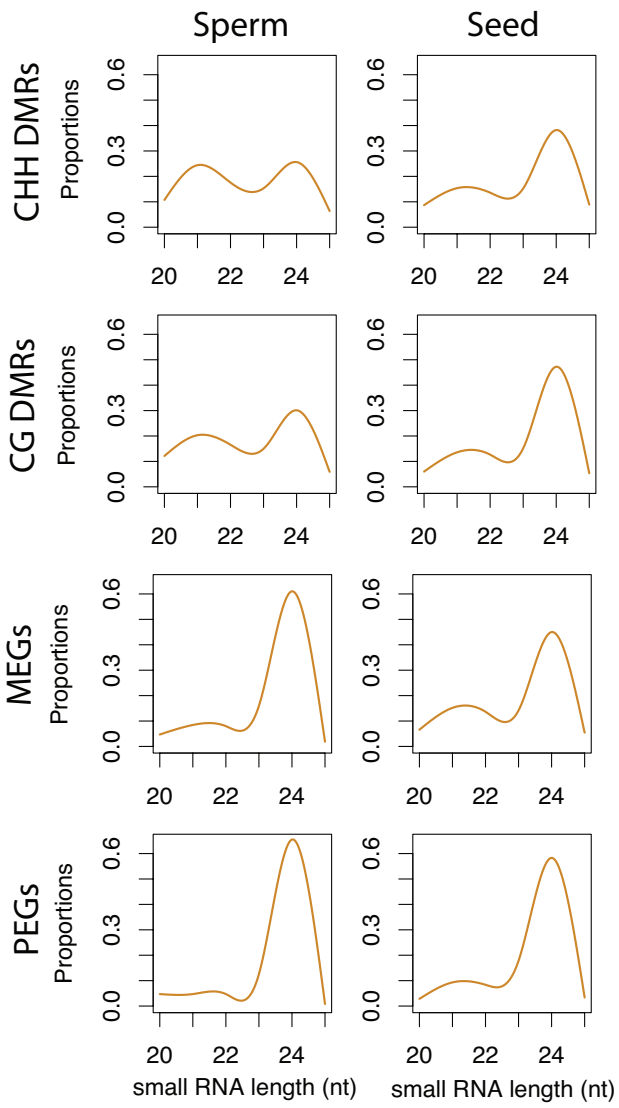


Figure 4. Small RNA from Differentially Methylated Regions (DMRs). Small RNA in sperm cells (Slotkin et al., 2009) and seeds (Lu et al., 2012) were mapped to DMRs and plotted according to size. CHH DMRs were hypomethylated in sperm cells, while CG DMRs were hypermethylated. CG DMRs flanking Maternally and Paternally Expressed imprinted Genes (MEGs and PEGs) were also analyzed separately. Relative abundance of size classes is shown as proportions.

In somatic cells, the activity of DNA glycosylases such as ROS1, DML2 and DML3 results in loss of siRNA production as well as loss of DNA methylation, so that RDD targets tend to gain small RNAs in *rdd* mutants (Lister et al., 2008; Ortega-Galisteo et al., 2008). Sperm cells do not express DME, ROS1, DML2 and DML3 resembling *rdd* mutants in this respect, and we found that many DMRs that lost CG methylation in the VN accumulated siRNA in sperm cells (Figure 4). As many of these CG DMR flank imprinted genes (Gehring et al., 2009), we examined methylation patterns in repeats flanking the imprinted Maternally Expressed Gene (MEG) *SUPPRESSOR OF DRM2/CMT3 (SDC)* and the imprinted Paternally Expressed Gene (PEG) *PHERES1* (*PHE1*) (Figure 5A). *SDC* is expressed when flanking repeats are unmethylated (Henderson and Jacobsen, 2008), but *PHE1* is only expressed when a tandem repeat downstream of the coding sequence is methylated, likely because methylation prevents inhibition by the MEA/FIS2 Polycomb Group (PcG) complex (Makarevich et al., 2008). We found that tandem repeats flanking both genes lose CG methylation in the VN (Figure 5A). However, CHH methylation was only detected at *SDC* and not at *PHE1*. Furthermore, *SDC* accumulated 24nt siRNA in sperm cells (Figure 5A) unlike *PHE1*. The siRNA accumulated to even higher levels in total pollen grains, indicating they may (also) be generated in the VN.

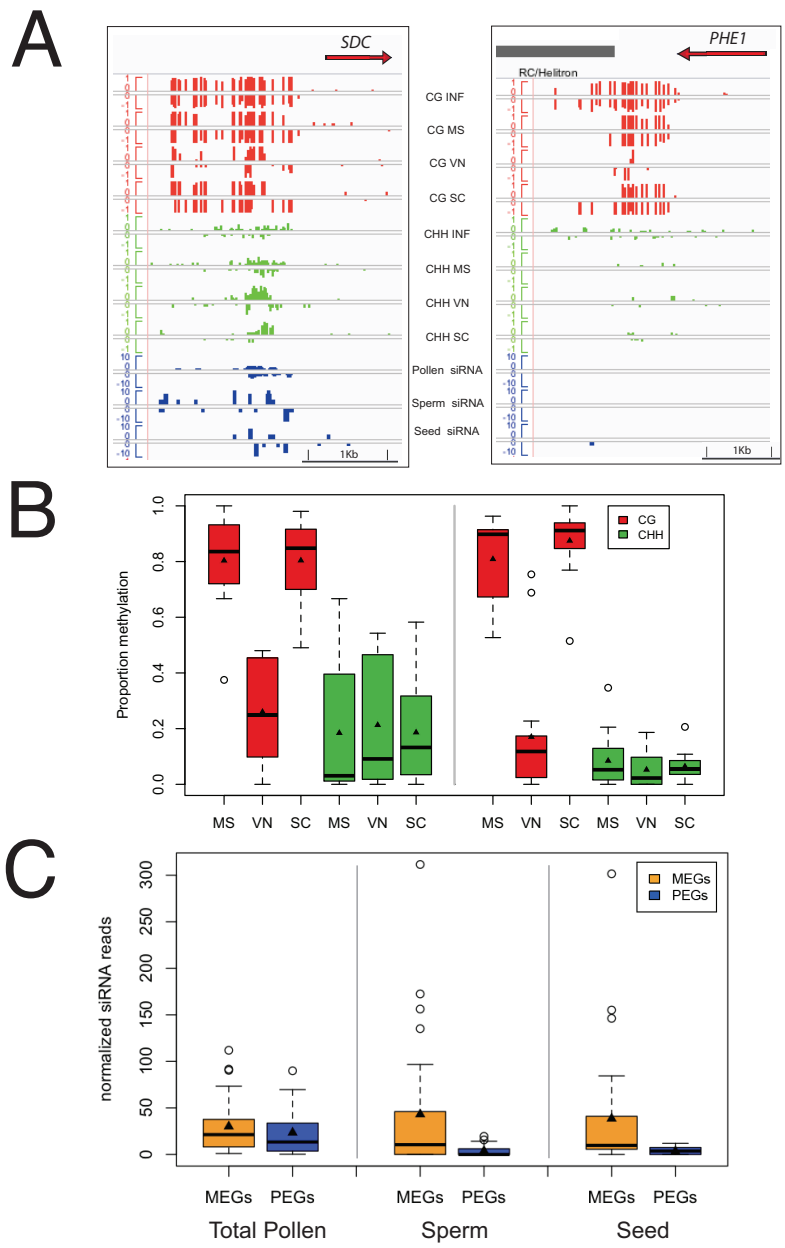


Figure 5. **DNA methylation and small RNA abundance at imprinted genes in pollen**. **(A)** Genome browser view of the Maternally Expressed Gene (MEG) *SDC* and the Paternally Expressed Gene (PEG) *PHE1*. Tracks display CG (red) and CHH (green) methylation as well as 24nt siRNAs (blue) from pollen, seeds and purified sperm cells. Methylation is represented on a scale of 0-100% and siRNAs for total normalized reads from 0-20 RPM (reads per million). MS (microspore), SC (sperm cell), VN (vegetative nucleus), INF (Inflorescence). **(B)** Box-plot representation of DNA methylation percentages at MEGs and PEGs. TEs neighboring both MEGs and PEGs are demethylated in the CG context specifically in the vegetative nucleus. Higher CHH methylation levels were detected at MEGs in comparison with PEGs. **(C)** Box plot representation of 24nt siRNA corresponding to TEs surrounding PEGs and MEGs in total pollen, sperm cells, and seeds. Boxes represent lower and upper quartiles surrounding the median (line). Triangles represent the mean.

We extended these observations to a larger number of putative imprinted genes (Gehring et al., 2011; Hsieh et al., 2011; McKeown et al., 2011; Wolff et al., 2011) filtered to include only experimentally validated PEGs and MEGs, resulting in 28 imprinted loci (12 MEGs and 16 PEGs) that passed our filter for methylation calls and had a TE within 2kb of the coding sequence (Table S3). All 28 TEs lost CG methylation in the VN relative to the progenitor microspore, but interestingly only those surrounding MEGs were targeted by siRNA and CHH methylation in pollen (Fig 5B). We plotted the size distribution of siRNA corresponding to CG DMRs, and found that while CG DMRs accumulated both 21 and 24nt siRNA in sperm cells, MEGs and PEGs accumulated only 24nt siRNA in sperm and seeds (Fig 4). siRNA levels for MEGs were higher than PEGs in sperm cells and in seeds, but not in total pollen (Fig 5C). We conclude that 24nt siRNA from repeats surrounding MEGs accumulate preferentially in sperm cells. It is possible that these are derived from the VN, resembling 21nt siRNA in this respect (Slotkin et al.)

Reprogramming leads to spontaneous epigenetic variation

In plants, epigenetic changes in gene expression are frequently inherited from one generation to the next, and gains and losses of DNA methylation arise as spontaneous epigenetic variation (Martienssen and Colot, 2001). In two recent studies, more than 100 loci (DMRs) were found to gain DNA methylation sporadically in young leaf tissue after 30 generations of inbreeding by single seed descent (Becker et al., 2011; Schmitz et al., 2011). Methylation gains were recurrent, occurring at the same loci in multiple independent lines, leading to the proposal that methylation gains and losses might be pre-programmed in the germline (Schmitz et al., 2011). Among 100 hypervariable loci that gain methylation, we identified several ROS1/DML2/DML3 (RDD) targets that were completely re-methylated in *rdd* mutants compared to wild-type (Lister et al., 2008). Most of the remaining hypervariable loci already showed high methylation levels in wild-type inflorescence tissue (Lister et al., 2008). Remarkably, we observed that 56 of these 100 variable DMRs were hypermethylated in wild-type sperm cells. An example of a RDD target, corresponding to one of the hypervariable epialleles, is shown in Figure 6. This target is contained within a COPIA element (*Atg409455)* that is heavily methylated at CG sites in sperm cells, and less so in the microspore and VN (Figure 6). Further examples are shown in Figure S3. *DME*, *ROS1, DML2* and *DML3* are expressed at low levels in the microspore (Honys and Twell, 2004), and high levels in the VN (Schoft et al., 2011), accounting for differential CG methylation observed in sperm. Importantly, CG methylation found in sperm cells was removed in the embryo, reflecting the restoration of ROS1 activity after fertilization (Fig 6).

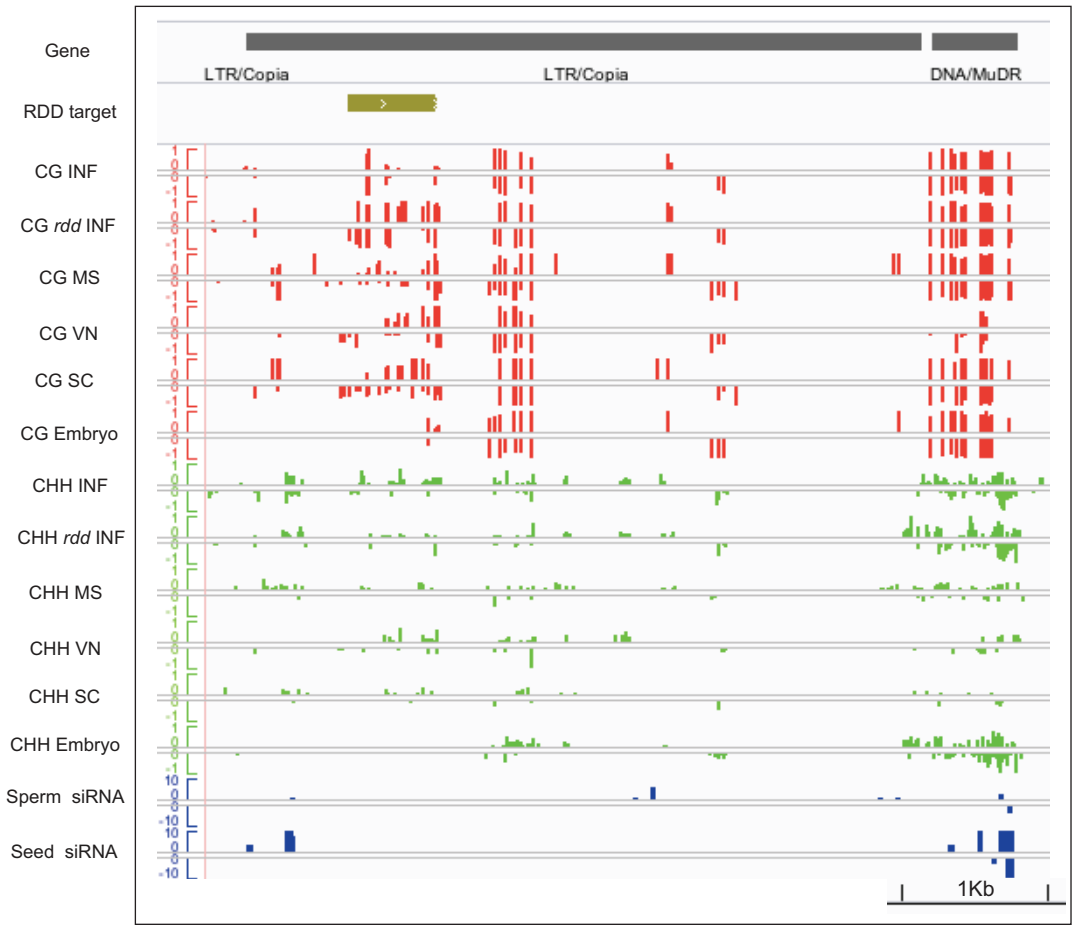


Figure 6. DNA methylation at hypervariable recurrent epialleles. 100 hypervariable epialleles gain DNA methylation recurrently in plants propagated by single seed descent (Becker et al., 2011; Schmitz et al., 2011). Many are targets of ROS1 and its homologs DML1 and DML2 (RDD). An example is shown (*ATCOPIA51, At4g09455*), along with a neighboring MuDR element for comparison. Tracks represent the RDD target region, and methylation levels in CG and CHH contexts in microspores (MS), vegetative nucleus (VN), and sperm cells (SC), along with inflorescence (INF) and embryo. CG methylation at the RDD target site is found in *rdd* triple mutant inflorescence (*rdd* INF) (Lister et al., 2008) and in pollen, but not in inflorescence or embryo. Small RNA from sperm cells (Slotkin et al., 2009) and seed (Lu et al., 2012) are also shown.

Discussion

In mammals, 5-methylcytosine occurs mainly in symmetric CG dinucleotides, and is depleted in male primordial germ cells by loss of DNA methyltransferases and by active demethylation (Feng et al., 2010; Popp et al., 2010) resulting in TE activation (Castaneda et al., 2011). Methylation is restored in mature round spermatids (Feng et al., 2010; Popp et al., 2010) and then extensively modified by hydroxylation just before fertilization (Salvaing et al., 2012; Zhang et al., 2012). Further rounds of methylation and demethylation occur in the blastocyst and early embryo (Feng et al., 2010) resulting in a complex pattern of DNA methylation that is reset in each generation (Bartolomei and Ferguson-Smith, 2011). In pollen, we have found that symmetric CG and CHG methylation are largely retained in the germline (Figure 7). This may account for the prevalence of epigenetic inheritance in plants, compared with mammals. Strikingly, however, asymmetric CHH methylation of transposons is reduced in the microspore, accompanied by down regulation of the RdDM methyltransferase DRM2, a homolog of the mammalian Dnmt3 (Figure 7). CHH methylation is restored in the embryo, and may reflect an ancient mechanism for transposon recognition.

Transposon reprogramming in pollen

The loss of asymmetric CHH methylation in sperm cells means that paternal retrotransposons are delivered to the zygote stripped of CHH methylation. Restoration of DNA methylation in the embryo (Hsieh et al., 2009), indicates that CHH methylation must occur during or after fertilization, when the RdDM pathway is active (Jullien and Berger, 2012) (Figure 7). We demonstrate that 24nt siRNAs in seeds match retrotransposons that have lost CHH methylation in sperm (Figure 6). CHH methylation is restored in seeds (Hsieh et al., 2009), guided by these 24nt siRNAs (Jullien and Berger, 2012). It has been proposed that most 24nt small RNA in seeds are maternal in origin, especially in the seed coat and the endosperm (Mosher et al., 2009), and target retrotransposons (Lu et al., 2012). We can speculate that paternal retrotransposons that have lost CHH methylation, but do not match maternal siRNA, might escape silencing immediately after fertilization (Josefsson et al., 2006).

DRM2 expression is restored in the VN, and retrotransposons are remethylated in these companion cells (Figure 7), most likely at the bicellular stage when DCL3 and other components of the 24nt siRNA biogenesis pathway are expressed (Grant-Downton et al., 2009). However,

TEs are strongly activated in the VN and give rise to mobile 21nt siRNA that accumulate in sperm cells (Slotkin et al., 2009). CHH methylation by RdDM would not be expected to prevent transcription in the absence of the chromatin remodeler DDM1 (Teixeira et al., 2009), which is not expressed in the VN, accounting for transposon activation (Slotkin et al., 2009). Loss of chromatin remodeling can result in transposon transcription even in the presence of DNA methylation (Lorkovic et al., 2012; Mittelsten Scheid et al., 2002; Moissiard et al., 2012; Vaillant et al., 2006). Furthermore, the VN undergoes extensive histone replacement, with the loss of many canonical histones including the centromeric histone CENH3, which may contribute to transposon activation (Berger el al, 2011; Schoft et al., 2009). It is possible therefore that CHH methylation in the VN compensates for the loss of pericentromeric heterochromatin(Schoft 2009).

Reprogramming of imprinted genes

Although CG methylation was globally retained, a subset of DNA transposons, some non-LTR retrotransposons, and intergenic regions lost CG methylation in the VN (Figure 7). These TEs are targets of the DNA glycosylases DME, ROS1, DML2/3, which are expressed in the VN. In sperm, these enzymes are not expressed, and 24nt siRNA corresponding to some of their targets accumulate, resembling *ros1/dml2/dml3* triple mutants in this respect (Lister et al., 2008). This is particularly true for TEs that flank imprinted genes which are expressed from the maternal allele in the endosperm (MEGs), and imprinting at the *SDC* locus is lost in the endosperm when inherited from mutant pollen impaired in RdDM (Vu et al., 2012). These results indicate that 24nt siRNA in sperm cells contribute to RdDM and transcriptional silencing before fertilization in at least some cases (Fig 7). Like 21nt siRNA, these specific 24nt siRNA may also be derived from the VN, although this has not been tested directly (Fig 7). In this way, imprinted genes are protected from the global loss of methylation, reminiscent of mammalian imprinted genes, which regain methylation in the germline before fertilization (Feng et al, 2010).

Many imprinted genes are expressed in pollen (Table S3), and *dme* mutants are transmitted poorly because of defective pollen germination (Schoft et al., 2011). Similarly, *ros1* mutants exhibit severe fertility defects after 3 generations of inbreeding (Gong et al., 2002). It is likely therefore that the targets of DME and ROS1 play a role in fertilization when the vegetative nucleus supports pollen tube growth (Berger and Twell, 2011). Silencing in sperm cells would restrict expression to the pollen tube, as well as resulting in imprinting in the endosperm. A small number of targets were also demethylated in the microspore, and may have a function earlier in pollen development (Figure S2B).

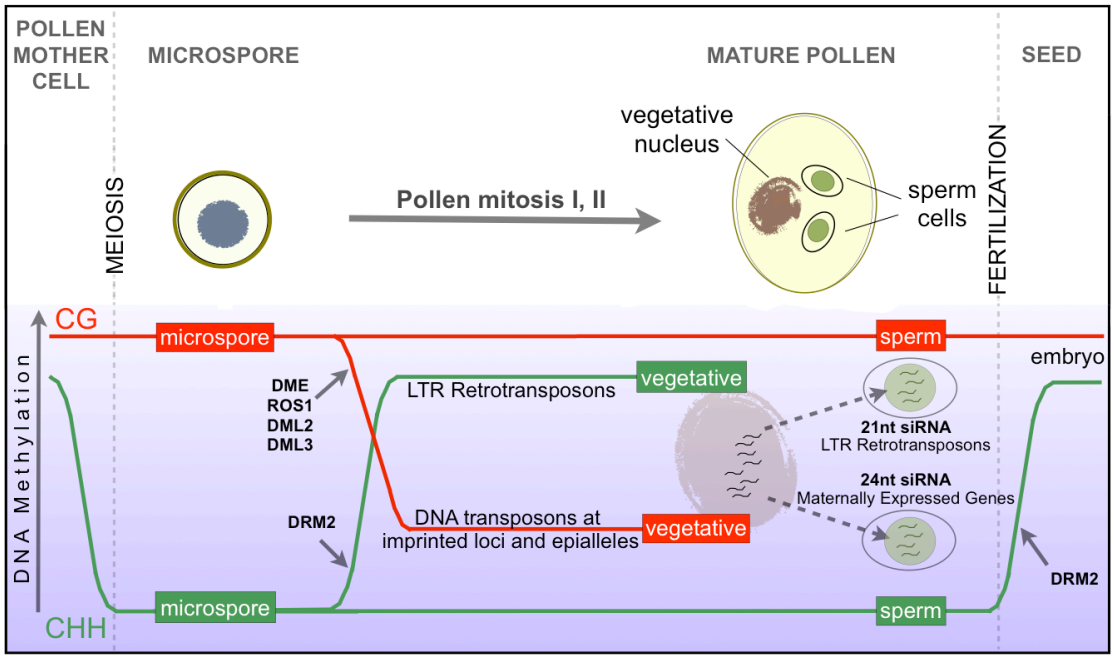


Figure 7. Genome reprogramming during pollen development. Differential expression of DME, ROS1, DML2 and DML3 are depicted together, being present only in the vegetative cell nucleus after mitosis I. DRM2, on the other hand shows an large increase in expression in the vegetative cell nucleus compared to the VN and SN. The expression of these particular genes is reflected in the levels of DNA methylation observed in each cell type, with CHH methylation (green line) increasing in the vegetative cell while it remains at a low level in the sperm cell. CG methylation (red line) instead shows a specific and targeted decrease in methylation but only at DNA transposons surrounding imprinted loci and epialleles. Also depicted are 21nt siRNAs, which tend to correlate to regions that showed CHH re-methylation in the VN and 24nt siRNAs that tended to map to regions that surrounded maternally expressed imprinted loci.

Epigenetic inheritance in the plant germline

Similar silencing mechanisms may account for the methylation we observe in sperm at hypervariable epialleles. These epialleles acquire heritable methylation sporadically on inbreeding, prompting speculation that they might be reprogrammed in sperm (Becker et al., 2011; Schmitz et al., 2011). Some of these epialleles are silenced in *ros1/dml2/dml3* mutants, and many correspond to TEs (Schmitz et al., 2011). We show that these variable epialleles are indeed methylated in sperm cells, and that many of them are methylated already in the inflorescence (Lister et al., 2008). Sperm cells do not express ROS1 and its homologs, accounting for higher methylation of RDD targets in sperm, and providing a mechanism for gain of heritable methylation if 24nt siRNA accumulate after fertilization to prevent demethylation by ROS1.

When transposon methylation is lost, it can be regained through RNAi (Teixeira et al., 2009) which seems to occur stepwise in subsequent generations consistent with its occurrence in the germline (Teixeira and Colot, 2010). Loss and gain of class II DNA transposon activity in maize occurs over generations (McClintock, 1965), during development (Li et al., 2010; Martienssen and Baron, 1994; Martienssen and Colot, 2001) and is inherited in the germline resembling the epialleles recently described in *Arabidopsis* (Becker et al., 2011; Schmitz et al., 2011). Our results suggest that similar epigenetic mechanisms silence epialleles and imprinted genes in pollen, which escape reprogramming in subsequent generations because of the retention of DNA methylation in sperm.

References

Bartolomei, M. S., and Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. Cold Spring Harb Perspect Biol *3*, a002592.

Becker, C., Hagmann, J., Muller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245-249.

Berger, F., and Twell, D. (2011). Germline specification and function in plants. Annu Rev Plant Biol *62*, 461-484.

Boavida, L. C., Becker, J. D., and Feijo, J. A. (2005). The making of gametes in higher plants. Int J Dev Biol *49*, 595-614.

Borges, F., Gardner, R., Lopes, R. T., Calarco, J. P., Slotkin, R. K., Martienssen, R. A., and Becker, J. D. (2012). FACS-based Purification of Arabidopsis Microspores, Sperm Cells and Vegetative Nuclei. BMC Plant Methods *In press*.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijo, J. A., and Becker, J. D. (2008). Comparative Transcriptomics of Arabidopsis thaliana Sperm Cells. Plant Physiol *148*, 1168-1181.

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

Castaneda, J., Genzor, P., and Bortvin, A. (2011). piRNAs, transposon silencing, and germline genome integrity. Mutat Res *714*, 95-104.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

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

Feng, S., Jacobsen, S. E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

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

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PLoS One *6*, e23687.

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldan-Arjona, T., David, L., and Zhu, J. K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

Grant-Downton, R., Hafidh, S., Twell, D., and Dickinson, H. G. (2009). Small RNA pathways are present and functional in the angiosperm male gametophyte. Mol Plant *2*, 500-512.

Haag, J. R., and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol *12*, 483-492.

Henderson, I. R., and Jacobsen, S. E. (2008). Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. Genes Dev *22*, 1597-1606.

Honys, D., and Twell, D. (2004). Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biology *5*, R85.

Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science *324*, 1451-1454.

Hsieh, T. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., Hashimoto, M., Kirkbride, R. C., Harada, J. J., Zilberman, D., and Fischer, R. L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A *108*, 1755-1762.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Current Biology *16*, 1322-1328.

Jullien, P. E., and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Opin Plant Biol *12*, 637-642.

Jullien, P. E., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Current Biology *In press*.

Jullien, P. E., Kinoshita, T., Ohad, N., and Berger, F. (2006). Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell *18*, 1360-1372.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Li, H., Freeling, M., and Lisch, D. (2010). Epigenetic reprogramming during vegetative phase change in maize. Proc Natl Acad Sci U S A *107*, 22184-22189.
Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D.*, et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471-476.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Lorkovic, Z. J., Naumann, U., Matzke, A. J., and Matzke, M. (2012). Involvement of a GHKL ATPase in RNA-directed DNA methylation in Arabidopsis thaliana. Current Biology *22*, 933-938.

Lu, J., Zhang, C., Baulcombe, D. C., and Chen, Z. J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. Proc Natl Acad Sci U S A *109*, 5529-5534.

Makarevich, G., Villar, C. B., Erilova, A., and Kohler, C. (2008). Mechanism of PHERES1 imprinting in Arabidopsis. J Cell Sci *121*, 906-912.

Martienssen, R., Barkan, A., Taylor, W. C., and Freeling, M. (1990). Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. Genes Dev *4*, 331-343.

Martienssen, R., and Baron, A. (1994). Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. Genetics *136*, 1157-1170.

Martienssen, R. A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science *293*, 1070-1074.

McClintock, B. (1965). The control of gene action in maize. Brookhaven Symposia in Biology *18*, 162-184.

McKeown, P. C., Laouielle-Duprat, S., Prins, P., Wolff, P., Schmid, M. W., Donoghue, M. T., Fort, A., Duszynska, D., Comte, A., Lao, N. T.*, et al.* (2011). Identification of imprinted genes subject to parent-of-origin specific expression in Arabidopsis thaliana seeds. BMC Plant Biol *11*, 113.

Mittelsten Scheid, O., Probst, A. V., Afsar, K., and Paszkowski, J. (2002). Two regulatory levels of transcriptional gene silencing in Arabidopsis. Proc Natl Acad Sci U S A *99*, 13659-13662.

Moissiard, G., Cokus, S. J., Cary, J., Feng, S., Billi, A. C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B. R., McCord, R. P.*, et al.* (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science *336*, 1448-1451.

Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J., and Baulcombe, D. C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature *460*, 283-286.

Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R., and Roldan-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol *67*, 671-681.

Pina, C., Pinto, F., Feijo, J. A., and Becker, J. D. (2005). Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol *138*, 744-756.

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Radford, E. J., Ferron, S. R., and Ferguson-Smith, A. C. (2011). Genomic imprinting as an adaptative model of developmental plasticity. FEBS Lett *585*, 2059-2066.

Raissig, M. T., Baroux, C., and Grossniklaus, U. (2011). Regulation and flexibility of genomic imprinting during seed development. Plant Cell *23*, 16-26.

Salvaing, J., Aguirre-Lavin, T., Boulesteix, C., Lehmann, G., Debey, P., and Beaujean, N. (2012). 5-methylcytosine and 5-hydroxymethylcytosine spatiotemporal profiles in the mouse zygote. PLoS One *7*, e38156.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet *34*, 65-69.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., and Ecker, J. R. (2011). Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. Science *334*, 369-373.

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J. S., Park, G. T.*, et al.* (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A *108*, 8042-8047.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., and Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461-472.

Teixeira, F. K., and Colot, V. (2010). Repeat elements and the Arabidopsis DNA methylation landscape. Heredity *105*, 14-23.

Teixeira, F. K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M. F.*, et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Vaillant, I., Schubert, I., Tourmente, S., and Mathieu, O. (2006). MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep *7*, 1273-1278.

Vu, T. M., Nakamura, M., Calarco, J. P., Susaki, D., Lim, P. Q., Kinoshita, T., Higashiyama, T., Martienssen, R., and Berger, F. (2012). RNA-directed DNA methylation controls parental imprinting in Arabidopsis. Submitted.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Kohler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. PLoS Genet *7*, e1002126.

Zhang, P., Su, L., Wang, Z., Zhang, S., Guan, J., Chen, Y., Yin, Y., Gao, F., Tang, B., and Li, Z. (2012). The involvement of 5-hydroxymethylcytosine in active DNA demethylation in mice. Biol Reprod *86*, 104.

Chapter 4
DDM1 purification, antibody design and attempted crystallization

DDM1 is involved in heterochromatin silencing in Arabidopsis

Once mapped after its initial discovery, DDM1 was surprisingly found to encode a putative SWI2/SNF2 chromatin remodeler (Jeddeloh et al., 1999). The screen it was originally isolated in was designed to uncover genes necessary for maintaining centromeric DNA methylation (Jeddeloh et al., 1999) (Vongs et al., 1993) and as such has a molecular phenotype indicating a necessary role for CG methylation maintenance at the centromere and throughout the genome, which is why DDM1 was originally thought to be a DNA methyltransferase. This observation is not surprising, as molecular phenotypes in *ddm1* are not limited to DNA methylation, as additionally defects are observed in centromeric H3K9 methylation, which results in a loss of transposon silencing, as well as perturbations in the small RNA pathways. Together, these phenotypes suggest that DDM1 is necessary for heterochromatin regulation in Arabidopsis (Lippman et al., 2004) (Soppe et al., 2002) (Figure 1). Interestingly, backcrossing cannot restore defects in *ddm1* plants, as heterochromatin silencing perturbations and epimutations are inherited by the offspring (Lippman et al., 2003), and only restored to wild type levels after multiple generations of backcrossing. This phenotype is reminiscent of *met1* mutants and points to the phenomenon in plants where CG methylation is not properly restored after it is lost, even after multiple generations (Mathieu et al., 2007). A recent study investigating the trans-generational effects after loss of CG methylation suggests that while CG methylation may not be properly restored after it is lost, the RNAi machinery will try to re-establish heterochromatin silencing/function through CHH (where H is any nucleotide except G) DNA methylation (Mathieu et al., 2007). This same observation was extended to be reliant on DDM1, as *de novo* methylation guided by siRNAs only drove the re-establishment and spreading of DNA methylation in the presence of DDM1 (Teixeira et al., 2009).

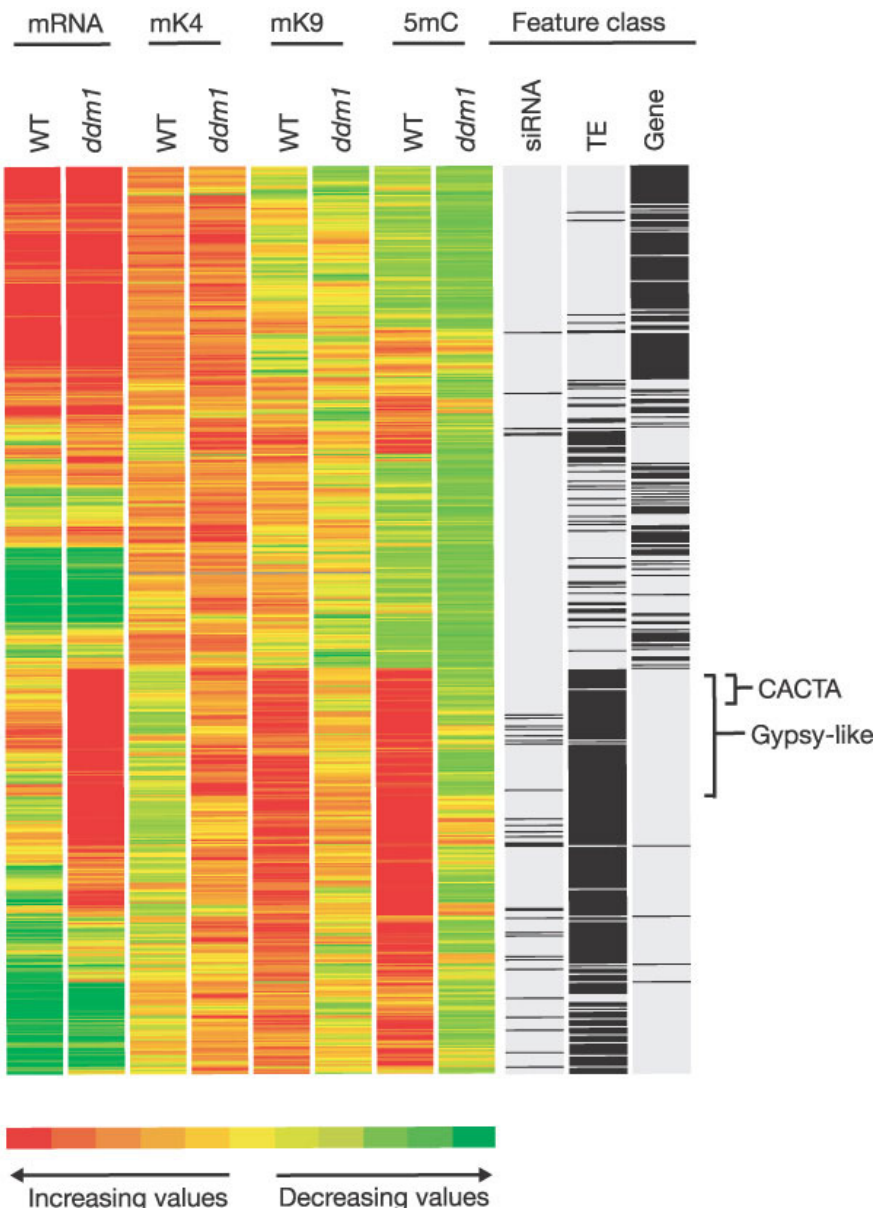


Figure 1- *ddm1* vs. wt molecular phenotypes Cluster analysis of wild type vs. ddm1 mutants for heterochromatic knob on Chromosome 4. Increased mRNA expression of TEs, which correlates with increased H3K4me3, decreased H3K9me3 and 5mC marks. Genic regions are less affected. **Figure from Lippman *et al. 2004*

Over 20 years without a mechanism

Currently there is very little evidence to suggest a model for DDM1's mechanistic role in heterochromatin silencing. While numerous genetic studies have investigated *ddm1* phenotypes to try and uncover its specific mechanistic role, the conclusions have been limited. This is mostly due to the interplay between different silencing machinery (DNA methylation, histone modification, and small RNAs) (Lippman et al., 2004; Lippman et al., 2003; Soppe et al., 2002), suggesting that DDM1's role is upstream of all of these processes. However, there is some evidence from studies on DDM1, and its homologs, to suggest a molecular mechanism. The most accepted model of DDM1 function is that it is recruited to specific sites on the chromatin to remodel the nucleosomes and allow access of the DNA to MET1 for CG methylation. This model is supported by the strong mCG perturbations of *ddm1*, the similarity of *ddm1* and *met1* phenotypes, and nucleosome sliding assays that have suggested a weak activity of DDM1 to reposition nucleosomes *in vitro* (Brzeski and Jerzmanowski, 2003; Lippman et al., 2004; Lippman et al., 2003; Mathieu et al., 2007; Saze et al., 2003). However, this model appears to be at least partially inaccurate, as *ddm1* and *met1* phenotypes do not completely overlap and the double mutant shows more severe defects than either single mutant (Kankel et al, 2003) . Additionally, DDM1 has been implicated in DNA repair as it's yeast homolog, YFR038W, interacts with a subunit of the mediator complex and may play a role in meiotic recombination (Robinson and Schultz, 2005). Furthermore, LSH1, the mouse homolog of DDM1, has been shown to interact with DNMT3a and 3b, homologs of the DNA methyltranserases implicated in RNAi mediated DNA methylation in plants (Zhu et al., 2006). Curiously, LSH1 did not interact with DNMT1, the MET1 homolog (Zhu et al., 2006). While there has been a lot of progress in understanding the function of DDM1 in the cell, there is still much more to be investigated.

It is thought that SWI/SNF chromatin-remodeling complexes regulate gene expression by disrupting histone-DNA interactions, thus permitting a variety of proteins to access DNA. These ATP binding proteins fall into at least five highly conserved subfamilies, three of which are the yeast SWI2/SNF2, *Drosophila* ISWI, and mouse CHD1 subfamilies. They have been implicated in both transcriptional activation and repression. DDM1 in Arabidopsis, LSH1 (mouse), and YFR038W (yeast) represent a fourth subfamily that lacks C-terminal SLIDE and HAND DNA binding domains which are found in ISWI proteins. They also lack the chomodomains found in CHD1. Essentially we have a gene that is involved in CG DNA methylation, histone methylation and affects transposon movement, yet has no obvious known protein motifs that can provide any insight into how it influences these processes. As such, we attempted to employ structural, biochemical and genetic techniques to try and tease apart DDM1's role in these processes.

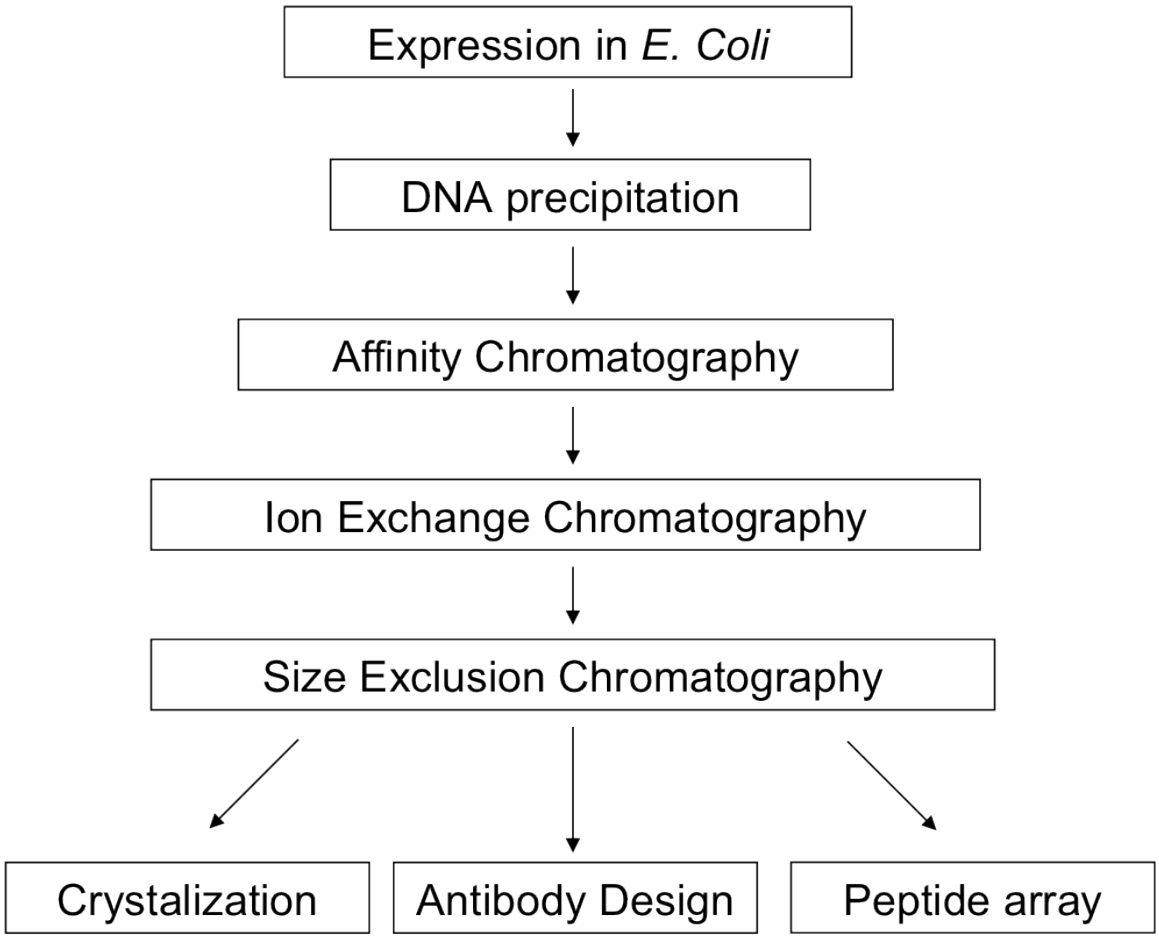


Figure 2 – Schematic of final purification strategy for recombinant DDM1. Essential to a functional purification strategy was a chemical based DNA precipitation that limited the formation of large aggregates and increased the yield of functional DDM1 obtained after affinity chromatography. Ion exchange chromatography was essential in separating non-specific binders to both DDM1 and the affinity resin. Size exclusion chromatography ensured a final sample of pure unaggregated protein. The final purified product was then used for a variety of downstream applications.

Recombinant DDM1 purification

Essential to these proposed studies would be sufficient quantities of purified DDM1 protein for use in both biochemical and structural studies, both of which would aid in the creation of an anti-DDM1 antibody with high specificity. To go about this I made a fusion of DDM1 to an N-terminal hexahistidine tag in a pET28 expression vector for *E. coli.* The final expression/ purification protocol involves inducing *E. coli* cultures with 100mM IPTG at and optical density of ~0.6 and expressing protein overnight at 18^{o}C. Pelleted cells are re-suspended in 20mM HEPES, 1M KCl pH 8.0 in order to dissociate any potential DNA-protein ionic interactions. I then perform a polyethylenimine treatment which precipitates DNA and load the cleared lysate on a Ni-NTA column. A number of washes are performed with increasing concentrations of imidizole (up to 20mM in the wash buffers) and then finally eluted into 20mM HEPES, 200mM KCl pH 7.5 with 200mM imidizole.

The eluate is then passed through a cation exchange column, which greatly increases the purity of the sample by eliminating aggregated and misfolded protein. Specifically, a Mono S 5/50 GL column (GE life sciences) was run at a flow rate of 2mL/min for 30 column volumes. The sample was loaded in 20mM HEPES, 200mM KCl pH 7.5 with 200mM imidizole and run against a buffer containing 20mM HEPES, 200mM KCl and 1M NaCl pH 7.5. The eluate containing the most pure, unaggregated DDM1 appeared at volume 112-114 ml (denoted by arrow Figure 3).

The purified 2mL fraction from the cation exchange column was loaded on a gel filtration column, the Superdex 200 10/300 GL. The column was equilibrated with a buffer containing 20mM HEPES, 200mM KCl pH 7.5. The sample was run at a flow rate of 0.5ml/min and protein absorbance was followed at 280nm. The first peak (orange arrow, figure 4) which arrives ~15mins represents the void volume of the column (aggregated protein sample which does not enter the matrix). The sample which eluted at ~30mL (denoted by a black arrow, figure 4) represents protein of ~80 kDa, corresponding the size of DDM1 protein with HIS tag. Purity here is assessed by a combination of peak architecture and coomassie staining (figure 4B).

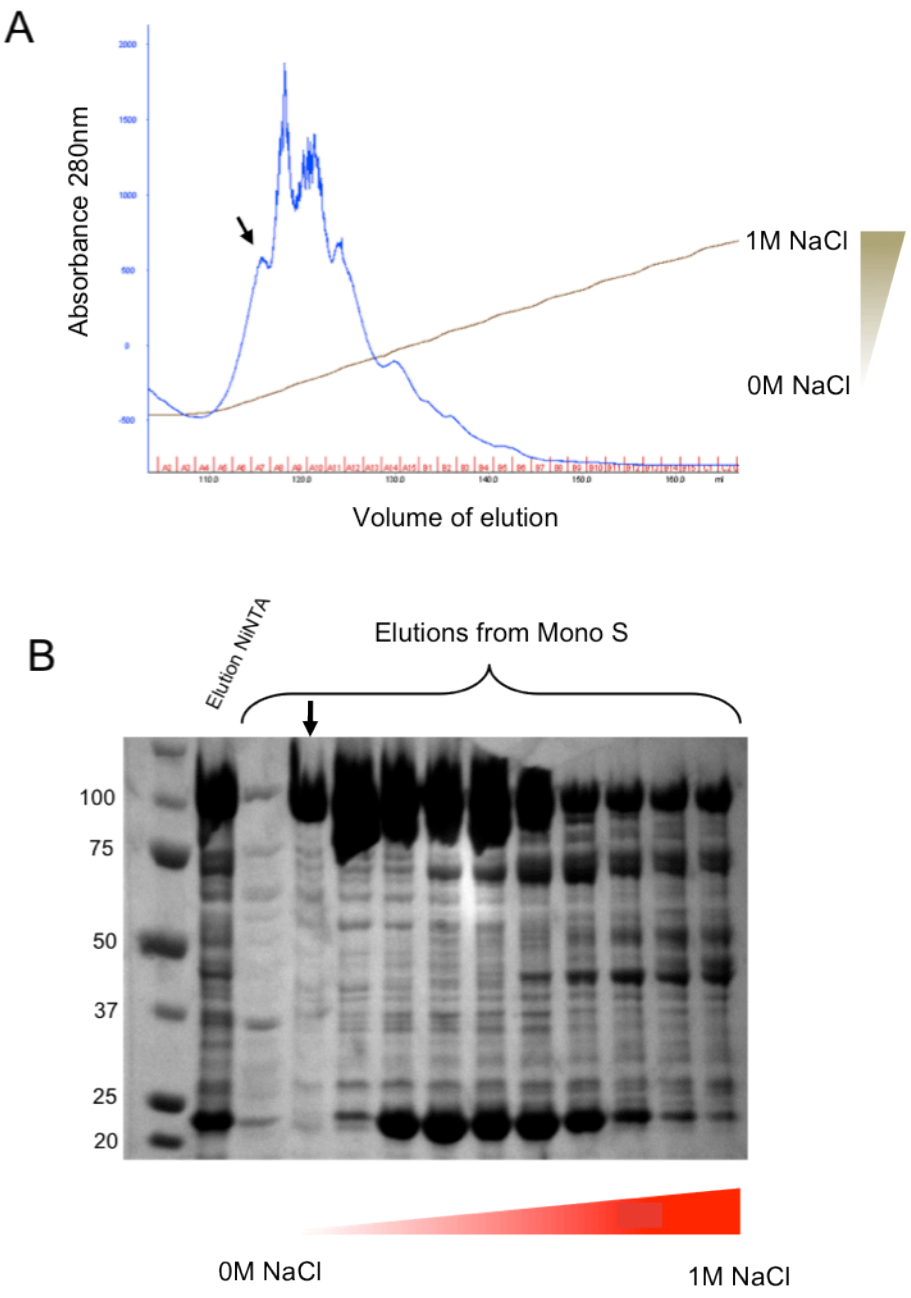


Figure 3 – Mono S Cation Exchange chromatography – Induced, cleared and affinity purified DDM1 is loaded onto a GE Mono S 5/50 GL cation exchange column and eluted with increased NaCl, up to 1M. The first peak, annontated with a black arrow in A and B indicates the peak containing DDM1. The gel in (B) shows the fractions of protein obtained. The purist fraction based on coomassie staining of the eluted fractions from the column

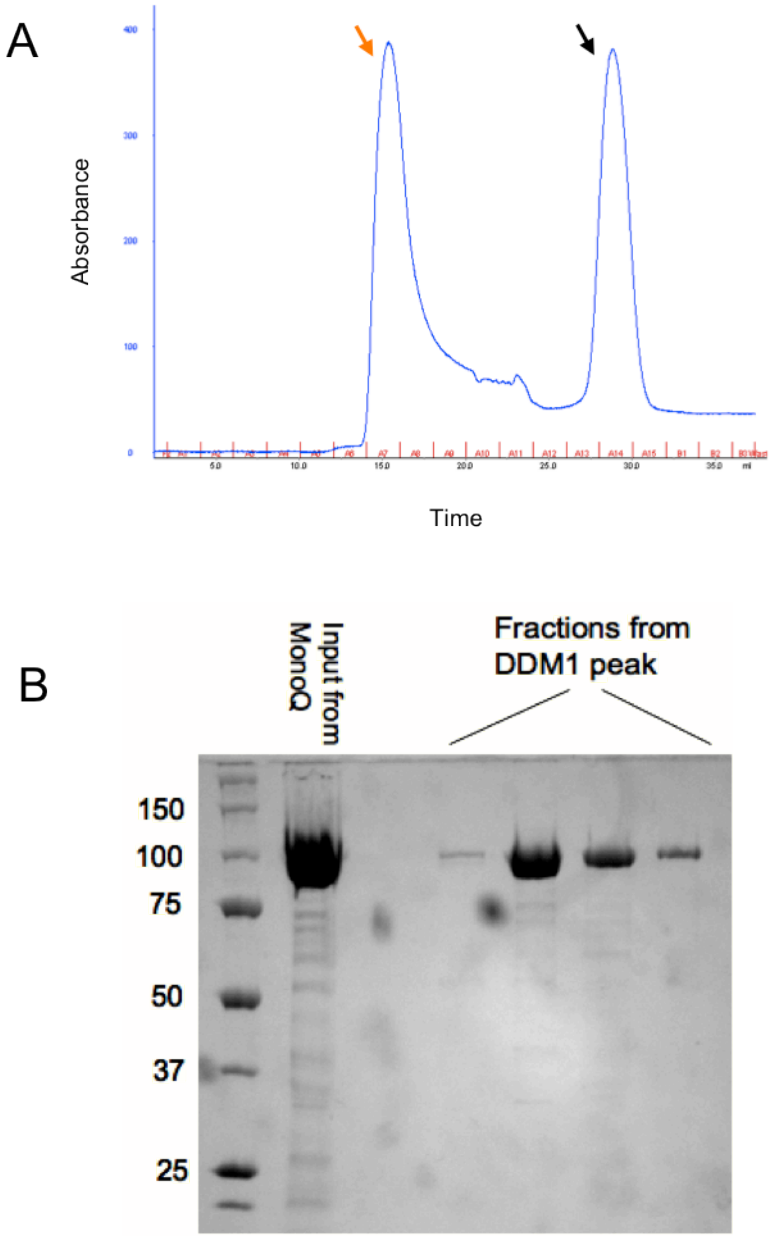


Figure 4 – Size Exclusion Chromatography. The DDM1 enriched fraction from the cation exchange column, represented on the gel in (B) is loaded on the size exclusion column, Superdex 200 10/300 GL. Part of the protein is aggregated and comes in the void fraction (orange arrow, A) while a pure DDM1 fraction that is monodispersed comes in the second peak (black arrow) annotated in (A).

Attempted Crystallization

With purified protein obtained, hanging drop crystallization screens were set up with a number of commercially available screens routinely used in the lab. A range of protein concentrations were used, specifically 5mg/ml, 7.5mg/ml and 10mg/ml. Drops were set up under a number of different conditions, with and without DNA and with and without DNA and AMP-n-P-n-P (a non hydrolysable analogue of ATP). When added, DNA was present in a 1:1.2 molar ratio as used in the published structure of the SWI/SNF domain of Rad54 bound to DNA (Dürr et al. 2006). The DDM1 input for this first set of screens contained a 6x His tag and an intact TEV cleavage site on the N terminus. Though not always an issue for most screens, the His tag does result in a non-natural charge on the terminus of the protein, which could effect crystallization. Further, intact TEV cleavage sites have the potential to interact with each other and cause the protein to precipitate; as such, I set about to cleave the tag in subsequent crystallization screens. Cleaving the tag turned out to be rather inefficient. This was especially true when attempting on column cleavage (treating with protease while the sample is immobilized on Ni-NTA beads), leaving the majority of the protein still bound to the beads. When it works well, on column cleavage is the ideal method for tag cleavage as it usually results in a much cleaner starting product in purification. In light of this I decided to swap the full-length DDM1 transcript into a vector with a thrombin site between the N-terminal His tag and transcript, instead of the TEV site present on the initial vector. The construct yielded good expression of the recombinant protein and the on column cleavage worked well, resulting in clean elutions from the Ni-NTA column. Subsequent cation exchange and gel filtration columns were run resulting in product that looked similar to those observed in figures 3 and 4. This protein was used in crystallization screens as well.

A range of both concentration of protein and DNA was used in the crystallization trials, with both the 6xHis tag containing protein and new thrombin cleaved protein. Sadly, no crystals other than ones that contained only salt came out of the trials. There are two avenues to undertake when proteins fail to crystallize, one being to uncover interacting partners that could help stabilize the protein of interest, making it more likely to crystallize. The other option is to make truncations of protein of interest, removing portions of the protein which may be unstructured under non physiological (*ie* in cell) conditions.

To this end, I performed limited proteolysis experiments combined with published secondary structure prediction software to attempt to uncover stable, well-folded portions of the protein. These experiments involve incubating a standardized amount purified protein of interest and incubating it for increasing amounts of time. The less structured, more disordered portions of the protein will be degraded more quickly, leaving well folded portions of the protein intact for longer amounts of time. The protease treated protein is then separated by gel electrophoresis and the resulting bands are excised, digested further and sent for mass spectrometry analysis. With ample coverage, one is able to determine which portion of the protein a particular fragment belonged to and hence identify which portion of the protein is stable. The results of the proteolysis experiments correlated well with the outputs from secondary structure prediction software, suggesting the N terminus of the protein, upstream of the predicted coiled coil domain are unstructured (Figure 5)

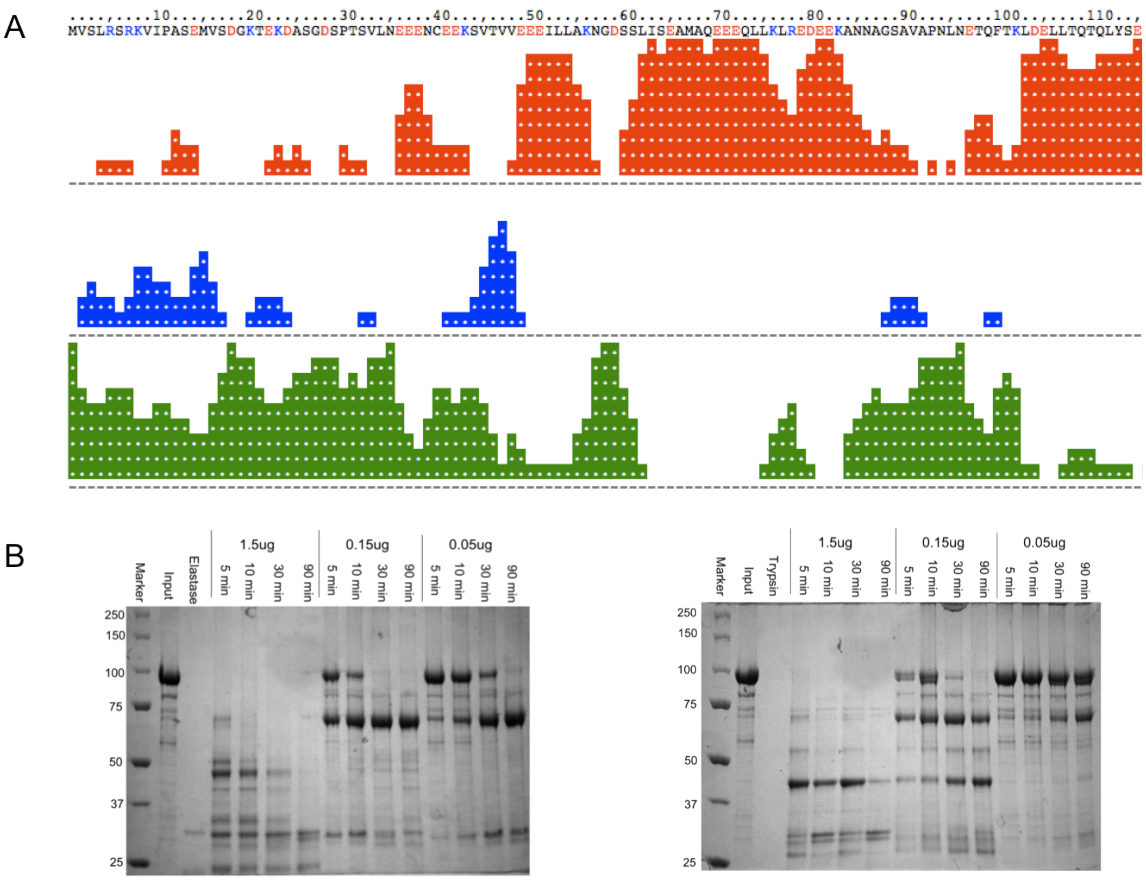


Figure 5 – Secondary Structure Prediction and Limited Proteolysis – (A) shows the N terminal sequence of DDM1 and its predicted alpha helical (red), beta sheet (blue) and unstructured (green) secondary structures. (B) illustrates the results of limited proteolysis experiments (using trypsin or elastin), with four time points of increasing incubations (5,10,30,90mins) and three separate enzyme concentrations. The starting concentration of protein was 100ug/mL.

Based on the abovementioned limited proteolysis experiments, I cloned and tested expression of a truncated form of DDM1. The construct was a truncation of the N-terminus of the protein, including the predicted coiled coil domain up towards the predicted SWI/SNF domain. The protein expressed quite well under the same expression and purification conditions used for the full length. Sadly, crystallization screens yielded no results.

Other than truncations, another avenue to pursue in a crystallization study is to attempt to co-crystallize the protein of interest with known interacting partners. Many times interacting proteins can stabilize their interacting partner and sometimes is the key to obtaining crystals. Zemach et al (2005) reported an interaction between DDM1 and MBD5, a methyl cytosine binding protein, with DDM1 being suggested to interact specifically with the MBD domain. I attempted to test this interaction *in vitro* to check the validity of these claims. Each protein was purified separately from E. coli and eluted from Ni-NTA beads via on-column cleavage. The two proteins were incubated in a 1:1 ratio with 100uM of each component for 30min and then run on a gel filtration column. Both proteins eluted off the column separately. Further experiments were performed in the presence of DNA but the reported interaction was not reproducible. This was not surprising, as the validity of this interaction has been a rather contested result in the DDM1 community, as the result has not been reproduced by other groups who have attempted it (Personal communication with Eric Richards).

Production of DDM1 antibody

As mentioned above, though much work has been done with DDM1, mechanistic insight into the specific function of the protein has proved to be elusive. This is in large part due to the gross phenotypic effects a loss of function mutant of DDM1 results in, making it very difficult to tease apart a specific function based solely on genetic experiments. Two important pieces of data are lacking, that of solid interacting partners and specific genome localization. The results of the limited proteolysis experiments uncovered stable subunits of the DDM1 protein that appear to be well structured. These observations were given more weight by the secondary structure prediction computational output. With this we felt it would be possible to create a polyclonal antibody to DDM1, which will prove to be a useful tool for future experimentation. I decided upon a region just outside of the predicted SWI/SNF domain to raise an antibody against. This region was predicted to be alpha helical and appeared in a portion of protein that was quite stable after prolonged proteolysis experiments with either trypsin or elastin. This particular region was also chosen because it is found outside of the highly conserved SWI/SNF domain, hence limiting the probability of non-specific recognition. A twenty-one amino acid sequence was submitted to an antibody production service provided by *Covance* who raised a polyclonal antibody against a synthetic peptide synthesized from that particular sequence. The antibody has been tested for specificity via western blot (Figure 7) against both purified recombinant DDM1 protein as well as protein extracted from Arabidopsis inflorescence tissue. The antibody is currently being used for both IP-mass spec and ChIP experiments, the results of which should be obtained in the coming months. In addition to the antibody, a plant expressing a GFP-DDM1 fusion under the control of the endogenous DDM1 promoter is being employed as well for ChIP-seq analysis.

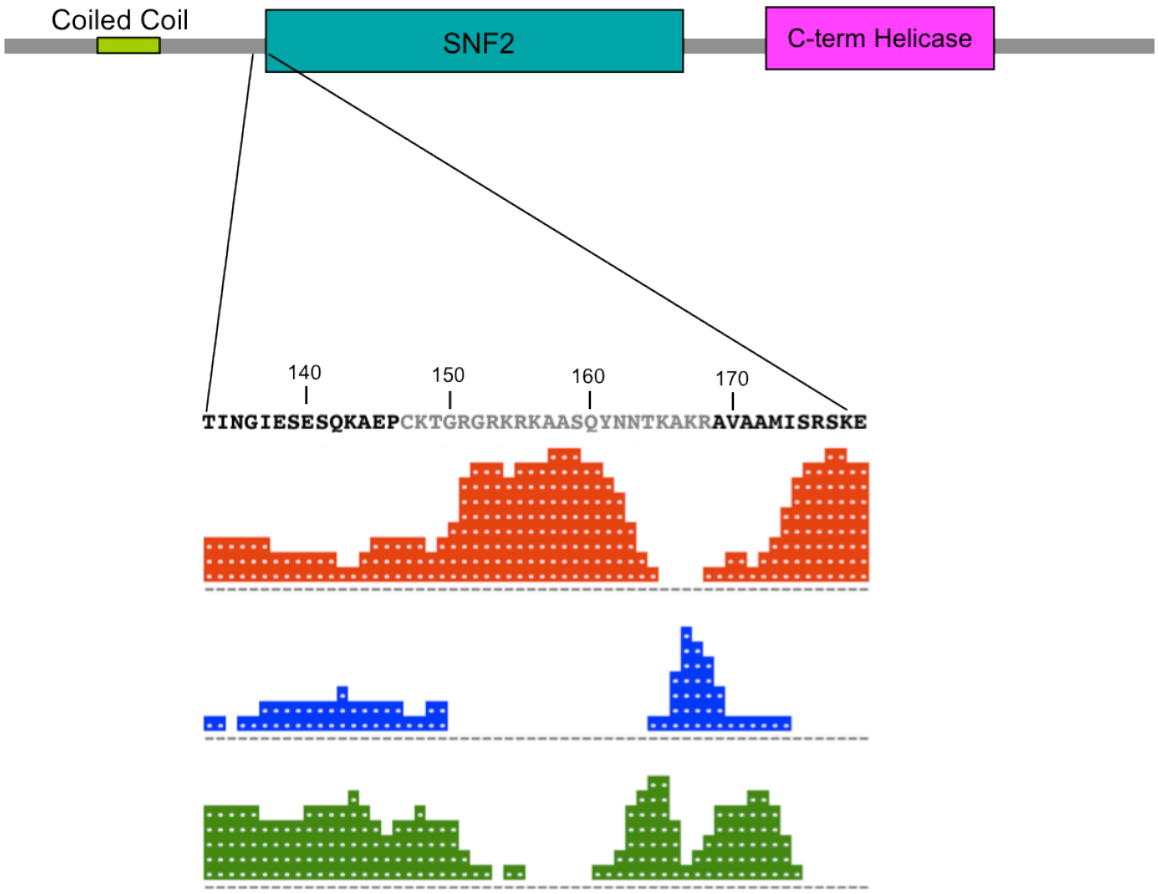


Figure 6 – Antibody design. A Schematic representation of the DDM1 primary protein structure, indicating the relative positions of the N terminal coiled coil domain, SWI/SNF domain and the C terminal Helicase Domain. The inset focuses on the specific region to which the antibody is rasied. This 21 amino acid sequence is outside of the SWI/SNF domain and has a predicted helical secondary structure.

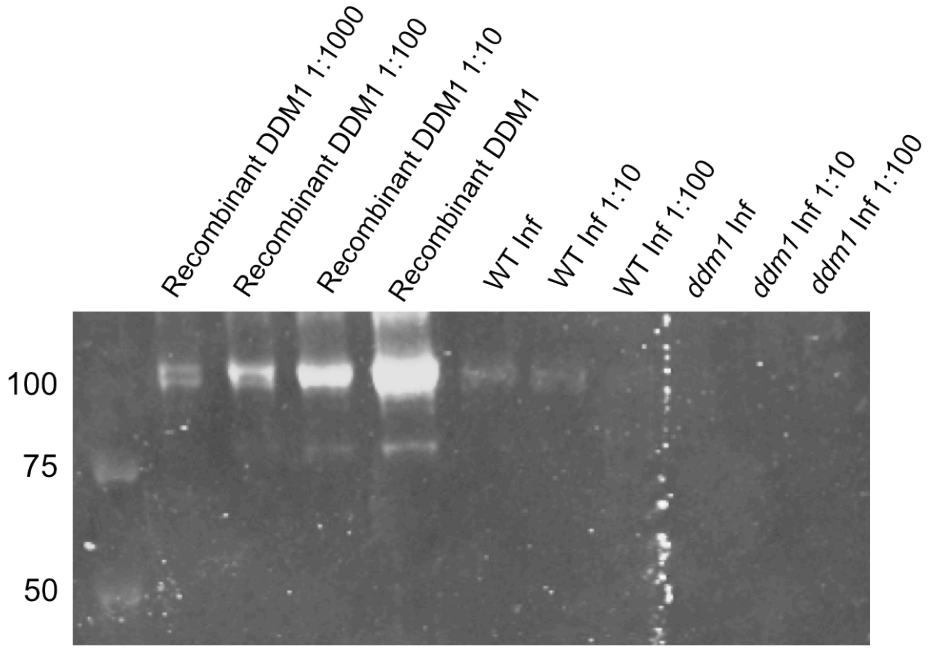


Figure 7 – Western Blot with purified antibody. A western blot visualized using chemiluminescence using our specially raised polyclonal antibody against a particular 21 amino acid sequence of DDM1. Dilutions of the recombinant protein are shown in the first 4 lanes, and three dilutions of total protein extracted from inflorescence tissue from both WT plants (lanes 5-7) and *ddm1* mutant inflorescence (lanes 8-10)

Brzeski, J., and Jerzmanowski, A. (2003). Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors. J Biol Chem *278*, 823-828.

Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X., and Meyerowitz, E. M. (2000). Ectopic hypermethylation of flower-specific genes in Arabidopsis. Curr Biol *10*, 179-186.

Jeddeloh, J. A., Stokes, T. L., and Richards, E. J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet *22*, 94-97.

Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D.*, et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471-476.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol *1*, E67.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell *130*, 851-862.

Robinson, K. M., and Schultz, M. C. (2005). Gal4-VP16 directs ATP-independent chromatin reorganization in a yeast chromatin assembly system. Biochemistry *44*, 4551-4561.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet *34*, 65-69.

Soppe, W. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., Jacobsen, S. E., Schubert, I., and Fransz, P. F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J *21*, 6549-6559.

Teixeira, F. K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M. F.*, et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Vongs, A., Kakutani, T., Martienssen, R. A., and Richards, E. J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926-1928.

Zhu, H., Geiman, T. M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. EMBO J *25*, 335-345.

Chapter 5
Future implications and ongoing work

Sections of this chapter are taken from **Calarco JP and Martienssen RA. DNA Methyltransferases Illuminate Reprogramming. (2012) *Current Biology*.

Some data is taken from Vu TM, Nakamura M, **Calarco JP**, Susaki D, Lim PQ, Kinoshita T,Higashiyama T, Martienssen RA, Berger F. RNA-directed DNA methylation controls parental genomic imprinting in Arabidopsis. (2012) *submitted*

The remainder is some of my own unpublished data, along with work from others in the lab

The importance of chromatin modification has been well established in the context of gene regulation and transposon repression. Histone tail modifications and DNA methylation are two major hallmarks of epigenetic regulation, and both have an intricate relationship with small RNA biogenesis and function (Van Ex et al., 2011). Whereas epigenetic changes in somatic cells can have important consequences on an individual, similar modifications in the germline can affect multiple individuals throughout many generations.

The epigenetic marks defined during development must be reset in the germline in order for the zygote to acquire pluripotency and subsequently to initiate embryonic development. This principle is clearly illustrated in mammals, where DNA methylation marks are reset once during primordial germ cell development, and again during early embryo development (Popp et al., 2010) (Feng et al., 2010). Comparably less is known about resetting epigenetic marks during plant development, which is what I have begun to address in this thesis. Typically, progress in understanding of many essential and basic questions tends to be driven by technical advances, which enable experiments that were previously impossible or too difficult to interpret. To this end, the protocol outlined in chapter two has served as a launching point from which advancements in the field of epigenetic inheritance can now be made. One example of this is the study presented in chapter three, where insights into DNA methylation dynamics have provided some rather surprising results, especially with regards to genomic imprinting and epiallele inheritance. The remainder of this dissertation will present further interpretation of the data presented in chapter three, framed in the context of other recently published and unpublished work in the developing embryo and endosperm, and its consequences for imprinting. Furthermore, I will outline experiments that are ongoing which utilize the system outlined in chapter two.

DNA methylation dynamics during gametogenesis and its consequences in embryogenesis

As introduced in detail in chapter one, different types of DNA methylation might reflect different aspects of transposon and gene regulation. As mentioned before, these different types of methylation can be classified based on the context of the methylated cytosine (CG, CHG and CHH methylation, where H represents any residue other than G). A recent study (Jullien et al., 2012) provides more insights into the process of DNA methylation-dependent reprogramming through embryo development. The authors combine the use of fluorescent reporters and bisulfite sequencing on a number of loci in the developing embryo to show a progressive increase in DNA methylation through embryogenesis, specifically in the CHH context, which is especially

interesting in light of our observations on CHH methylation in sperm. They are also able to show a correlation, through the use of fluorescent translational fusions, between the expression of DNA methyltransfereases and the observed methylation landscape in the embryo. As a reminder, symmetric CG DNA methylation is maintained in a replication-dependent manner, during which a hemi-methylated parent strand is used as a template to direct methylation on the daughter strand. MET1, the plant homolog of Dnmt1 performs this process. Furthermore, CMT3, a plant-specific methyltransferase, performs CHG methylation in a mechanism intertwined with lysine 9 methylation of the tail on histone H3 (Law and Jacobsen, 2010). Though these are the proteins principally involved in symmetric DNA methylation, there are some other homologous genes that participate in this process and are redundant with one another (specifically, MET2, MET3, CMT1, CMT2). This redundancy provides an additional challenge to the elucidation of the complete methylation mechanism. Additionally, plants have another type of DNA methylation, not as prevalent as the symmetric methylation described above, which is replication-independent, and relies on RNA called RNA-directed DNA methylation (RdDM). The DNA methyltransferase DRM2 (and its less ubiquitous homolog DRM1) are required for this process, as well as specific RNA polymerase (Pol IV and Pol V) subunits, NRPD1a, and NRPE1a – both of which are also involved in production and utilization of 24nt siRNA (Haag and Pikaard, 2011). Although recent mechanistic insights have greatly increased our current understanding of this type of methylation, many questions still remain.

Fluorescent DNA methyltransferase DRM1 and DRM2 reporter lines have revealed specific expression of DRM1 in the egg cell and increased expression of DRM2 in the egg cell and developing embryo. By contrast, only DRM2 is expressed in pollen (F. Berger personal communication) and its expression pattern shows a peak of expression in the VN at the bicellular stage (See Chapter 3). Insertional mutants for DRM1 show no obvious defect; however, its specific expression in the egg cell as reported by Jullien et al. (Jullien et al., 2012) suggests a particular role for DRM1 at a discrete stage in female germline development. This would be an interesting avenue for future studies in embryo development, especially in light of our observations on CHH methylation and sperm development. The authors also previously showed the absence of maintenance DNA methyltransferases in the egg cell and the absence of both maintenance and *de novo* methyltransferases in the central cell (Le et al., 2010). This finding is in accordance with the current view that DNA methylation levels are dramatically reduced in the endosperm. The situation is much different in the developing embryo however, where the expression of methyltransferases dramatically increases. This points to a very important role for RdDM in *de novo* CG, CHG and CHH methylation in the embryo.

The difficulty in isolating the different cells of the female gametophyte has resulted in an incomplete view of the methylation landscape throughout its development, especially in the egg cell. In contrast, the male germline is more readily accessible to tissue collection, and has therefore provided us with major insights regarding epigenetic marks prior to fertilization (Calarco et al., 2012) (Schoft et al., 2011) (Schoft et al., 2009) (Slotkin et al., 2009). Studies reported here indicate that both CG and CHG methylation are maintained in sperm cells, while CHH methylation is greatly reduced. These observations are in accordance with the observations in the embryo, as CG and CHG methylation appear quite steady through embryo development, while CHH methylation is progressively re-established in accordance with DRM2 expression (Figure 1). The only reported reductions in methylation during female gametogenesis and embryogensis is a loss of CG methylation observed in the central cell and subsequently in the endosperm, mediated by the repression of MET1 (Jullien et al., 2008) and the activation of DME (Hsieh et al., 2009), but not in the embryo. The results are still somewhat unclear, as methyltransferases seem to be downregulated in the egg cell even though symmetric methylation, at least in the CG context, appears to be quite high even in the early embryo. Advances in tissue isolation techniques (Deal and Henikoff, 2011) will undoubtedly address the lack of megagametophyte methylation data, and, perhaps – in combination with 'third-generation' sequencing (nanopore sequencing) technology – will provide us with new insights in the methylation landscape at the single-cell level.

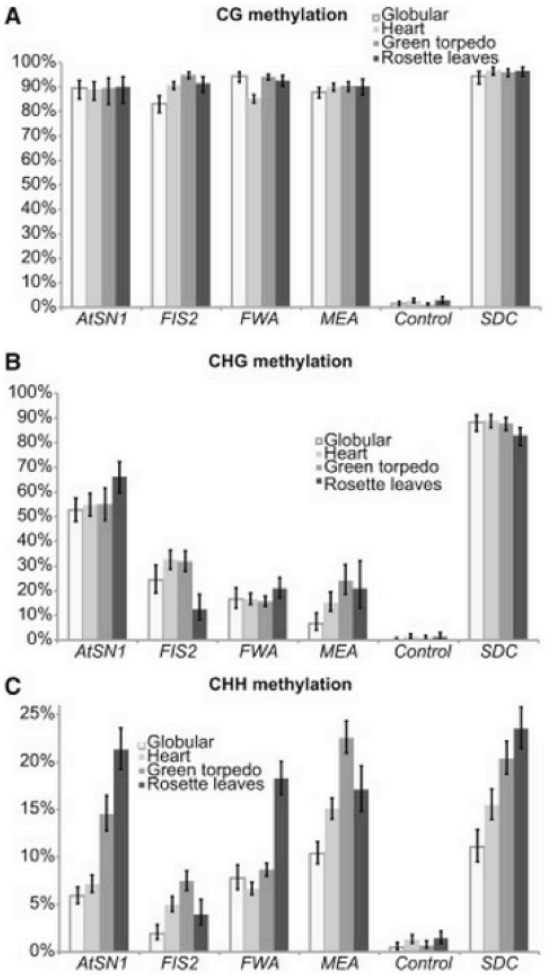


Figure 1 - DNA Methylation during Embryogenesis. Levels of DNA methylation at five loci during embryogenesis. Levels of DNA methylation in CG context (A), CHG context (B), and CHH context (C) at the methylation-sensitive loci *AtSN1*, *FIS2*, *FWA*,*MEA*, and *SDC* are compared with the unmethylated *Control* locus At2g20610. Rosette leaves provide a control for levels of DNA methylation in adult plants. *Figure modified from Jullien et al, *Current Biology 2012*

One of the most striking observations from this study in embryo (Jullien et al., 2012) is that asymmetric CHH methylation is re-established throughout embryogenesis (Figure 1) and provides the foundation for further mechanistic study into how this occurs. Analysis of Pol IV and Pol V expression profiles, as well as sequencing of 24nt small RNAs in the egg cell and embryo would provide major insight into this process. A previous study (Mosher et al., 2009) reported a very low abundance of 24nt small RNAs produced by Pol IV in the embryo, in contrast to the endosperm and seed coat, where much higher levels of these small RNAs were detected (Mosher et al., 2009). Moreover, these endosperm siRNAs, when examined in light of ecotype-specific crosses, proved to be maternal in origin. This result suggests that either these small RNAs move from the neighboring tissue types, that they were under the detection limit in the embryo at the time stage they examined, or that the siRNAs in the embryo come from the sperm cell. These observations are made all the more interesting in light of the DNA methylation and methyltransferase expression data in the embryo, and the siRNA profiles presented in Chapter 3 of this thesis (Calarco et al., 2012).

The use of translational fusions for the study of DNA methyltransferases localization proved to be very informative and should be extended to other components of the RdDM pathway. It would be particularly interesting for NRPDa/b (subunits of Pol IV) to assess in which part of the ovule the biogenesis of the 24nt small RNAs required to direct CHH methylation occurs. If biogenesis of these small RNAs occurs in the central cell or endosperm instead of, or in addition to, the embryo itself, this would imply movement of a specific class of small RNA in the female gametophyte, reminiscent of the proposed movement of small RNA in pollen (Slotkin et al., 2009) (Calarco et al., 2012). With regards to pollen, Pol IV reporter lines would also allow us to determine whether Pol IV is expressed in the sperm cells, opening up the possibility that both male and female gametophytic cells contribute 24nt small RNAs to the developing embryo.

A role in imprinting?

In both plants and mammals, mechanisms exist to implement parental genomic imprinting, which restricts expression of certain loci to one parental allele (Feil and Berger, 2007). Imprinting in mammals relies on sex-dependent *de novo* deposition of DNA methylation by specific methyltransferases Dnmt3a and Dnmt3b (Ferguson-Smith, 2011) (Bartolomei, 2009). Once established, the asymmetric patterns of DNA methylation on each parental allele are further maintained in the embryo by the maintenance DNA methyltransferase Dnmt1. In plants, the role of RNA directed *de novo* DNA methylation in imprinting remains unclear. It is known though

that MET1, the plant homolog of mammalian Dnmt1, is involved in the mechanism of parent-specific gene activation and also in maintenance of imprinted expression (Jullien et al., 2008). Furthermore, histone methylation via the Polycomb group of proteins also influences imprinting both in plants and mammals (Feil and Berger, 2007), but these two mechanims might not account for all of the monoallelic expression observed.

As outlined in detail in the introduction, during plant reproduction two sperm cells fertilize two distinct female gametes, the egg cell and the central cell, producing the embryo and the endosperm (Berger and Twell, 2011). Genome wide surveys of parental allele specific expression in Arabidopsis, have uncovered new series of candidate imprinted genes and shown that imprinting is largely confined to genes expressed only in endosperm (Gehring et al., 2011; Hsieh et al., 2011; Wolff et al., 2011), which is derived from the central cell. MET1 expression is important to maintain cytosine methylation in CG contexts (Feng et al., 2010), and its action results in the silenced status of many imprinted genes in sporophytic vegetative cells and in male gametes (Jullien and Berger, 2009). Importantly then, the modulation of MET1 expression in tissues where monoallelic expression occurs is at least partially responsible for the endosperm-specific expression of imprinted genes, activating for example the expression of the maternal allele. In the central cell the transcriptional repression of MET1 (Jullien et al., 2008) together with the active demethylation by the DNA glycosylase DEMETER (Gehring et al., 2006) cause the loss of methylated CGs, resulting in transcriptional activation in the endosperm precursor, the central cell. This may not account for all regions which show imprinting, and a precedent has been set for a role for RdDM in imprinting in *Maize* after fertilization (Jahnke and Scholten, 2009), though insights into the mechanism are lacking. This suggests that additional mechanisms, other than ones relying on DME or MET1 should exist to regulate the maintenance and establishment of an imprint. Some potential pathways are the RdDM *de novo* methylation pathway or of course silencing through polycomb. Its not known whether these will exist in tandem or only act on specific imprints and not others (Calarco et al., 2012; Gehring et al., 2011; Hsieh et al., 2011; Wolff et al., 2011). The first indications of RdDM's role in imprinting came from our work, but this evidence is only indirect. Work of collaborators has set about doing this (Vu et al *submitted*), focusing on two particular imprinted loci.

It has been established that *SUPPRESSOR OF drm1 drm2 cmt3* (*SDC*) is a maternally expressed imprinted gene. Through the use of fluorescent reporter constructs in various RdDM mutant backgrounds Vu *et al.* are able to show SDC is influenced by RdDM. *SDC* is not expressed in pollen and accordingly, our genome wide bisulfite sequencing data in sperm cells illustrate that DNA methylation is found in all contexts, including CHH in the region containing

the *SDC* promoter (Calarco et al., 2012). This observation suggested that RdDM actively represses *SDC* expression in male gametes, but the effect of RdDM on the expression of the paternal allele of *SDC* in endosperm remained unknown. Using the reporter or SNPs from different accessions, ectopic transcriptional activity of the endogenous *SDC* paternal allele and of the *SDC* reporter when paternally inherited from *nrpd2a* and *drm1,drm2* mutants was detected. These data show that the RdDM pathway is involved in silencing the *SDC* paternal allele that is inherited in endosperm.

In the end, Vu et al were able to highlight the importance of RdDM in establishing monoallelic expression (Vu et al 2013). Though this mechanism cannot be extended to all imprinted loci, it underscores how complicated the process of imprinting is and how much mechanistic work needs to be done, and highlights how important an understanding of the chromatin modification in sperm cells are. In plants, the opposition of DNA methylation patterns between male and female gametes appears to be sufficient to create a blueprint that causes imprinted expression at certain loci. Hence the maternal expression in the endosperm becomes a mere consequence of what was primed in gametes. As highlighted extensively in chapter three, close to 50% of imprinted loci are surrounded by TEs. It has also been well established that TE silencing is typically dependent on RdDM (Martienssen et al., 2008). All of this helps to define a reason why TE insertion close to a gene can effect it's monoallelic expression, as these elements can produce siRNAs which can in turn direct RdDM and silence a locus. It would be feasible to believe that a mechanism of RdDM associated to TEs direct the silencing of imprinted genes with transposons surrounding them, while imprinted loci without TEs surrounding them would be controlled by other mechanisms (Polycomb etc). The association between transposable elements (TEs) and several imprinted loci in plants and in mammals suggesting TEs playing a role in establishing imprinting expression (Wolff et al., 2011) (Gehring et al., 2011). This hypothesis was further supported by the fact that DME might have a preference for de-methylating TEs (Gehring et al., 2009), although it is now apparent that DME is also expressed in other cell types (Calarco et al., 2012a) (Schoft et al., 2011) and is dispensable for imprinting of certain loci (Jullien et al., 2008). TEs rather, are silenced by RdDM, which makes MET1 unlikely to be involved in the initial silencing event required in creating a new imprinted locus, strengthening the idea of imprinting independent of RB or DME based mechanisms. It would be interesting to see this observation extended to all other imprinted loci surrounded by TEs.

A Role in Epiallele Formation?

Even broader implications can be drawn from the important finding of DNA methylation re-establishment in the embryo with regards to the transmission of 'epialleles', which are genetically identical genes that differ only in their covalent chromatin modifications (Feng et al., 2010). Two recent studies provide additional genome-wide DNA methylation profiling data over multiple generations and uncover a number of epialleles. These studies uncovered fluctuations in the methylation landscape and expression of the epialleles based on different levels of the modification over multiple generations. The origins of this fluctuation, however, are unclear and it is proposed that the gains and losses in methylation might be pre-programmed in the germline (Schmitz et al., 2011) (Becker et al., 2011). The reduction and re-establishment of asymmetric methylation during gamete development and early embryogenesis respectively, could represent a specific pathway for the erasure or establishment of repressive marks on epialleles. These epigenetic marks could perhaps be influenced by external factors such as abiotic stress or genome clashes as a result of interspecies hybridization. This intricate response mechanism relying on RdDM could provide plants a certain flexibility to adapt to both internal and external stresses, and with the observations made from our work in male gametogenesis (Calarco et al., 2012) and recent studies in the developing embryo (Jullien et al., 2012) give further weight to this idea. These observations provide the necessary frame work to address the important questions of epigenetic reprogramming with regards the transmission of epialleles, and to see how they behave under stress conditions. This is an active area of research and the roles of abiotic stress (in this case drought) are being undertaken currently in the lab in collaboration with Dr. Frederic Van Ex.

Modulation of methylation via argonaute?

Focusing on mechanism, our studies on DNA methylation content in pollen have uncovered a dynamic change specifically in CHH methylation. This is seen most strikingly in particular transposon subfamilies, being most prevalent in LTR/retrotransposons (Calarco *et al* 2012) where CHH methylation decreases consistently to close to 0%, compared to other tissue types, which fluctuate between 15 and 20% (Cokus et al., 2008). The fact that only CHH methylation is reduced in these regions suggests that the loss of methylation is not active, as glycosolyases act in a context independent manner and should target both CG and CHG as well. Furthermore, through the use of published data examining DNA methylation in various DNA glycosolyase mutants we have defined regions that are targeted for active demethylation. None of these regions overlap with the retroelements which show loss of CHH methylation in the microspore and SN. Most likely then, the RdDM pathway would simply be modulated in such a way to limit its recruitment to these regions resulting in CHH methylation being passively lost. Importantly, other transposon classes, as well as genic regions do not show such dynamic changes in CHH methylation and instead show consistent levels all through pollen development. Altering the scale on the heatmap to $\log_{10}$ greatly accentuates this trend, illustrating that CHH methylation still present in the microspore population and the purified sperm cells (Figure 2). All of this suggests that during microspore development some regions recruit the RdDM machinery normally and CHH methylation is maintained, while the machinery is instead either excluded or specifically not recruited to the regions that lose CHH methylation in the microspore.

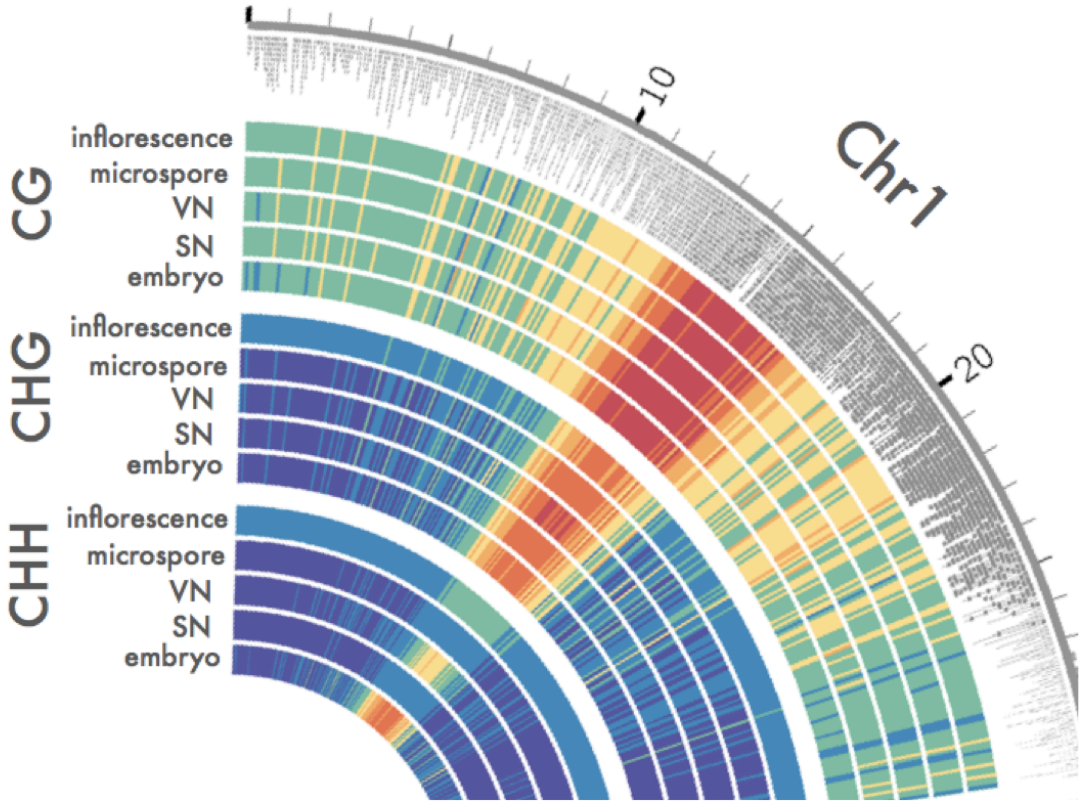


Figure 2 – CHH methylation is still retained at a number of loci. Using a heatmap representation of methylation over chromosome 1 and accentuating the scale by placing the data on a $\log_{10}$ scale, we see that a number of loci retain some CHH methylation in the microspore and sperm cell. Suggesting the methylation is not lost genome wide but modulated on at specific regions.

This suggests that regions of the genome would still retain RdDM directed DNA methylation all throughout development and others would not. Though not completely understood, many components of the RdDM pathway have been uncovered (Law and Jacobsen, 2010) (Figure 3) and one of the parts of the mechanism that could be easily modulated is the AGO component in this pathway (Figure 4). Its has been well established that argonaute genes have some of the most diversity in the Arabidopsis RNAi system, with ten different copies roughly sectioned into three different functional groups (Figure 5) (Borges et al., 2011; Vaucheret, 2008). Argonaute 4 (AGO4) was the first AGO protein to come up in a screen for RdDM genes, and is thought to be the primary AGO protein involved in RdDM, being loaded with 24nt siRNAs.

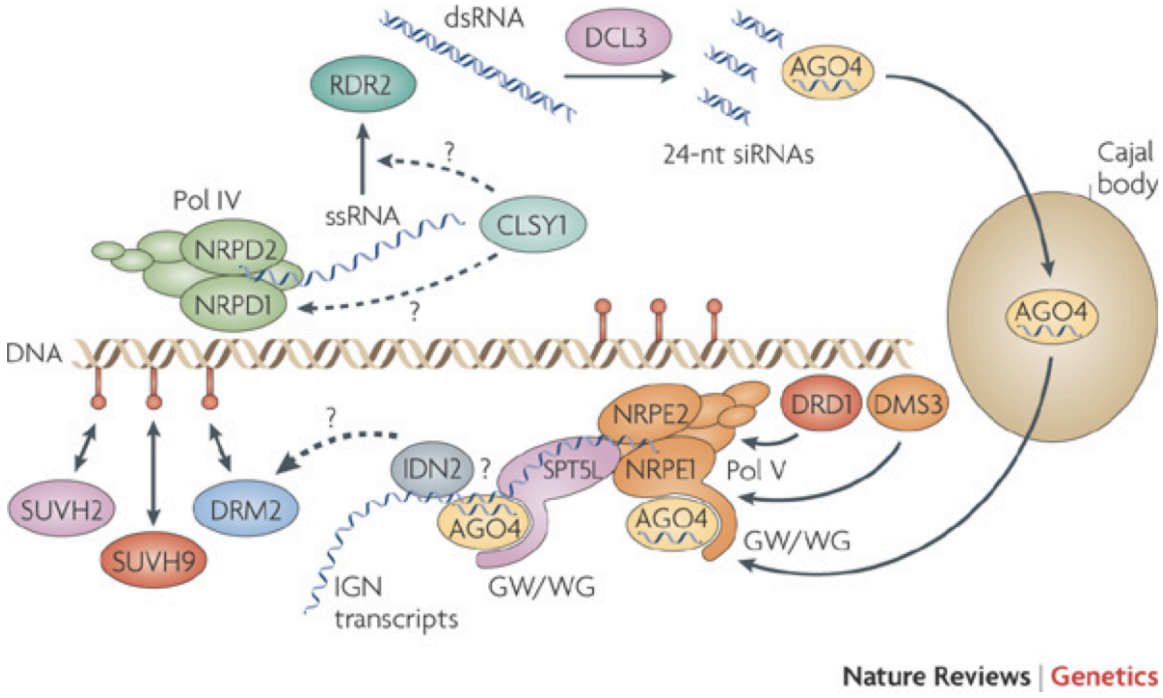


Figure 4 - Model for RNA directed DNA methylation (RdDM)
Single-stranded RNA transcripts (ssRNA) corresponding to transposons and repeat elements are hypothesized to be generated by Pol IV. RDR2, an RNA dependent RNA polymerase, is proposed to generate double-stranded RNA (dsRNA) from the ssRNA transcripts. DCL3, a dicer-like protein, is thought to process the dsRNAs into 24nt siRNAs, which are bound by an argonaute protein, AGO4. AGO4 also co-localizes with two Pol V subunits, NRPE1 and NRPE2, and DRM2 at a distinct nuclear foci and may represent a site of active RdDM. Pol V is thought to transcribe intergenic noncoding regions throughout the genome and help target the RdDM complex.
**Figure from (Law and Jacobsen, 2010)

Evolutionary analysis suggests both AGO6 and AGO9 are the most related to AGO4, and various immunoprecipitation studies have shown that both also bind 24nt siRNAs. A role for AGO9 in female germline development has already been well established (Olmedo-Monfil et al., 2010), and roles for AGO6 are starting to emerge (JP Vielle-Calzada, personal communication). To this end, we are attempting to understand the potential interplay between AGO4, AGO6 and AGO9 in controlling DNA methylation dynamics through pollen development, specifically with regards to understanding how retrotransposons are expressed in the VN (Slotkin et al., 2009). To this end, there are number of experiments currently being performed. The first is to uncover the population of 24nt siRNAs bound to AGO9, AGO6 and AGO4 in pollen. RNA IP-seq (RIP-seq) experiments have been performed in pollen with a GFP tagged version of AGO9 (F. Van Ex) while FLAG-tagged AGO6 RIP-seq data is available from total flower (Havecker et al., 2010). Though not the perfect comparison, as the AGO6 data has male germline tissue, as well as female and somatic, the trend already does suggest an interesting correlation (Figure 6). Data from the AGO9 RNA-IP shows an enrichment for 24nt siRNAs mapping to Gypsy elements, when compared to the 24nt siRNAs bound to AGO6. Just as a reminder, these are the regions that show massive loss and regain of CHH methylation in the VN. The opposite correlation is seen for RC/Helitron elements, where 24 siRNAs bound to AGO6 tend to map here to greater amounts than AGO9 bound 24nt siRNAs (Figure 6). I have constructed a line with a GFP-AGO6 under the endogenous promoter to repeat these RIP-seq experiments in pollen. This has lead to a hypothesis that in the male germline, RdDM is controlled by a combination of AGO6 and AGO9, each providing specificity for particular regions of RdDM. To further test this hypothesis I have construct genome wide bisulfite genomic libraries for *ago4, ago6* and *ago9* purified SC and VN currently being sequenced. *ago6/9* SN and VN will be analyzed next. Additionally, we are profiling TE expression in these same mutants, to understand if this is enough to modulate TE expression and siRNA production(Slotkin et al., 2009).

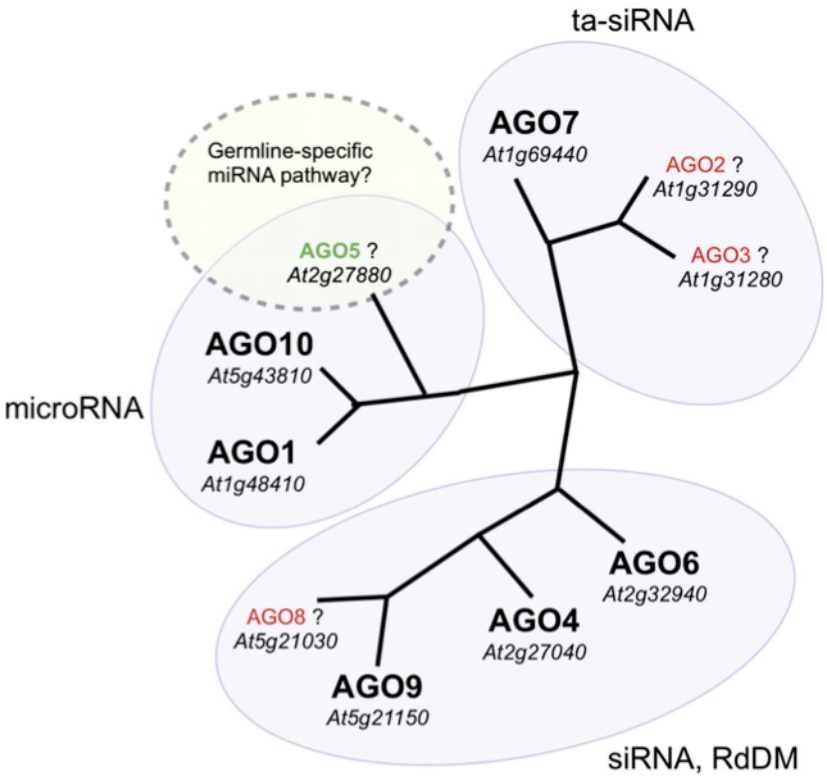


Figure 5 - Argonaute protein family in *Arabidopsis*. Phylogenetic tree illustrating the 10 Argonaute proteins inArabidopsis, subdivided into the three main functional classes based on sequence homology: Of interest here is the clade at the at the bottom of the figure which participates in chromatin remodelling via siRNA-directed DNA methylation (RdDM) and is made up of AGO4, AGO6 and AGO9 **Figure adapted from (Borges et al., 2011)

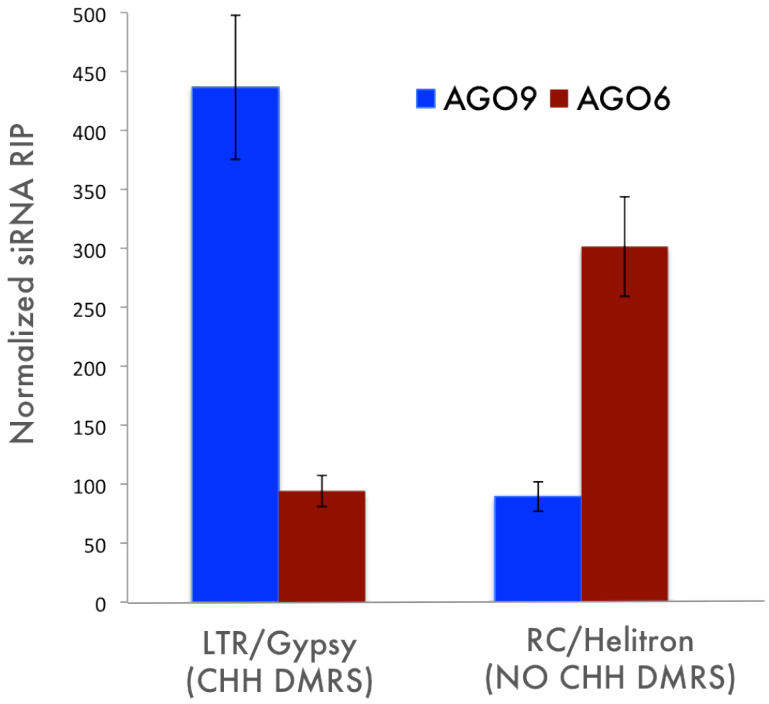


Figure 6 – siRNA populations bound to AGO9 and AGO6 at regions of differential methylation. RIP-seq data from AGO9 IPs in pollen and AGO6 in total flower over TE regions which did or did not overlap with CHH DMRs. 24nt siRNAs bound to AGO9 tended to be enriched in regions of CHH DRMs while AGO6 tends to be enriched in regions that maintained CHH methylation all throughout pollen development. Error bars represent different sub types of TEs within a specific TE class.

The surprising results of our study of DNA methylation dynamics through pollen development have led to an interest in exploring the possibility of other factors being increasingly important in mediating chromatin compaction and potential transposable element de-repression observed in pollen. Furthermore, as our system for specific cell type purification has become increasingly more efficient, more avenues of study are experimentally feasible. To this end we have decided to observe both histone variant dynamics through pollen development and dynamics of histone tail modifications and attempt to understand the potential role they may play. This work is being performed with the collaboration of Dr. Yannick Jacob, who is focused mostly on histone variants and modifications specific to sperm cells, where my interest lies mostly in the vegetative nucleus.

Chromatin Modification through Pollen Development

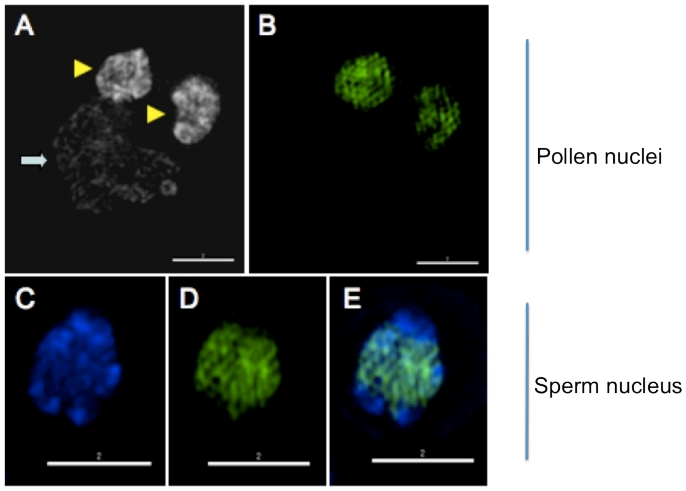
When we initiated our work on DNA methylation dynamics through pollen development we struggled with tissue collection protocols but nonetheless were able to collect enough tissue for bisulfite sequencing. Since these initial studies, a number of technical improvements in sorting has increased our efficiency and hence the scope of our work, and will facilitate the completion of a number of informative experiments. Firstly, we are focusing on Chromatin Immunoprecipitation experiments (ChIP) for a number of histone tail modifications. These would be very informative and made increasingly more powerful when taken together with genome wide bisulfite seq data. Furthermore, because we can sort simply with dye and have no need for markers, we can look at various mutants easily. Currently, we are pooling purified, fixed nuclei from sorted wild type plants, preparing for ChIP-seq experiments for H3K4 trimethylation, H3K27 di/tri methylation in sperm and VN.

Histone Variants through Pollen Development

Essential to this work is the fact that Arabidopsis contains 15 different Histone Three Related (HTR) genes that can be grouped into four classes, H3.1, H3.3, H3.3-like, and CENH3. (Ingouff and Berger, 2010) A typical somatic cell will contain 9 H3 variants while sperm nuclei only contain four: HTR5 (H3.3), HTR10 (H3.3-like), HTR12 (CENH3) and MGH3. The vegetative nucleus only contain HTR5 (H3.3), HTR8 (H3.3), and HTR14 (H3.3-like) (Ingouff et al., 2007). HTR5 is shared between both the sperm cells and vegetative cell, while HTR6 (H3.3-like) is a redundant copy of HTR14 which shows no expression in SN or VN. Immunoflorescence experiments looking for the presence of the H3K9me2 suggests its absence from the VN, but

presence in the SN (Schoft et al., 2009). This observation, combined with the surprising DNA methylation results in the VN have lead to the idea that histone variant localization might have an influence on transposon expression. With this in mind I am uncovering the position of these pollen specific histone variants in the VN of mature pollen making a high resolution map. Then, I will compare the location of these proteins and see their correlation with DNA methylation and siRNA populations.

The localization of these particular variants would be accomplished though ChIP-seq, using fusion constructs that have been previously produced and published and are present in the lab. Specifically then I am profiling HTR5, HTR10, HTR12, HTR8 and HTR14, which are fused to GFP or RFP and driven under the expression of their own promoters. To determine the feasibility of these experiments I developed a protocol focusing first on working with just one of the lines, HTR10-GFP, a variant whose localization pattern was well defined via microscopy, being excluded from the pericentomeric, transposon rich region (Figure 7). An exhaustive literature search yielded no results for a ChIP experiment from pollen, so I devised a protocol modified from a yeast ChIP and in the end was able to develop a protocol that uses ~10,000,000 pollen grains, well within the available means of growth and collection. The results correlate quite well with the observed microscopic localization.



F) Localization of MGH3 (HTR 10) via ChIP-qPCR

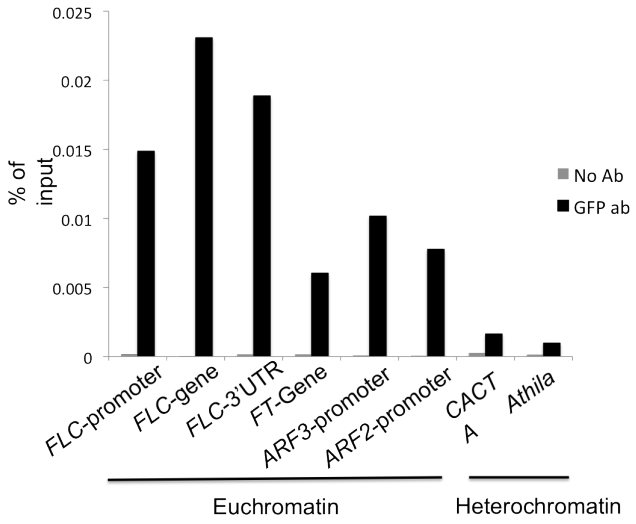


Figure 7 – MGH3 localization, a proof of principle. MGH3-GFP translational fusion under the control of its endogenous promoter shows a distinct euchromatic localization pattern in sperm cells. A) DAPI stain of pollen grain highlighting sperm nucelus (yellow arrow) and vegetative nucleus (white arrow). B shows MGH3 (HTR10) localization is specific to sperm cells, while C-E (DAPI, GPF, MERGE) suggest HTR10 is excluded from DAPI dense, heterochromatin loci. Subsequent ChIP-qPCR experiments (F) at a number of euchromatic and heterochromatic regions correlates with the observed localization patterns from the microscopy experiments. **Image prepared by Chantal LeBlanc/Zsolt Lazar, ChIP performed by Joe Calarco and Yannick Jacob

To try to understand the role of histone variants in regulating transposons in the VN we've completed several trials to attempt to uncover a representation of TE expression in *htr14* and WT pollen, with multiple replicates of RNA isolated from total pollen and multiple tested and published primer pairs. After an extensive array of qPCR experiments, the general conclusion is that *htr14* doesn't seem to effect transposon expression in pollen.

We thought HTR14 could be critical for transposon reactivation and sRNA biogenesis in pollen based on its specific localization in the VN, and also based on its amino acid sequence. We hypothesized that the substitution at position 11 of HTR14 (T11H) might interfere with H3K9 methylation on this histone, based on previous structural studies showing that T11 is critical for K9 specificity of the K9 methyltransferases. Thus, if HTR14 is deposited at heterochromatic loci specifically after mitosis I, which we are sure it is, and cannot undergo H3K9 methylation, this might lead to transposon reactivation.

To this end, we performed *in vitro* histone methyltransferase assays using recombinant Kryptonite (H3K9 Methyltransferase) on reconstituted nucleosomes containing HTR14 (Y. Jacob). Our results indicate HTR14 cannot be methylated at K9. However, when we did ChIP-qPCR on pollen we found out that HTR14 is not that abundant in the VN, and that it seems to be equally distributed in euchromatin and heterochromatin loci. It is possible that HTR14 localization is very specific, and that we are missing its peak areas of localization by picking specific loci for qPCR. We tried unsuccessfully so far to perform ChIP followed by sequencing but have so far have been limited by input DNA amounts. Experiments are currently being repeated with an increased input DNA concentration.

In tandem, we have recently made a double mutant between HTR14 and HTR6 to check for redundancy, as both these histone H3 variants contain the same T11H substitution, and they have similar promoter sequences. Upon genotyping, the expected amount of *htr6-htr14* double mutant plants were found in the F2 population, suggesting there are no transmission defects. Furthermore, these plants appear to be normal phenotypically, even with regards to VN structure in mature pollen. Once we have ample seed of the double mutant I will profile the expression pattern of various TEs we know are de-repressed in WT pollen to understand if HTR6 and HTR14 possibly act redundantly.

At this point, we are still unsure of the roles of HTR14 (and HTR6 potentially). ChIP-seq will provide more information about its role and be made increasingly more powerful once combined with our DNA methylation data.

Future directions with DDM1

The work completed in chapter four of this thesis, which resulted in the production of a functional antibody, has provided the lab with a useful tool to attempt to further our knowledge of DDM1 and its role in maintaining correct chromatin modification profiles. As outlined in chapter 4, a DDM1 antibody, which recognizes both recombinant DDM1 and DDM1 from an inflorescence lysate will be of critical importance in understanding the localization pattern of the protein molecularly via ChIP experiments. Though expected to be largely pericentromeric, some unexpected regions of localization may be uncovered. More importantly, the antibody would be useful for attempting to uncover interacting partners for DDM1 via IP-mass spec experiments.

DDM1 has a very specific expression pattern when examining different tissues of the plant. It appears to have a basal expression in all somatic tissue types while its expression peaks rather highly in sperm cells. Interestingly, it is absent in the VN of pollen (Slotkin et al., 2009). It has come to light that the DNA methylation reduction observed in *ddm1* plants can be reversed if it is combined with an *h1 (histone1)* mutation (Zemach et al, 2013), suggesting that DDM1's role is to act somehow in concert with linker histones, and a potential role in mediating higher order chromatin structure. As H1 is expressed in the VN, experiments are underway to drive the expression of DDM1 in the VN, to determine if this is sufficient to limit the expression of TEs in mature pollen. Results are pending.

Bartolomei, M. S. (2009). Genomic imprinting: employing and avoiding epigenetic processes. Genes Dev *23*, 2124-2133.

Becker, C., Hagmann, J., Muller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245-249.

Berger, F., and Twell, D. (2011). Germline specification and function in plants. Annu Rev Plant Biol *62*, 461-484.

Borges, F., Pereira, P. A., Slotkin, R. K., Martienssen, R. A., and Becker, J. D. (2011). MicroRNA activity in the Arabidopsis male germline. J Exp Bot *62*, 1611-1620.

Calarco, J. P., Borges, F., Donoghue, M. T., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijo, J. A., Becker, J. D., and Martienssen, R. A. (2012a). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194-205.

Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Deal, R. B., and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. Nat Protoc *6*, 56-68.

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

Feng, S., Jacobsen, S. E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. Science *330*, 622-627.

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet *12*, 565-575.

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

Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell *124*, 495-506.

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. PLoS One *6*, e23687.

Haag, J. R., and Pikaard, C. S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol *12*, 483-492.

Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., Schwach, F., Doonan, J. H., and Baulcombe, D. C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell *22*, 321-334.

Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., and Zilberman, D. (2009). Genome-wide demethylation of Arabidopsis endosperm. Science *324*, 1451-1454.

Hsieh, T. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., Hashimoto, M., Kirkbride, R. C., Harada, J. J., Zilberman, D., and Fischer, R. L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A *108*, 1755-1762.

Ingouff, M., and Berger, F. (2010). Histone3 variants in plants. Chromosoma *119*, 27-33.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. Curr Biol *17*, 1032-1037.

Jahnke, S., and Scholten, S. (2009). Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol *19*, 1677-1681.

Jullien, P. E., and Berger, F. (2009). Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Opin Plant Biol *12*, 637-642.

Jullien, P. E., Mosquna, A., Ingouff, M., Sakata, T., Ohad, N., and Berger, F. (2008). Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol *6*, e194.

Jullien, P. E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol *22*, 1825-1830.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Le, B. H., Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S*., et al.* (2010). Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci U S A *107*, 8063-8070.

Martienssen, R. A., Kloc, A., Slotkin, R. K., and Tanurdzic, M. (2008). Epigenetic inheritance and reprogramming in plants and fission yeast. Cold Spring Harb Symp Quant Biol *73*, 265-271.

Mosher, R. A., Melnyk, C. W., Kelly, K. A., Dunn, R. M., Studholme, D. J., and Baulcombe, D. C. (2009). Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature *460*, 283-286.

Olmedo-Monfil, V., Duran-Figueroa, N., Arteaga-Vazquez, M., Demesa-Arevalo, E., Autran, D., Grimanelli, D., Slotkin, R. K., Martienssen, R. A., and Vielle-Calzada, J. P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature *464*, 628-632.

Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., and Ecker, J. R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science *334*, 369-373.

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J. S., Park, G. T.*, et al.* (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A *108*, 8042-8047.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J. D., Feijo, J. A., and Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461-472.

Van Ex, F., Jacob, Y., and Martienssen, R. A. (2011). Multiple roles for small RNAs during plant reproduction. Curr Opin Plant Biol *14*, 588-593.

Vaucheret, H. (2008). Plant ARGONAUTES. Trends Plant Sci *13*, 350-358.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Kohler, C. (2011). High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. PLoS Genet *7*, e1002126.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. (2013) The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell 153(1):193-205

Acknowledgements

To Dr. Rob Martienssen, whose creativity and constant optimism in work is infectious. As an advisor he allowed me the freedom to explore the questions I was interested in, and supported me when things were not proceeding as planned. I hope to always be able to rely on him as an advisor, peer and friend.

To the talented scientists who I had the pleasure of working directly with. Dr. Filipe Borges, first during his time at the IGC and now recently here at CSHL, the modesty with which you work has always amazed me. Working with you has made me a better scientist. Dr. Mark Donoghue, for your patience in teaching and dedication in working I am forever indebted to you. Dr. Frederic Van Ex, your cleverness and attitude have always been an inspiration for me. Our projects together have always been some of my favourite. To Dr. Yannick Jacob and Chantal LeBlanc, your dedication and work ethic has always amazed me. I thank you sincerely for the things you have taught me. To all of the above mentioned people, who are co-workers and most importantly friends. I wish you nothing but the best.

To Dr. Jose Feijo and Dr. Jorg Becker. For a successful collaboration. For hosting me in Lisbon. For constant fruitful discussions and advice. For helping to teach me to love science. Your influence will not be forgotten.

To all the members of the Martienssen lab, both past and present for creating a vibrant, creative, collaborative and inspiring environment in which to do science. To the former members, Keith Slotkin, Dani Irvine, Mikel Zaratiegui, Milos Tanurzic who were kind and wise, always quick in proving friendship and advice. To Stephane Castel, Evan Ernst, Andrea Schorn, *et al*, current members who are just as wonderful.

To Leemor Joshua-Tor, as a former co-advisor and Dean. For proving me with trust and mentorship all throughout my time in her group. To members of Leemor's lab, specifically Drs. Rajesh Kumar, Chris Faehnle, Thomas Schalch and Claus Kuhn for their advice during my time there, and continued friendship afterwards.

Last and certainly not least, to my family. My parents Joe and Joanne, for providing me all the opportunities in the world and for teaching me the value of hard work. Your love and support has made me a better person. Any of the meager success I've had in life is all thanks to you. To my brother John, for always making me strive to be better, both in life and science. Though your success has nothing to do with me, I am always proud of you.