DNA methylation reprogramming during sexual reproduction in land plants

James Walker

September 2019

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

Abstract

"DNA methylation is just decoration the cell uses to make its genome look pretty."

James Walker, 2016

Chromatin modifications are important for transcriptional regulation and development in eukaryotes. One such modification, DNA methylation, is extensively reprogrammed to regulate mammalian development. However, we are only beginning to understand the changes in methylation during the development of plants and the importance of those changes for cell function.

This thesis focuses on when, how and why dynamic DNA methylation appears during the sexual reproduction of land plants. Firstly, a specific gene-targeted methylation profile observed in the male sexual lineage of *Arabidopsis thaliana* is described. It is established that this methylation is caused by the RNA-directed DNA methylation (RdDM) pathway and that this methylation has implications for gene regulation and development. Secondly, a detailed mechanistic understanding of the specific methylation is explored to establish that gene-targeted methylation in *Arabidopsis* is likely due to small RNA produced at canonical RdDM targets. Finally, the methylation profiles during sexual development in the basal land plant, *Marchantia polymorpha*, are considered. A similar gene-targeted methylation profile is observed in the meiotic phase of development of this plant, separated from *A. thaliana* by 450 million years, while a unique and distinctive global methylation pattern occurs during the late stages of sperm maturation.

Together, these works provide a fundamental understanding of DNA methylation dynamics in plant sexual development.

Contents

Abstract 2
Contents 3
Published work and notices 5
Acknowledgements 6
Chapter One Introduction 9
1.1 Chromatin 9
1.2 Transposons 10
1.3 RNA-directed DNA methylation 11
1.4 DNA demethylation 13
1.5 Roles of chromatin modifications in sexual reproduction 14
1.6 Suppression of transposons in sexual reproduction 15
1.7 Small RNA in sexual reproduction 15
1.8 *Arabidopsis thaliana* and *Marchantia polymorpha* 16
1.9 Focus of this study 17
Chapter Two Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis thaliana* 19
2.1 Abstract 19
2.2 Introduction 20
2.3 The germline of *A. thaliana* features high CG and low CHH methylation compared to other tissues 21
2.4 The male sexual lineage exhibits specific regions of hyper and *de novo* RdDM activity 23
2.5 SLHs lose methylation through both active and passive processes 27
2.6 In soma, residual CG methylation at SLMs is maintained by MET1 28
2.7 RdDM-induced sexual-lineage-specific methylation regulates gene 30
expression in meiocytes 30
2.8 An SLM regulates the splicing of MPS1 and is important for meiosis 34
2.9 Discussion 36
2.10 Materials and Methods 39
2.11 Supplementary Figures 42
Chapter Three Mechanisms underlying sexual-lineage-specific methylation in *Arabidopsis thaliana* male meiocytes 49
3.1 Abstract 49
3.2 Introduction 50
3.3 Meiotic sRNAs are associated with a specific subset of RdDM loci 51
3.4 Specific CLSY expression is likely the cause of meiotic RdDM 54
3.5 Meiotic RdDM is associated with helitrons but not gypsy elements 57

3.6 SLM sRNAs originate from meiotic RdDM loci 59
3.7 Discussion 61
3.8 Materials and Methods 65
3.9 Supplementary Figures 67
Chapter Four Dynamic DNA methylation reprogramming during sexual development in the basal land plant *Marchantia polymorpha* 69
4.1 Abstract 69
4.2 Introduction 69
4.3 The genome of M. polymorpha is heavily methylated in sperm 75
4.4 Transposons in sperm have distinguishable CAH methylation 79
4.5 Gene-associated methylation exists in sperm and sporophytes 83
4.6 Low-methylated regions are associated with genes expressed during sperm development 87
4.7 Methylation patterns are established during sperm development and are associated with the expression of specific methylases 89
4.8 Discussion 93
4.9 Materials and Methods 97
4.10 Supplementary Figures 100
Chapter Five Main Discussion 106
5.1 Importance of SLMs associated with tRNA genes 107
5.2 Importance of de novo methylation in meiosis 107
5.3 A mechanistic understanding of sexual-lineage-specific methylation 108
5.4 Evolution of SLMs 108
5.5 Female sexual-lineage-specific methylation 109
5.6 Concluding remarks 110
List of abbreviations 111
References 114

Published work and notices

This thesis includes research material from the following published work:

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D. and Feng, X., 2018. Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. Nature genetics, 50(1), p.130.

All Supplementary Tables are provided on a CD attached to the end of this thesis but are also available at: https://github.com/jimmy-303/JW_Sept_2019 at the time of submission.

Acknowledgements

The best part of carrying out this PhD has not been the scientific ideas and discoveries (although they have been a highlight!) but the immeasurable privilege to meet such fantastic people along the way, from my placement with Yihui Cui, Tetsuya Hisanaga and the others in the Nakajima lab in Japan, to the conferences I've attended with Xiaotian Zhang and Jonathan Cocker, the latter of whom took a trip to Japan to say hello despite working in the same building in Norwich! Without this PhD I wouldn't have had the opportunity to experience such diversity and international wonder and I am hugely thankful for this opportunity. A shout out to Aaron Ang, Viktoria Hristova, Tom Casey, Lira Palmer, Aileen Magilin, and Dulce Gomez-Hernandez who have passed through the lab and expanded horizons!

I'd like to thank my family, who've been my reminders of what really matters in life. From helping me to buy a washing machine or toaster to Sunday roast dinners and folk music festivals, I will always appreciate the love from you all! This appreciation extends to all my friends mentioned here or otherwise, including Zoë Demetriou, Cilla Hughes (and her awesome son Neo), Nikki Wade, Jade Doughty, Erin Cullen, Zoe Stevens, Laura Connolly, David Lloyd-Williams, James Wilkinson, Adam Mcleod, Ali Sutherland, Stephen Futers, Ola Abayomi, Jess Eastwell, and of course Mr. Bear.

I'd like to thank my mates in the rock band NR4 and from bridge on Thursdays at the Alexandra pub for providing me with much-loved hobbies. You all showed impressive amounts of patience with my slow ability to learn songs off by heart or inability to count points while trying to navigate my PhD and you provided me with a way to switch off from work among friends. It's been so much fun! I'd especially like to thank Roy Dunford, lab manager of the Biffen building, whose charm and kindness has been a notable influence on me these past four years, who invited me to take part in these activities and somehow convinced me to be the lead singer of a rock band!

I'd like to thank the Gatsby network, who provided me with the confidence to be the scientist I am today. The support and training with such people are irreplaceable and the people I've met and interacted with during the network and training meetings have provided me with the reassurance and advice I didn't know I needed every step of the way. I'd particularly like to thank Chiara Perico, my Italian partner in crime as the other Gatsby funded student in my year. I'm so glad that we clicked

the moment we met, and I so look forward to seeing what you do in the future! Being a member of this organisation has been a source of pride for me, and I hope that it inspires me to continue helping the career of others – an endeavour that I've discovered is the most rewarding aspect of science!

I'd like to give a massive thank you to the people who have supported me in the last four years at the John Innes Centre. To Beccy Corkhill and Liz Chapman, who have been there through the PhD system with me and helped me so much I cannot put into words! To Tom Bradley, my small RNA buddy who's always great to chat to (even if he does work on animal miRNAs). To Billy Aldridge, Toby Buttress, and Sam Deans, my fellow Feng lab PhD students who have been great fun with the "lad" banter and horseplay. I may not understand it half the time, but it's always been a good laugh! To Martin Vickers, whose patience as my bioinformatics teacher (and colleague who deals with my messy desk being next to his) has been astounding – such a great mate with stories of his many jobs in life! To Hongbo Gao, who worked with me in the first few years with calmness and assuredness that took me through the process of publishing a paper, and whose humming in the lab was always a welcome sound. To Jincheng Long, Wenjing She, and Yalin Liu, postdocs who've put up with my chaotic nature and been such great friends in the office. I'd like to say a special thank you to Jingyi Zhang and Shengbo He, the postdocs who have supported me throughout this whole experience. We joke that you're my adoptive parents and honestly, you've made me feel so welcome and supported every single week here. From the playful teasing, strong advice, copious hugs, and such understanding, I'm so glad that your lives crossed mine. Your daughter Yiyi is the luckiest daughter in the world!

Finally, I'd like to give a massive thank you and always a hug to my supervisor Xiaoqi Feng. The first reason I chose to join this lab for my PhD was because I wanted to expand my research experience of transcriptional regulation on a genomic scale. Xiao has made that possible and more. She has taught me my strengths in puzzle solving and analysis and given me the permission to explore all avenues of science that I've wanted (even when it perhaps wasn't warranted!) and provided me with the tools needed to do so. The other reason for choosing this lab was the obvious care and joy that Xiao has. PhDs are tough, but the people around you can make it easier. I was worried that I hadn't met the new recruits to Xiao's lab before starting my PhD and felt a slight unease whenever a new colleague joined. I shouldn't have been so worried, as it's clear now that a lab's characteristics reflect a

supervisor’s personality. I am immensely proud to have done my PhD in this group, and so glad to be part of the family that Xiao has created!

Feng Lab, Summer 2019.

Chapter One
Introduction

"I'm freezing my sperm for future use!" – James Walker, 2016

1.1 Chromatin

The selective use of genetic information through transcriptional regulation is critical for cell function and differentiation. Most prokaryotic genomes primarily rely on a variety of transcription factors to regulate gene expression (Talbert et al., 2019). Eukaryotic genomes, on the other hand, have additional proteins known as histones (typically H2A, H2B, H3, and H4) that form an octamer by which DNA is wrapped around twice for a length of about 146 base pairs (bp) to form a nucleosome. Together with linker histones that sit between nucleosomes, this structure can affect the accessibility of DNA for transcription factors. By altering the strength of the association between histones and DNA or through nucleosome labels that are recognised by chromatin remodellers and/or transcription factors, regions of the genome can become more open (euchromatic) or closed (heterochromatic). In *A. thaliana* as with numerous other organisms, heterochromatin occurs primarily at the ends of chromosomes (i.e. the telomeres) and at the centre of chromosomes at centromeric and pericentromeric regions (Vanrobays et al., 2018). In other species such as maize and wheat, heterochromatin is also widespread along the chromosome arms and euchromatin is only seen at interspersed genes (Vanrobays et al., 2018).

One mechanism used to affect chromatin accessibility is to use histone variants that have different chemical properties, such as the intensively studied H2A.Z. This histone variant dynamically occupies nucleosomes in various regions across a genome and is thought to play numerous roles in chromatin structure (Henikoff and Smith, 2015; Talbert and Henikoff, 2010). A second mechanism to influence chromatin compaction and transcription is through chemical modifications of the histones themselves. A suite of modifications is associated with either repression, such as H3K9 trimethylation (H3K9 dimethylation in plants) and H3K27 trimethylation, or associated with activation, such as H3K4 trimethylation and H3K36 trimethylation (Feng and Jacobsen, 2011; Lawrence et al., 2016). In plants, H3K9 can be dimethylated by the action of SU(var)3-9 homologues SUVH4/KRYPTONITE (KYP), SUVH5 and SUVH6 (Bernatavichute et al., 2008; Ebbs and Bender, 2006; Jackson et al., 2002). In both plants and mammals, H3K9 methylation is particularly

associated with transposons and is important for transposon repression (Bulut-Karslioglu et al., 2014; Ding et al., 2007).

1.2 Transposons

Transposons are parasitic genetic elements which reside within host genomes and can copy themselves into new regions of DNA. Class I transposons transpose via an RNA intermediate (also known as retrotransposons), while class II transposons copy via an DNA intermediate (and hence are known as DNA transposons). A major family of class I transposons in plants are the long-terminal-repeat (LTR) retrotransposons, making up more than 75% of the maize genome (Baucom et al., 2009). These elements are characterised by the presence of extensive repeats in their flanking regions. Remarkably, it has been shown that transfer RNAs (tRNAs; i.e. the molecules used for translation in the host cell) are utilised by LTR retrotransposons as primers for reverse transcription as part of their transposition mechanism (Mak and Kleiman, 1997). This provides a clear advantage to ensure survival of these transposons within a host population as it is practically impossible to evolutionarily remove tRNAs. However, evidence suggests that host organisms instead cut a portion of tRNAs into tRNA fragments (tRFs), which act as competitors for intact tRNAs at the LTR retrotransposon primer binding sites and target LTR retrotransposons for repression at both the transcriptional and post-transcriptional level (Schorn et al., 2017).

A major family of DNA transposons are the helitrons, which are thought to replicate via a rolling circle mechanism (Kapitonov and Jurka, 2007). Interestingly, helitrons often contain fragments of other genomic regions - notably fragments of genes (Barbaglia et al., 2012; Hollister and Gaut, 2007). Several mechanisms have been proposed to explain this phenomenon such as synthesis-dependent strand-annealing (SDSA), whereby a double strand break occurring at a helitron during transposition leads to invasion of the broken DNA into another region of DNA (likely open DNA such as transcribing genes). The invaded DNA is subsequently replicated during DNA repair so that it is copied to the original helitron (Kapitonov and Jurka, 2007).

Due to their ability to cause harmful or lethal consequences for the host organism through mutation and subsequent loss of gene function, a myriad of mechanisms has evolved whereby a cell represses transposons such as through H3K9 methylation and the production of tRNA fragments. Methylation of the fifth carbon of cytosine (5mC, referred to simply as DNA methylation throughout most of this thesis) is also typically associated with transposons and transposon repression

(Deniz et al., 2019). This methylation is known to act in a self-reinforcing loop with H3K9 methylation so that the presence of each is typically dependent on the other (Stroud et al., 2014; Stroud et al., 2013). While repression of transposons is the primary purpose of chromatin modifications in these regions, the host organism can sometimes take advantage of such chromatin environments. For example, abiotic stresses such as high temperature are known to release repression of transposons, and such a release is also applied to neighbouring genes that are thought to be useful for stress responses (Joly-Lopez and Bureau, 2014). However, distinguishing the uses of transposons for host function from the need to repress transposons is often difficult.

The methylation patterns associated with transposons within a host genome are typically analysed using ends analysis (a form of metaplot; for example, Figure 2.1c; Zilberman et al., 2007), whereby transposons are aligned at the 5' and 3' ends, respectively, and the average methylation across these elements are plotted on a graph. The shoulders of these plots typically represent the methylation status for the edges of long transposons and short transposons combined. Both are more euchromatic and are targeted for repression by a pathway in plants known as RNA-directed DNA methylation (RdDM; Matzke and Mosher, 2014). The centre of ends analysis graphs represents the bodies of longer transposons. These regions are more heterochromatic and are targeted for methylation in plants by chromomethylases which recognise H3K9 methylation via their chromo and BAH domains (Stroud et al., 2014; Stroud et al., 2013).

1.3 RNA-directed DNA methylation

While RdDM is critical for the repression of transposons and mutants for RdDM in various plant species have pleiotropic defects, RdDM mutants in *A. thaliana* remarkably show no obvious morphological alterations (Matzke and Mosher, 2014). As a result, the underlying mechanisms of RdDM have been extensively studied within this well-known model organism. RdDM is carried out by DOMAINS REARRANGED METHYLTRANSFERASEs (DRMs; Figure 1.1), DNA methyltransferases ubiquitously found throughout land plants (Schmitz et al., 2019). DRMs methylate DNA in the contexts of CG, CHG and CHH (where H is A, C or T). RNA transcripts are produced by RNA polymerase IV (POLIV), processed into double stranded RNA by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) and then into sRNA by Dicer proteins (DCL3; Figure 1.1). These small RNAs (sRNAs) are then bound to Argonaute (AGO) proteins and direct DRMs to methylate DNA by binding to complementary scaffold RNA produced by RNA polymerase V (POLV;

Figure 1.1). Upon establishing methylation by RdDM, DNA METHYLTRANSFERASE 1 (MET1) has been observed to maintain CG methylation in all studied plants when recruited by VARIANT IN METHYLATION (VIM) proteins (Woo et al., 2008). In flowering plants, CHROMOMETHYLASE 3 (CMT3) maintains CHG methylation while a combination of CMT2 and RdDM is activity is used in flowering plants to maintain levels of CHH methylation (Figure 1.2).

RdDM is a self-reinforcing pathway as DNA methylation promotes the recruitment of POLIV by chromatin remodellers such as CLASSY and the homeodomain protein SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1), the latter of which also binds to H3K9 dimethylation (Figure 1.1; Johnson et al., 2014; Law et al., 2013; Liu et al., 2014; Zhou et al., 2018). Variations of core RdDM components exist (such as RDR6, DCL4, and specific AGO proteins) and these are thought to be important for functions such as the establishment of DNA methylation at active transposons rather than DNA methylation maintenance (Nuthikattu et al., 2013; Wu et al., 2012).

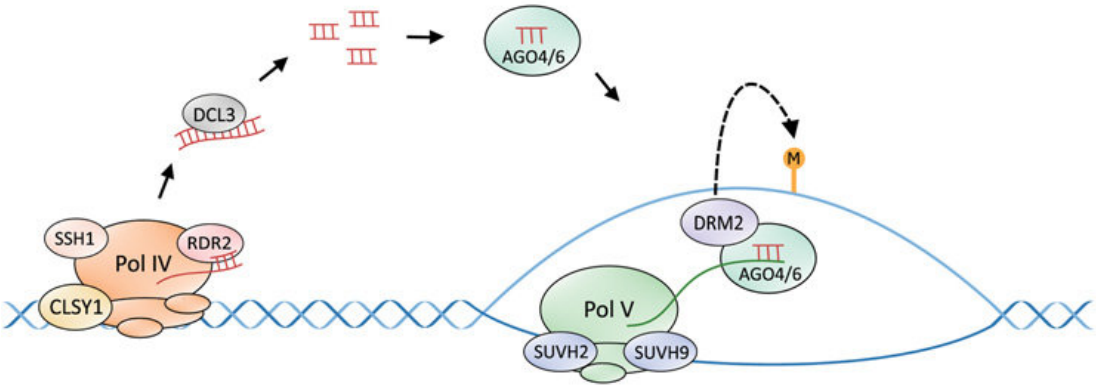


Figure 1.1 Simplified canonical RNA-directed DNA methylation pathway. CLSY proteins (typically CLSY1) and SSH1 are required for POL IV transcription of DNA. RDR2 produces double stranded RNA which is processed into small RNA molecules (typically 24 nt in length). AGO proteins (here AGO4 and AGO6) associate with the small RNA and direct DRM (typically DRM2 in *A. thaliana* while DRM1 function is unclear) to RNA scaffolds produced by POL V (the latter requiring SUVH proteins that bind methylated DNA). Taken from Kenchanmane Raju et al., 2019.

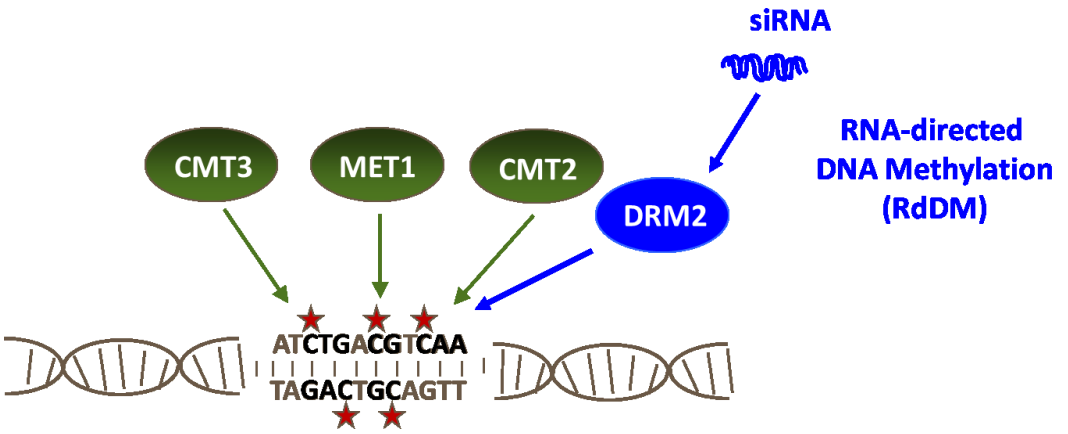


Figure 1.2. Methyltransferases in plants. MET1, CMT2, and CMT3 are responsible for CG, CHG, and CHH methylation, respectively (H = A, C or T, methylation depicted by red stars). DRM2 directs methylation in all three contexts via small interfering RNA (siRNA) via the RdDM pathway.

1.4 DNA demethylation

Loss of DNA methylation is known as DNA demethylation and has been observed in both plants and animals. DNA demethylation can be passive, such as with the loss of CHH methylation in the absence of RdDM in sperm cells (Calarco et al., 2012) or the loss of CG methylation when MET1 is downregulated in central cells (Jullien et al., 2008). Alternatively, methylation can be actively removed. In animals, this is achieved through the chemical modification of 5mC by TET enzymes into 5-hydroxymethylcytosine and other derivatives (Deniz et al., 2019). Plants lack TET enzymes and instead utilise DNA glycosylases that remove the methylated cytosine. An unmodified cytosine is subsequently inserted by DNA repair mechanisms (Wu and Zhang, 2014). In *A. thaliana*, the DNA glycosylases are known as DEMETER (DME), DEMETER-LIKE 1/REPRESSOR OF SILENCING 1 (DML1/ROS1), DML2, and DML3. While the role of DME is discussed below, it has been shown that ROS1, DML2, and DML3 stop transposon methylation (and hence repression) from spreading to neighbouring regions where gene expression could be affected (Zhang et al., 2007).

1.5 Roles of chromatin modifications in sexual reproduction

Although DNA methylation is typically associated with genome protection through the repression of transposons (Suzuki and Bird, 2008), cell-type specific methylation is employed by mammalian systems to direct transcriptional regulation and differentiation (Reik et al., 2001). Genome-wide chromatin reprogramming through DNA demethylation also occurs in mammalian primordial germ cells and in the early embryo following fertilisation through TET activity (Hemberger et al., 2009; Sasaki and Matsui, 2008; Surani et al., 2007). Both demethylation events are important for resetting a cell's ability to differentiate into other cell types and are also important for imprinting, a phenomenon whereby loci are differentially regulated in a parent-of-origin-specific manner (Reik et al., 2001). Global demethylation because of DME activity and MET1 repression also occurs in flowering plants in the vegetative cell and central cell, the companion cells to the sperm and egg, respectively (Calarco et al., 2012; Ibarra et al., 2012; Park et al., 2016). The endosperm formed after fertilisation of the central cell by one sperm cell is also demethylated. While evidence suggests that demethylation in these companion cells is important for reinforcing transposon silencing in sperm and eggs that carry genetic information to the next generation (Calarco et al., 2012; Ibarra et al., 2012; Park et al., 2016), demethylation in the endosperm is also involved in imprinting in a similar manner to the placenta of mammals (Rodrigues and Zilberman, 2015).

The competitive environment of sexual reproduction drives rapid evolution of traits that provide advantages particularly for fertilisation (regarding male sex cells this is often referred to as sperm competition; Gage, 2012). For example, sexual pressure put on beetle populations results in more competitive sperm with longer tails (Godwin et al., 2017). Thus, certain aspects of sexual biology have been found to be only recently derived within specific organisms. Chromatin compaction is often critical to allow swimming sperm to reach the egg with speed, as well as to package the DNA for safe delivery (Miller et al., 2010). In animals, chromatin compaction involves the radical replacement of most canonical histones with specific histone variants, and these in turn are replaced with transition proteins and finally protamines, the latter of which are arginine-rich proteins whose strong positive charge are important for the high condensation of DNA (Miller et al., 2010). Such compaction (which also requires a multitude of DNA strand breaks to allow topoisomerases to alter DNA structure into toroids with the protamines; Laberge and Boissonneault, 2005) results in transcriptionally inactive sperm cells (Casas and Vavouri, 2014). In flowering plants, sperm are delivered to the egg and central cell

via a pollen tube through the stigma and down the style. While compaction is still important, transcriptional activity is still present (Honys and Twell, 2003; Honys and Twell, 2004). Flowering plants lack protamines but do utilise histone variants that whose functions are still being determined (Borg and Berger, 2015). Basal land plants, on the other hand, have been found to harbour protamine-like genes that are expressed during sperm production (Higo et al., 2016).

1.6 Suppression of transposons in sexual reproduction

Transposon repression mechanisms are particularly notable during sexual reproduction as this is a critical period in which transposons can propagate throughout a host population (Bestor, 1999). Indeed, it has been asserted that the presence of transposons is closely intertwined with sex (Zemach and Zilberman, 2010). One example of transposon repression during sexual reproduction includes the demethylation in companion cells, allowing transcription of transposons to produce sRNAs that are transported into gametes for reinforced repression of the same transposons in the gametic genome (Calarco et al., 2012; Ibarra et al., 2012). tRFs are also notable in sexual reproduction and are reported to repress retrotransposons in pollen (Martinez, 2017). A specialised DNA methylase in mice has been found to be specifically important for the repression of unique transposons found within the genome of these mammals (Barau et al., 2016). Given the often-unique transposon populations of different organisms and the selective pressure of reproduction, it is perhaps unsurprising that different mechanisms for transposon repression during sexual reproduction have arisen between species.

1.7 Small RNA in sexual reproduction

As described above, sRNAs are used by plants to target transposons for repression via DNA methylation. sRNAs allow the recognition of specific mRNA or DNA sequences for targeted regulation through sequence complementarity. A prime example of such targeting in eukaryotes is the use of microRNAs (miRNAs) that are typically 21 nucleotide (nt) long and are produced from genomic DNA to target mRNAs for post-transcriptional regulation through mRNA degradation or translational repression (O'Brien et al., 2018).

Along with miRNAs, numerous specialised sRNA populations have been discovered, including the 24 nt sRNAs used in RdDM and the tRFs used in retrotransposon repression. Specific sRNA populations are often observed during sexual reproduction. For example, Piwi-interacting sRNAs (piRNAs) are enriched in the germlines of animals and are important for transposon repression as well as

gene regulation and germline viability (Weick and Miska, 2014). In plants, phased sRNAs (phasiRNAs) are produced primarily from miRNA-directed degradation of transcripts in the somatic cell layers of the anther and are present in both monocots and eudicots (Xia et al., 2019). However, the precise roles of these phasiRNAs are presently unknown (Patel et al., 2018). The selective recognition of transposon sequences through complementary sRNAs is a powerful defence mechanism at a host organism's disposal, particularly regarding the recognition of related transposon species that may be introduced from other individuals during sexual reproduction.

1.8 *Arabidopsis thaliana* and *Marchantia polymorpha*

In *A. thaliana* and all flowering plants, the diploid sporophyte generation is the dominant stage of life. Diploid meristematic cells generate somatic tissues and, following vegetative development, initialise the sexual lineage by developing into meiocytes in the flower. In the case of the male sexual lineage, these meiocytes give rise to haploid microspores via meiosis (Figure 1.3; Feng et al., 2013). The microspores subsequently divide mitotically to produce the vegetative and generative cells (Figure 1.3). The generative cell enters one more round of mitosis to generate two sperm cells, which are engulfed within the vegetative cell in the mature pollen grain (Figure 1.3). One sperm cell fertilises the egg cell to produce the diploid zygote that will go on to form the mature plant, while the other sperm fertilises the central cell to form the endosperm (Feng et al., 2013). Meiosis and pollen development occurs within the anther while surrounded by a somatic layer known as the tapetum, which supports the male sexual lineage and plays a critical role in pollen coat formation (Parish et al., 2012; Quilichini et al., 2015).

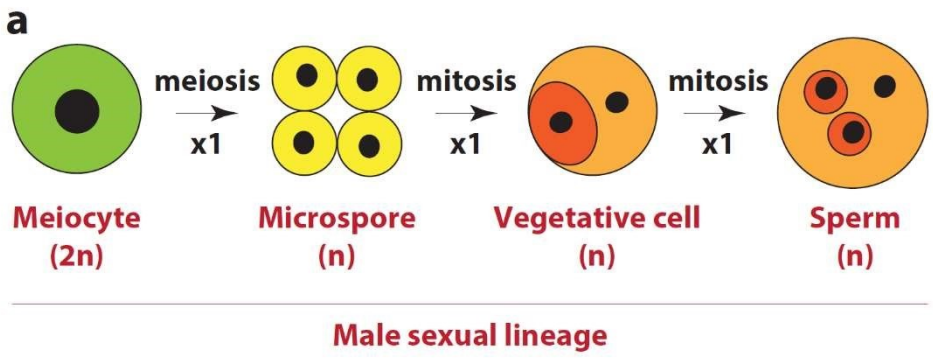


Figure 1.3 Model of male sexual lineage development in *Arabidopsis thaliana*.
n, the number of chromosomes in the haploid genome.

While the sporophyte is the dominant stage of life for flowering plants (with the gametophyte reduced to two mitotic divisions in the male sexual lineage from microspore to sperm), basal land plants such as the liverwort *Marchantia polymorpha* develop their vegetative tissues during the gametophytic phase of development. As such, the major stage of life for this organism is haploid. Critically, meiosis and the production of sperm are separated, allowing the DNA methylation profiles associated with these two processes to be examined independently. The development and methylation profile of *M. polymorpha* are described in more detail within Chapter Four.

1.9 Focus of this study

Aside from demethylation in specialised cells that support sexual development, DNA methylation within *A. thaliana* and flowering plants was thought not to be reprogrammed in the cells that contribute to the next generation as it does in animals due to the strong association of DNA methylation with transposons and the high fidelity of DNA methylation across generations (Feng et al., 2010; Pikaard and Mittelsten Scheid, 2014). Examples of CHH methylation reinforcement or loss have been observed in specific root cells (i.e. the columella), seeds, embryo, microspores, and sperm (Bouyer et al., 2017; Calarco et al., 2012; Kawakatsu et al., 2017; Kawakatsu et al., 2016), while CG and CHG methylation is reinforced in gametes in a manner that is thought to aid inheritance of DNA methylation marks to the next generation (Hsieh et al., 2016). However, more drastic alterations of DNA methylation targeting have been previously unobserved despite the remarkable sRNA biology that exists during plant sexual reproduction.

In this thesis, I assess the DNA methylation profile in close detail during plant development, with a focus on *A. thaliana* and the emerging model system *M. polymorpha*. Chapter Two focuses on published results (Walker et al., 2018), showing the presence of male sexual-lineage-specific hypermethylation and gene-associated *de novo* methylation in *A. thaliana* sex cells (i.e. meiocytes, microspores, and sperm cells) that is the result of RdDM activity and is important for gene repression, splicing and meiosis. In Chapter Three, the mechanisms underlying this male sex cell specific methylation are explored, with indications that the gene-associated *de novo* methylation is the result of trans targeting of sRNA produced from the hypermethylated loci, with the latter being a result of specific POLIV-recruiting CLASSYs that are highly expressed in the tapetum and meiocyte. Chapter Four explores the methylation profile of *M. polymorpha*, revealing a unique global methylation profile like vertebrate systems that is accompanied by transposon

methylation reinforcement. Chapter Four also identifies gene-targeted *de novo* methylation in *M. polymorpha* sperm and sporophytes that is unique to both tissues, with similarities between the *de novo* methylation of *M. polymorpha* sporophytes and *A. thaliana* male sex cells. Results from each of the three results chapters are discussed further in Chapter Five within the context of each other, the wider scope of the field and preliminary data emerging from the Feng lab.

Chapter Two

Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis thaliana*

"But I'm not interested in the female sex" – Multiple sources from the Feng lab

This chapter takes work from published work (Walker et al., 2018). The most significant modifications include section 2.4, further clarifying the distinction between canonical sexual-lineage-specific hypermethylation and *de novo* sexual-lineage-specific methylation, section 2.7 detailing the association of SLMs with chromatin genes, and the overall discussion.

2.1 Abstract

DNA methylation regulates eukaryotic gene expression and is extensively reprogrammed during animal development. However, whether developmental methylation reprogramming during the sporophytic life cycle of flowering plants regulates genes is presently unknown. Here we report a distinctive gene-targeted RNA-directed DNA methylation activity in the *Arabidopsis thaliana* male sexual lineage that regulates gene expression in meiocytes. Loss of sexual-lineage-specific RdDM causes mis-splicing of the MPS1 gene (also known as PRD2), thereby disrupting meiosis. Our results establish a regulatory paradigm in which *de novo* methylation creates a cell-lineage-specific epigenetic signature that controls gene expression and contributes to cellular function in flowering plants.

2.2 Introduction

Cytosine methylation is an ancient DNA modification catalysed by methyltransferases that are conserved across eukaryotes, including plants and animals (Zemach and Zilberman, 2010). DNA methylation patterns are faithfully replicated during cell division, thus allowing methylation to carry epigenetic information throughout cellular lineages (He et al., 2011; Law and Jacobsen, 2010). In the complex genomes of flowering plants and vertebrates, methylation heritably silences transposons, thereby maintaining genome integrity and transcriptional homeostasis (He et al., 2011; Law and Jacobsen, 2010). In agreement with this function, DNA methylation of regulatory sequences, especially those near transcriptional start sites, is strongly associated with gene silencing (Kim and Zilberman, 2014; Pikaard and Mittelsten Scheid, 2014).

Beyond its homeostatic function, DNA methylation can be reprogrammed during development, thus allowing it to regulate gene expression. In mammals, this phenomenon has been observed in numerous cellular lineages and appears to be a common regulatory mechanism (Dawlaty et al., 2014; Kubo et al., 2015; Slieker et al., 2015). In plants, gene expression in the transient extraembryonic endosperm tissue is controlled by active DNA demethylation, which occurs in the central cell (i.e. the companion cell of the egg) that is fertilised to give rise to the endosperm (Park et al., 2016; Rodrigues and Zilberman, 2015). As discussed in Chapter One, a similar active demethylation process also occurs in the vegetative cell, a terminally differentiated companion cell of the sperm (Calarco et al., 2012; Ibarra et al., 2012; Rodrigues and Zilberman, 2015). Beyond the endosperm and gamete-companion cells, there are intriguing examples of altered methylation levels and patterns in different cell types (Calarco et al., 2012; Hsieh et al., 2009; Ibarra et al., 2012; Kawakatsu et al., 2016) and during responses to biotic and abiotic stimuli (Bilichak et al., 2012; Dowen et al., 2012; Secco et al., 2015). However, it is unclear whether gene expression is controlled by developmental reprogramming of DNA methylation in plant cells that contribute to the next generation.

To investigate this question, we analysed DNA methylation in the male sexual lineage of *Arabidopsis thaliana*. This analysis allowed us to uncover a sexual-lineage-specific DNA methylation signature deposited by the RdDM pathway. We further demonstrated that this *de novo* methylation regulates gene expression and splicing, and is required for normal meiosis, establishing compelling links among DNA methylation reprogramming, gene expression and developmental fate.

2.3 The germline of *A. thaliana* features high CG and low CHH methylation compared to other tissues

To comprehensively understand DNA methylation reprogramming within the entire male sexual lineage, from the diploid meiocytes through to microspores and sperm cells, Dr. Hongbo Gao generated a genome-wide methylation profile for male *A. thaliana* meiocytes (Supplementary Table 2.1), which we compared with the previously sequenced profiles of the microspore, sperm and vegetative cell (Calarco et al., 2012; Ibarra et al., 2012). Contrary to previous speculations that DNA demethylation occurs in male meiocytes (Kawashima and Berger, 2014), we found that meiocyte methylation resembles that in the microspores and sperm, with high levels of CG and CHG methylation over transposons (where H = A, C, or T; Figure 2.1a-d). This result is consistent with robust transposon silencing in the germline, an essential function for ensuring genetic integrity across generations (Hsieh et al., 2016; Park et al., 2016). In the CHH context, the microspore and sperm cells of the germline have low levels of methylation, particularly in heterochromatic loci, relative to that of somatic tissues and especially vegetative cells (Figure 2.1a; Figure 2.1e; Calarco et al., 2012; Hsieh et al., 2016; Ibarra et al., 2012). However, the male meiocyte has even lower levels of CHH methylation than microspore and sperm cells (Figure 2.1e). This is most notable in euchromatic regions, suggesting a lack of global RdDM activity in meiocytes. Indeed, the methylomes of *drm1drm2* double mutant meiocyte and sperm cells (generated by Dr. Hongbo Gao along with the methylome of vegetative cells carrying the same mutations) show marginally less CHH methylation compared to wild-type meiocytes (Figure 2.1e; Supplementary Figure 2.1). Low levels of CHH methylation in microspore and sperm have been proposed to result from a lack of methylation maintenance during meiotic division (Calarco et al., 2012). Our results demonstrate that this is not the case; instead, CHH methylation is already low in meiocytes and undergoes an overall increase at euchromatic loci during the development of the male sexual lineage.

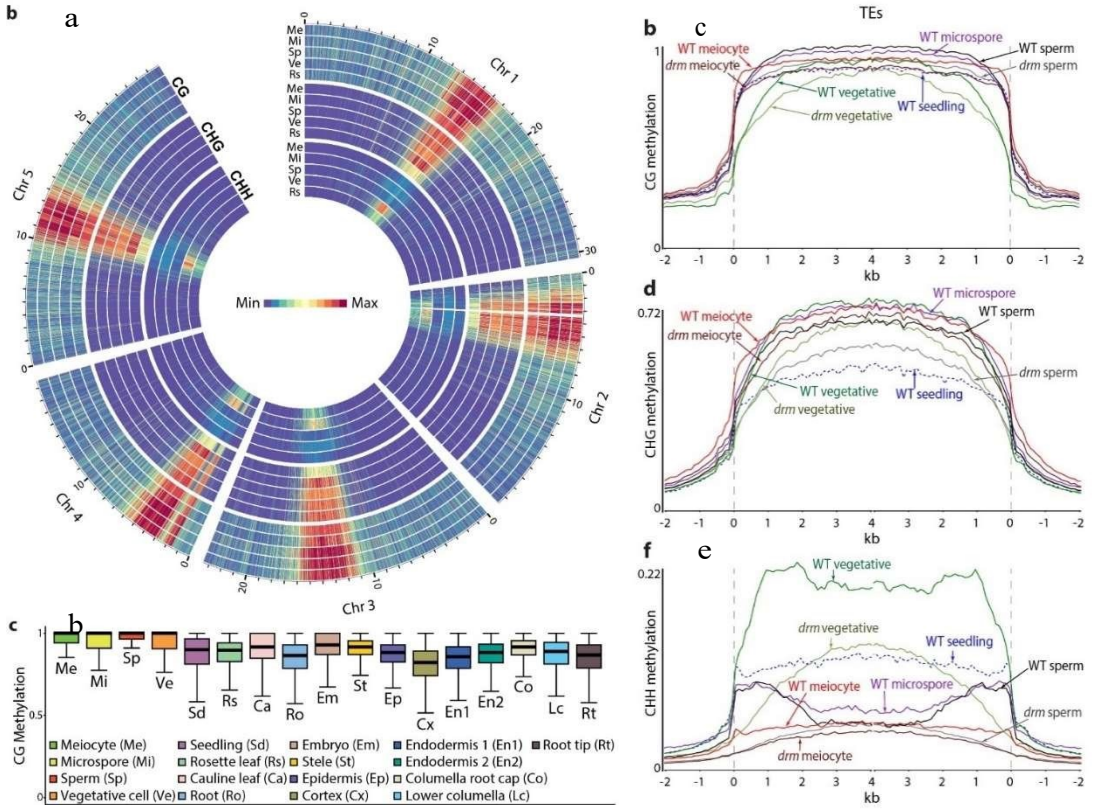


Figure 2.1 Male meiocytes exhibit high CG/CHG and low CHH methylation. a, Heat maps showing CG, CHG and CHH methylation of the male sexual lineage comprising the meiocyte (Me), microspore (Mi), sperm (Sp) and vegetative cell (Ve), in comparison to the rosette leaf (Rs). Methylation was calculated and is presented in 10 kb windows, with the maximum set at the highest value among the five tissues for each context. The region enriched in mitochondrial sequences on chromosome (chr) 2 (3.23–3.51 Mb) is removed. **b**, Box plots showing CG methylation for individual CG sites located within annotated transposons, and with methylation >50% and at least ten informative sequenced cytosines. Each box encloses the middle 50% of the distribution, with the horizontal line marking the median and vertical lines marking the minimum and maximum values falling within 1.5 times the height of the box. **c** – **e**, *A. thaliana* transposable elements (TEs) were aligned at the 5’ end (left panel) or the 3’ end (right panel), and average methylation levels in the CG (**c**), CHG (**d**) or CHH (**e**) context for each 100 bp interval are plotted. WT, wildtype. *drm*; *drm1drm2* double mutants. The dashed line at zero represents the point of alignment.

2.4 The male sexual lineage exhibits specific regions of hyper and *de novo* RdDM activity

In addition to the global methylation patterns exhibited within the *A. thaliana* male sexual lineage, a comparison between male sex cells and somatic tissues (seedlings, rosette leaves, cauline leaves and roots) revealed regions that were strongly hypermethylated in sex cells (Figure 2.2a; Supplementary Figure 2.1). Furthermore, loci hypermethylated in one male sex cell type tended to be hypermethylated in other sex cells (Figure 2.2a; Figure 2.2b; Supplementary Figure 2.1). This hypermethylation was most prominent in the CHH context but also encompassed other contexts (Figure 2.2a; Figure 2.2b; Supplementary Figure 2.1), meaning that the same locus was often hypermethylated at CG, CHG and CHH sites (Figure 2.2a; Supplementary Figure 2.1). Among the 1,301 identified loci that were consistently differentially methylated between sex cells and somatic tissues, most (1,265; 97%) were hypermethylated in sex cells (Supplementary Table 2.2). These loci are hereafter referred to as sexual-lineage-specific hypermethylated loci (SLHs).

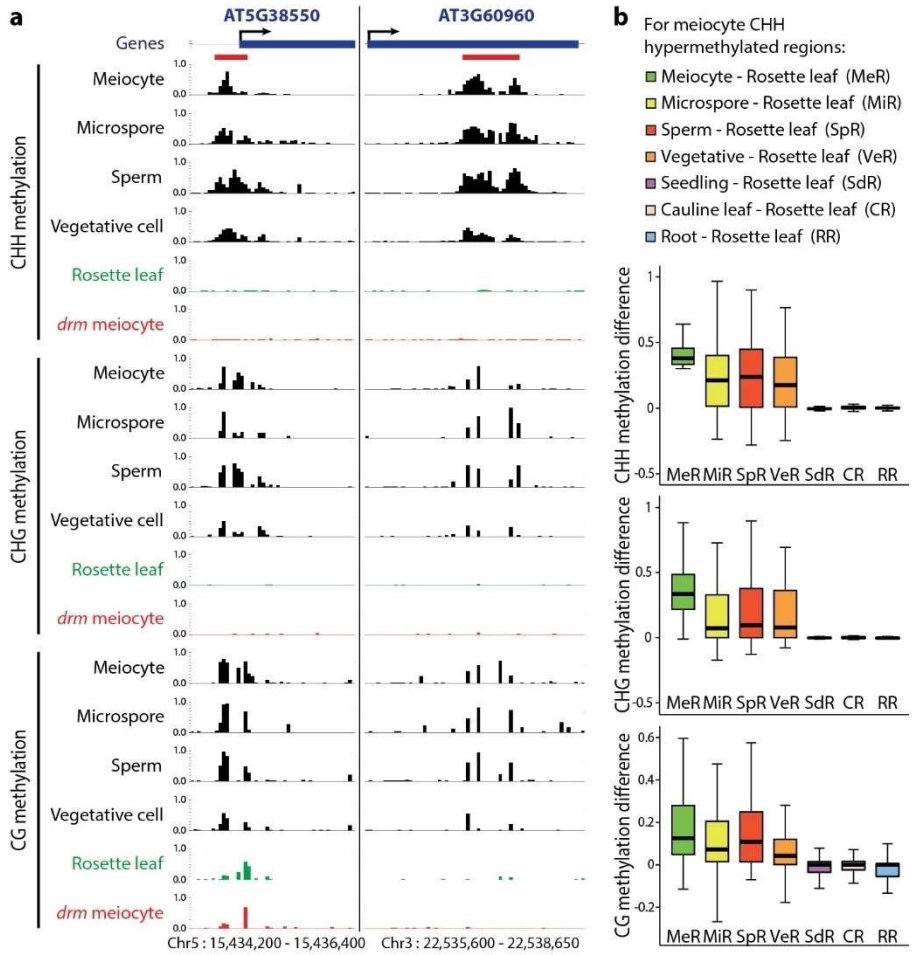


Figure 2.2 SLHs in *A. thaliana*. a, Snapshots of cytosine methylation in wild-type male sex cells (black), rosette leaves (green) and *drm1drm2* (*drm*) double mutant meiocytes (red) at two example SLHs. SLHs (full list in Supplementary Table 2.2) are underlined in red. Methylation patterns in other somatic tissues and *drm* sex cells are shown in Supplementary Figure 2.1. **b**, Box plots as in Figure 2.1b showing absolute methylation difference between specific cells/tissues and rosette leaves for 50 bp windows that are CHH hypermethylated in meiocytes in comparison to rosette leaves.

SLHs resemble targets of the RdDM pathway, which establishes and maintains methylation in all sequence contexts but is particularly important for CHH methylation (He et al., 2011; Matzke and Mosher, 2014; Pikaard and Mittelsten Scheid, 2014). Accordingly, the methylomes of *drm1drm2* double mutant meiocytes, sperm and vegetative cells, as well as the methylomes of *rdr2* sperm and *rdr2* vegetative cells (the latter two methylomes also being produced by Dr. Hongbo Gao) show hypomethylation in all cytosine contexts (Figure 2.2a; Figure 2.3a; Supplementary Figure 2.1). Furthermore, almost all SLHs (1257 loci) have significantly less methylation in *drm1drm2* double mutant sex cells compared to wild-type sex cells ($P < 0.001$; Fisher's exact test). Collectively, our results demonstrate that RdDM is the underlying mechanism producing SLHs.

SLHs can be due to increased RdDM activity at canonical targets, an expansion of RdDM into novel targets to create regions of *de novo* methylation, or both. Indeed, we separated the SLHs into canonical SLHs (724 loci) that have distinctive levels of CHG and/or CHH methylation in somatic tissues and sexual-lineage-specific methylated loci (SLMs; 533 loci), which lacked non-CG methylation in somatic tissues (Supplementary Figure 2.1; Supplementary Table 2.2). The size of canonical SLHs and SLMs (726 bp and 264 bp, respectively) are typical for RdDM loci and together encompass 0.6% of the nuclear genome.

To further evaluate the cell and tissue specificity of canonical SLHs and SLMs, we compared methylation of these regions to the methylation profiles of root cap columella cells, which exhibits high levels of RdDM-associated CHH methylation (Kawakatsu et al., 2016). Whereas canonical SLHs show substantial methylation in all contexts in columella cells (Figure 2.3b), SLMs have low amounts of CHG and CHH methylation (Supplementary Figure 2.2). Consistently, 76% (551 out of 724 loci) of the canonical SLHs overlap with published columella differentially methylated regions (DMRs; Kawakatsu et al., 2016), while 88% (469 out of 533 loci) of the SLMs do not overlap columella DMRs (Figure 2.3c). The canonical SLHs also have increased levels of methylation in the embryo – another tissue with reported CHH hypermethylation (Figure 2.3b; Hsieh et al., 2009) – while the 469 SLMs not overlapping columella DMRs again show no CHG or CHH methylation (Figure 2.3c). We used this group of highly specific SLMs in all subsequent analyses (Supplementary Table 2.2).

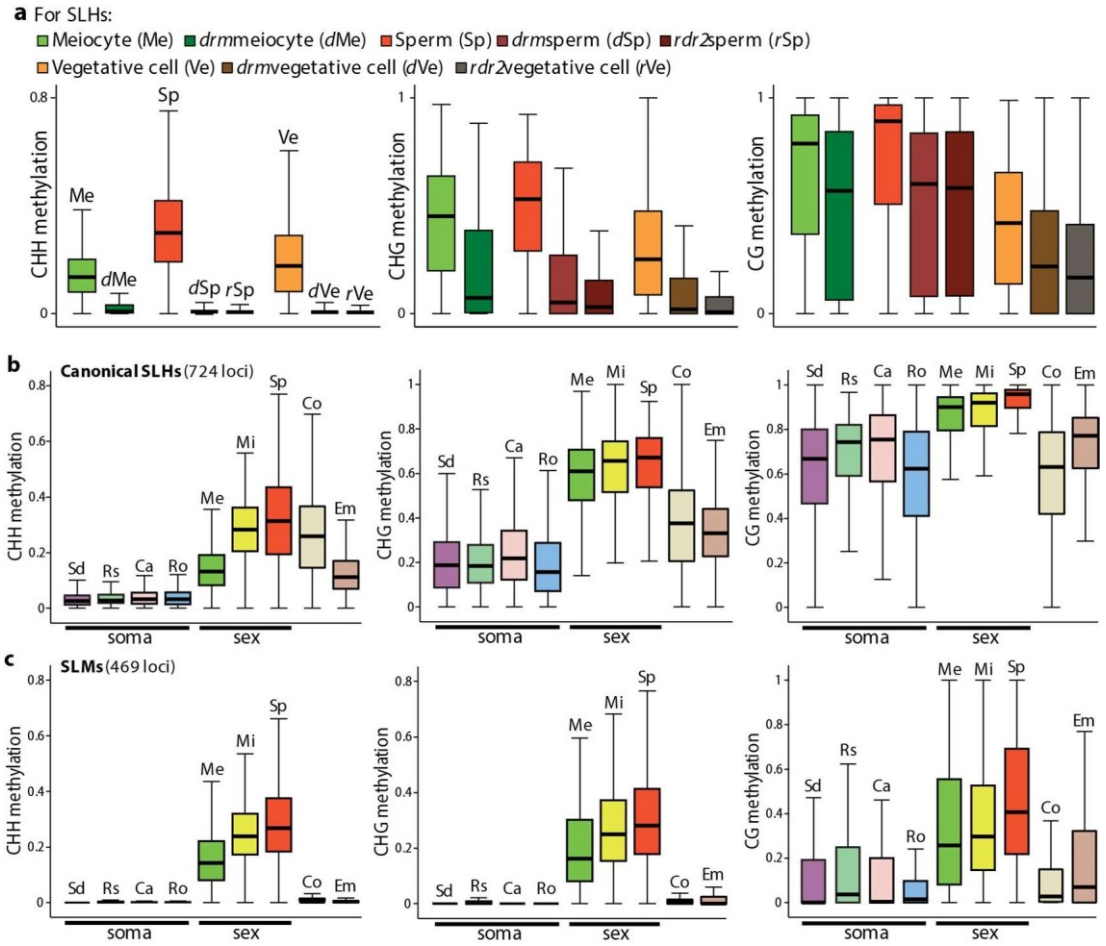


Figure 2.3 SLHs are produced by RdDM and are either hypermethylated (canonical SLHs) or *de novo* methylated (SLMs). a, Box plots as in Figure 2.1b, showing the absolute methylation at SLHs in wild-type meiocytes (Me), sperm (Sp) and vegetative cell (Ve) in comparison to *drm1drm2* (*drm*) double mutant meiocyte (*dMe*), sperm (*dSp*), vegetative cell (*dVe*) and *rdr2* mutant sperm (*rSp*) and vegetative cell (*rVe*). **b** - **c**, Box plots as in Figure 2.1b, showing the absolute methylation at canonical SLHs (**b**) and SLMs (**c**) in somatic tissues (Sd, seedling; Rs, rosette leaf; Ca, cauline leaf; Ro, root), sex cells (Me, meiocyte; Mi, microspore; Sp, sperm), columella root cap (Co) and embryo (Em).

2.5 SLHs lose methylation through both active and passive processes

DNA methylation in plants can be removed actively by DNA glycosylases that excise methylated cytosines (He et al., 2011). Alternatively, DNA methylation can be lost passively over several cell divisions if it is not maintained by methyltransferases. To test whether the canonical SLHs and SLMs are hypomethylated in somatic tissues due to active demethylation, we compared wild-type rosette leaf methylation to available methylation data for rosette leaves with mutations in the three somatically expressed demethylase genes (*rdd*; ROS1, DML2 and DML3; Stroud et al., 2013). CG and CHG methylation at SLMs in these *rdd* mutants is much higher than in wildtype control leaves (Figure 2.4). Similarly, CG and CHG methylation in *rdd* leaves is also higher at canonical SLHs than in wild-type, while minimal differences are observed for non-SLH methylated regions (Figure 2.4). CHH methylation is only marginally higher in *rdd* leaves compared to wild-type and much lower than in sex cells for both SLMs and canonical SLHs, however (Figure 2.4). Given that RdDM-established CG methylation, but not CHH methylation, is known to be maintained in the absence of RdDM (Matzke and Mosher, 2014), our data suggest that active demethylation removes some of the CG (and CHG) methylation that is introduced by RdDM in the sexual lineage.

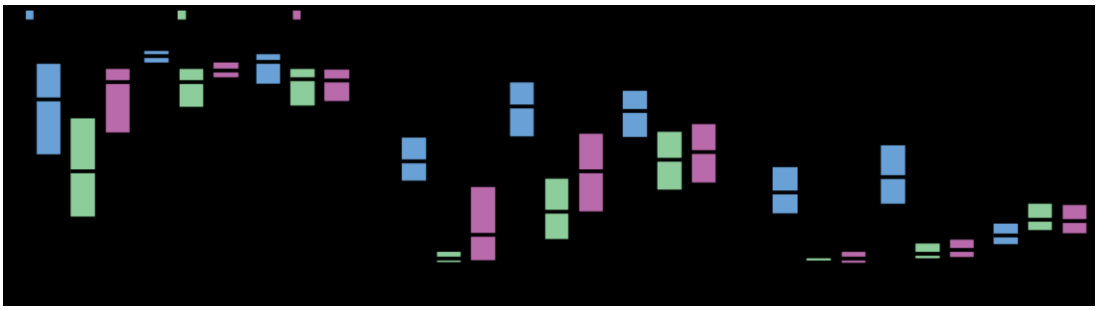


Figure 2.4 Active demethylation of SLHs. Box plots as in Figure 2.1b, showing the absolute methylation at SLM, canonical SLH and non-SLH 50 bp windows in ros1; dml2; dml3 (*rdd*)-mutant rosette leaves, wild-type rosette leaves and wild-type male sex cells (meiocyte, microspore and sperm).

2.6 In soma, residual CG methylation at SLMs is maintained by MET1

Although SLMs lack non-CG methylation in somatic tissues, some CG methylation is present (Figure 2.3c; Supplementary Figure 2.1). This remnant CG methylation may have been induced by sexual-lineage-specific RdDM and maintained in somatic tissues by MET1 or may have resulted directly from somatic RdDM activity. To distinguish between these possibilities, we analysed SLM CG methylation in RdDM mutant somatic tissues, which showed overall levels similar to those of wild-type somatic tissues (Figure 2.5a). Furthermore, SLM CG methylation in RdDM mutant somatic tissues (*drd1*, *drm2* and *rdr2*) correlated with that in wild-type tissues (Pearson's R = 0.80, 0.58 and 0.70, respectively; Figure 2.5b; Supplementary Figure 2.3), demonstrating that RdDM is not required to maintain somatic CG methylation at SLMs.

The hypothesis that CG methylation at SLMs is initiated by RdDM in sex cells and is maintained at lower levels by MET1 in the absence of RdDM in somatic cells yields several predictions. First, MET1 should be able to maintain CG methylation in sex cells without RdDM at levels like those in wild-type somatic tissues. Indeed, SLM CG methylation in *drm1drm2* double mutant sex cells was like that in wild-type somatic tissues (Figure 2.5a; Supplementary Figure 2.1g; Supplementary Figure 2.1h) and was strongly correlated with that in wild-type somatic tissues (Pearson's R = 0.76; Figure 2.5c). Second, somatic CG methylation at SLMs should be MET1 dependent. Indeed, this is what we observe (Figure 2.5d). Finally, CG methylation at SLMs should be re-established after it is erased, because CG methylation is known to be reconstituted at loci that are targeted by RdDM in a manner that is not dependent on pre-existing CG methylation (i.e., at loci where RdDM still functions in *met1* mutants (Catoni et al., 2017). Indeed, somatic CG methylation at SLMs was restored to wild-type levels through introduction of functional MET1 into *met1*-mutant plants (Figure 2.5d). Together, our analyses demonstrate that SLMs are products of sexual-lineage-specific RdDM activity, which establishes methylation in all sequence contexts. In somatic tissues, residual CG methylation at SLMs is maintained by MET1 in the absence of RdDM.

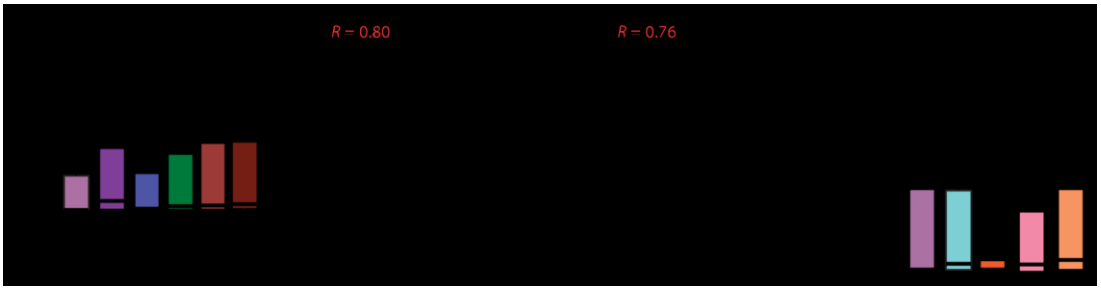


Figure 2.5 CG methylation established by RdDM at SLMs is retained in soma by MET1. a, Box plots as in Figure 2.1b, illustrating CG methylation at SLMs in wildtype (WT) seedlings (other somatic tissues are shown in Figure 2.4c), and seedlings and sex cells from RdDM mutants: *drm1drm2* (*drm*) and *rdr2*. **b**, Scatter plot showing the linear correlation between CG methylation in wild-type (y axis) and drd1 mutant (x axis) roots at SLMs (Pearson's R = 0.80). **c**, Scatter plot showing the linear correlation between average CG methylation in wild-type somatic tissues (cauline leaf, rosette leaf, root and seedling; y axis) and that in *drm* mutant sex cells (meiocytes, sperm and vegetative cell; x axis) at SLMs (Pearson's R = 0.76). **d**, Box plots as in Figure 2.1b, demonstrating the absolute CG methylation at SLMs in wildtype seedlings (same data as used in Figure 2.3d and Figure 2.5a), and published data (denoted by dagger † symbols; Catoni et al., 2017 including wild-type control seedlings (Sd†), *met1* mutants and MET1 reintroduction lines (T-MET1a T2 and TMET1b T5).

2.7 RdDM-induced sexual-lineage-specific methylation regulates gene expression in meiocytes

As RdDM is typically associated with transposons, we were curious whether there were any notable associations of SLHs with genomic features. We therefore analysed the location of SLMs and canonical SLHs compared to other RdDM targets in relation to genes and transposons. Canonical SLHs correspond mostly to transposons but overlap genes more frequently than other RdDM targets (Figure 2.6a). Furthermore, canonical SLHs were more likely to overlap annotated transposons than randomly selected sets of loci that were comparably located in relation to genes throughout the genome but were less likely to overlap transposons than other RdDM-target loci (Supplementary Table 2.3). Surprisingly, the majority of SLMs overlapped genes (268 SLMs; 57%; Figure 2.6a) and were even slightly less likely than random control loci to overlap annotated transposons (Supplementary Table 2.3). These results indicate that canonical SLHs are an extension of conventional transposon-targeted RdDM, a result consistent with their methylation in some somatic cell types, whereas SLMs represent a novel targeting of RdDM to gene-associated loci.

As SLMs are associated with genes and suppression of gene expression by DNA methylation is well known (He et al., 2011), we analysed mRNA levels in *drm1drm2* double mutant meiocytes and wild-type controls by RNA-seq (data produced by Dr. Hongbo Gao). The expression of meiosis-associated genes was substantially enriched in our data compared with published meiocyte RNA-seq results (Supplementary Table 2.4; Chen et al., 2010; Yang et al., 2011), suggesting high meiocyte purity. Among the 47 genes with a greater than four-fold change in expression between wild-type and *drm1drm2* double mutant meiocytes, all of which are activated in the mutant, seven overlap an SLM and one has an SLM within 20 bp (Figure 2.6b; Figure 2.6c; Supplementary Figure 2.4; Supplementary Table 2.5). This is a much higher fraction (17%) than expected by random chance (Fisher's exact test, $P = 1.40 \times 10^{-8}$), as only 0.9% of nuclear genes are within 20 bp of an SLM. All four of the SLM-associated genes that are overexpressed in *drm1drm2* double mutant meiocytes and significantly differentially expressed between meiocytes and leaves are suppressed in meiocytes compared to leaves (Figure 2.6c; Supplementary Table 2.5). Furthermore, expression levels of these genes in leaves are not elevated by RdDM mutation (Supplementary Table 2.5). Our data therefore demonstrates that the *de novo* targeting of RdDM in the male sex cells towards genes regulates male sex cell gene expression.

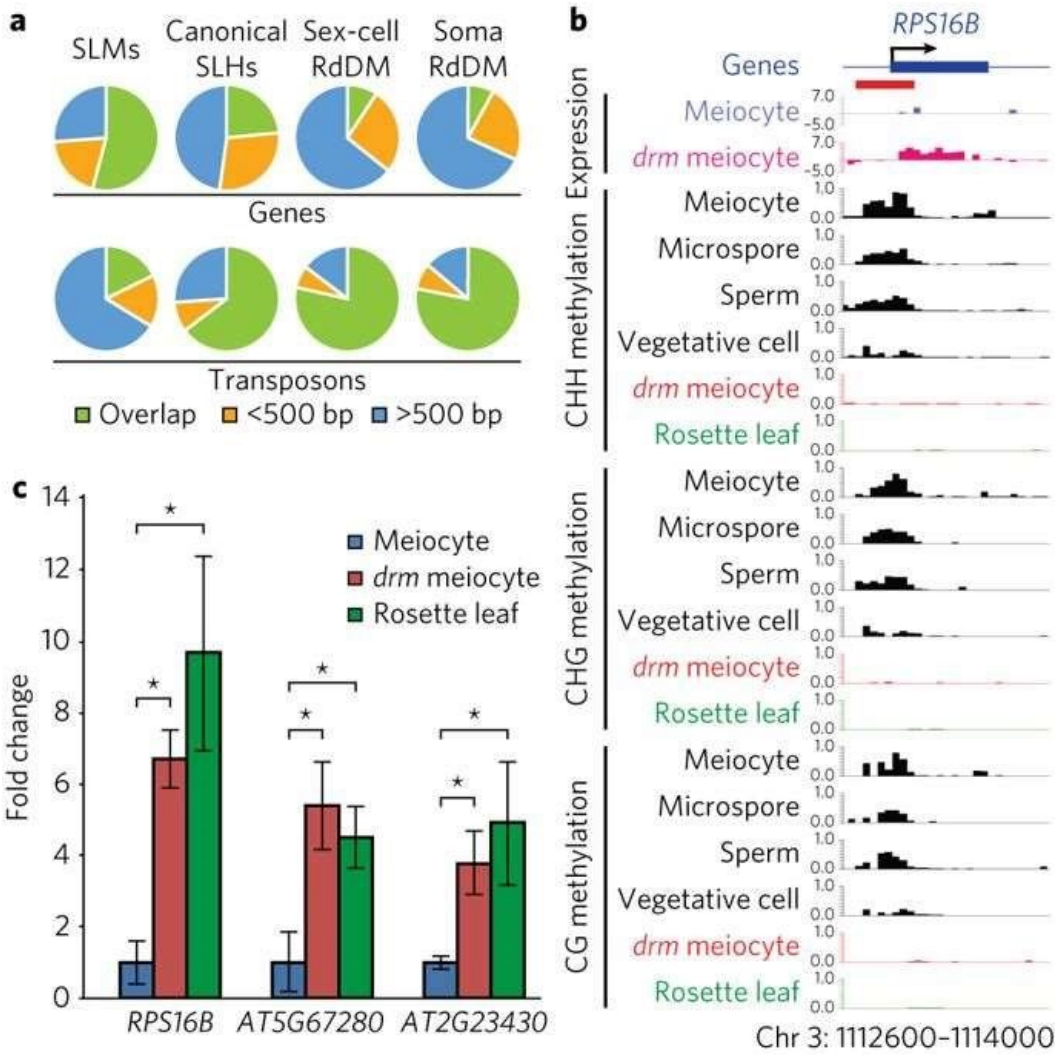


Figure 2.6 SLMs target genes and regulate gene expression in meiocytes. a, Pie charts illustrating percentages of SLMs, canonical SLHs and other RdDM-target loci overlapping (green) within 500 bp of (yellow) and more than 500 bp from (blue) genes or transposons (numbers shown in Supplementary Table 2.3). **b**, Snapshots of transcription (in $\log_2$ RPKM) and DNA methylation (similar to Figure 2.2a) at the RPS16B gene. *drm*, *drm1drm2*; SLM is underlined in red. **c**, Quantitative RT–PCR showing the expression of three SLM-regulated genes. *P < 0.02 (two-tailed t test; n = 3 RNA replicates extracted from independently isolated meiocytes or leaves).

While assessing the genes associated with SLMs and canonical SLHs, I identified SLMs in the gene bodies of the methylase genes DRM2, CMT3 and CMT1, as well as the demethylase genes ROS1 and DML3 (Supplementary Figure 2.5). A GO analysis by Dr Billy Aldridge confirmed an enrichment for chromatin genes in the list of genes associated with SLMs, further identifying DEFECTIVE IN RNA-DIRECTED DNA METHYLATION (DRD1), a chromatin remodeller required for POLV activity (Law et al., 2010), as well as VIM2, VIM3 and VIM4 (required for the maintenance of CG methylation; Feng et al., 2010; Woo et al., 2008). However, the specific role of these SLMs remains to be seen as the associated genes were not differentially expressed between wild-type and *drm1drm2* double mutant meiocytes.

Among the other genes containing SLMs, an unexpected group comprised genes encoding pre-tRNAs (denoted pre-tRNA genes). 24 pre-tRNA loci overlapped SLMs and showed a preference for specific anticodons: for example, 75% and 21% of the phenylalanine and methionine pre-tRNA genes, respectively, were covered by SLMs (Figure 2.7a; Supplementary Figure 2.6; Supplementary Table 2.6a); these numbers were substantially higher than expected by chance (both $P < 2.63 \times 10^{-6}$, Fisher's exact test). As our criteria for calling SLMs were highly stringent, we performed a genome-wide analysis to specifically detect sexual-lineage hypermethylation of pre-tRNA genes. We found an additional set of 60 pre-tRNAs with significantly more CHH and CHG methylation in at least two of the sex cells in comparison to somatic tissues, and in wild-type sex cells in comparison to *drm1drm2* double mutant sex cells (both $P < 0.001$, Fisher's exact test; Figure 2.7b; Supplementary Figure 2.7; Supplementary Table 2.6b). Together, the 84 hypermethylated loci included 100%, 75%, 73% and 42% of the phenylalanine, valine, cysteine and methionine pre-tRNA genes, respectively (Supplementary Table 2.6b). Consistently, 24 nt sRNAs were enriched at these pre-tRNA genes in pollen but not in shoots (Figure 2.7c). The preferential hypermethylation of certain pre-tRNA genes, together with the recent discovery of small tRNA fragments in *A. thaliana* pollen (Martinez, 2017), suggests that tRNA biology may have interesting aspects particular to sex cells (see section 2.9 for discussion).

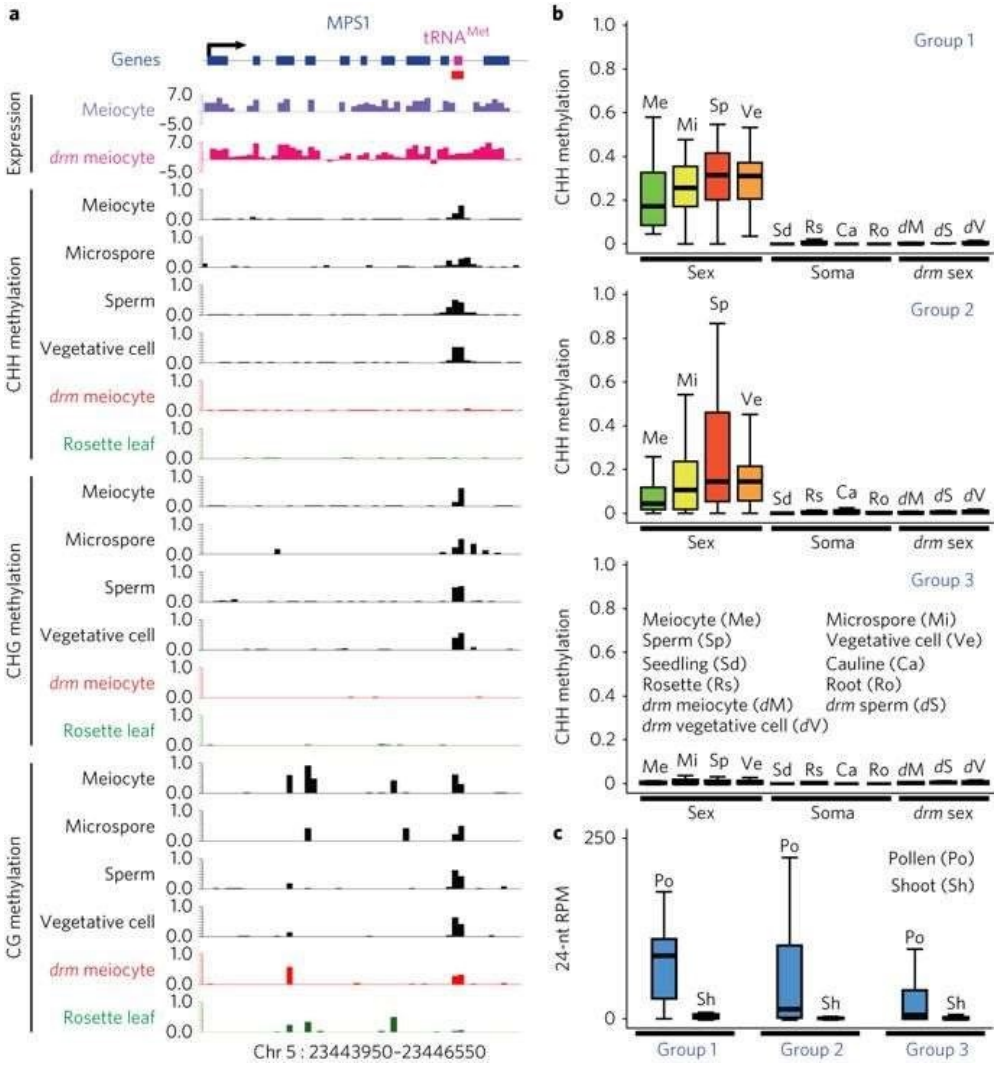


Figure 2.7 Pre-tRNA genes are hypermethylated in the male sexual lineage. a, Snapshots of transcription and DNA methylation (similar to Figure 2.6b) at the methionine pre-tRNA locus (magenta box) in the last intron of the MPS1 gene. *drm*, *drm1drm2*; SLM is underlined in red. Methylation patterns in other somatic tissues and *drm* sex cells are shown in Supplementary Figure 2.6b. **b**, Box plots as in Figure 2.1b, showing the absolute CHH methylation at three groups of pre-tRNA genes in sex cells (Me, meiocyte; Mi, microspore; Sp, sperm; Ve, vegetative cell), somatic tissues (Sd, seedling; Rs, rosette leaf; Ca, cauline leaf; Ro, root), and *drm* mutant sex cells (dM, *drm* meiocyte; dS, *drm* sperm; dV, *drm* vegetative cell). Group 1, the 24 pre-tRNA genes that overlap SLMs; group 2, the additional 60 genes hypermethylated in the sexual lineage by RdDM; group 3, the remaining 605 nuclear pre-tRNA genes. **c**, Box plots as in Figure 2.1b, demonstrating the abundance of 24 nt sRNA in the pollen (Po) or shoot (Sh) among the three groups of pre-tRNA genes described above.

2.8 An SLM regulates the splicing of MPS1 and is important for meiosis

I noticed that one SLM-covered methionine pre-tRNA gene occurs within the last intron (between exons 9 and 10) of another gene, MULTIPOLAR SPINDLE 1 (MPS1; also known as PUTATIVE RECOMBINATION INITIATION DEFECTS 2 (PRD2); Figure 2.7a; Figure 2.8a; Supplementary Figure 2.6b). DNA methylation has been shown to influence splicing in plants (Regulski et al., 2013; Wang et al., 2016), and animals (Lev Maor et al., 2015), and after we examined our RNA-seq data we detected cDNA reads that indicate incorrect splicing of MPS1 at the last intron in *drm1drm2* double mutant meiocytes (Figure 2.7a; Figure 2.8a). Quantitative RT-PCR analysis carried out by Dr. Hongbo Gao demonstrated that 28% of the mature MPS1 mRNA retains the last intron in *drm1drm2* double mutant meiocytes, whereas no such retention occurs in wild-type (Figure 2.8b), confirming that the SLM within the intron is required for correct splicing of the MPS1 transcript.

MPS1 is required for *A. thaliana* meiosis, with mutants causing polyads – meiotic products numbering other than the typical four microspores (De Muyt et al., 2009; Jiang et al., 2009). Retention of the last intron introduces a premature stop codon (Figure 2.8a) and one of the described loss-of-function alleles affects splicing between exons 9 and 10, while another is an insertion in the intervening intron, indicating that exon 10 is essential for MPS1 activity (De Muyt et al., 2009). Accordingly, a significant occurrence of cellular triads in RdDM mutants was found by Dr. James Higgins (7.1% and 7.8% in *drm1drm2* and *rdr2*, respectively; Figure 2.8b; Figure 2.8d). Pentads in *drm1drm2* and *rdr2* mutants were also found (Supplementary Figure 2.8d; Supplementary Figure 2.8e), as has been reported for other RdDM mutants (Oliver et al., 2016), whereas triads or pentads were not observed in wild-type plants. Introduction of an MPS1 transgene lacking the last intron into the *drm1drm2* double mutant background reduced the number of meiotic triads (4.1% and 3.6% for two independent complementation lines; Figure 2.8c), but not to the undetectable level of wild-type plants, suggesting interference by the unspliced product. Consequently, the introduction of an MPS1 transgene with mutations that prevent splicing of the last intron into wild-type plants resulted in a substantially higher percentage of meiotic triads than *drm1drm2* or *rdr2* mutants (13.4% and 15.1% for two independent interference lines; Figure 2.8c). Thus, our results indicate that loss of methylation at the SLM within the last intron of MPS1 causes intron retention, the product of which interferes with meiosis.

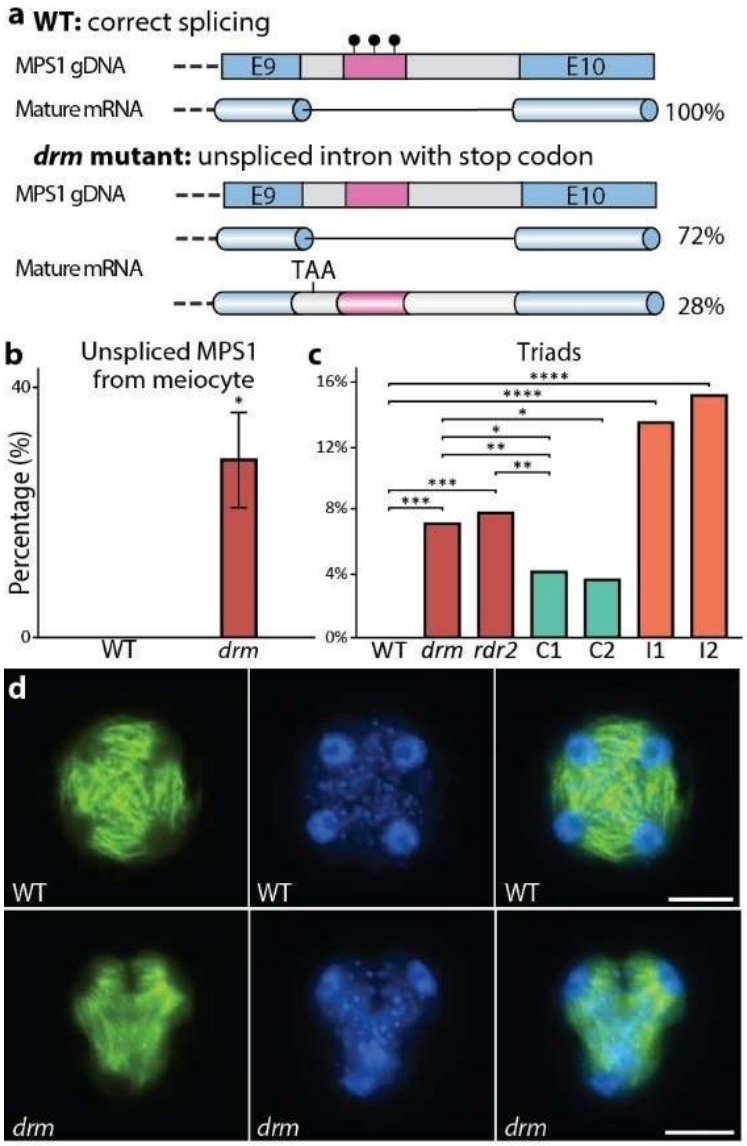


Figure 2.8 RdDM is important for the splicing of MPS1 and normal meiosis. a, Gene model illustrating that the methionine pre-tRNA SLM (magenta bar) located in the last intron of MPS1 affects the splicing of this intron. E, exon; black lollipops, DNA methylation. **b**, Quantitative RT–PCR showing the percentage of unspliced MPS1 transcript in wild-type (WT) and *drm1drm2* (*drm*) double mutant meiocytes. *P < 0.02 (two-tailed t test; n = 3 RNA replicates extracted from independently isolated meiocytes). **c**, Percentage of meiotic triads in wild-type, *drm* and *rdr2* mutants, two complementation lines (C1 and C2) and two interference lines (I1 and I2) (***P < 1 × 10−7; **P < 0.02; *P < 0.05; ****P < 1 × 10−14; two-tailed Fisher's exact test; C1, 514 observations; C2, 167 observations; numbers of observations for other genotypes are listed in Supplementary Figure 2.8. **d**, Spindles (green) and nuclei (blue) of wild-type (tetrad) and *drm* (triad) meiotic products at the tetrad stage. Scale bars, 10 μm.

2.9 Discussion

Our results reveal the presence of a specific DNA methylation signature mediated by the RdDM pathway in the male sexual lineage of *A. thaliana*. SLMs suppress gene transcription and promote the splicing of a gene essential for meiosis and thus are required for normal meiotic progression. This demonstrates that developmental gene regulation through DNA methylation reprogramming is not confined to gamete companion cells in flowering plants and can occur through the deposition as well as the removal of methylation. As RdDM activity appears to be ubiquitous in plant tissues, modulation of the RdDM pathway that achieves cell-specific methylation patterns can plausibly occur in any cell type. The epigenetic regulatory paradigm described here might therefore be broadly applicable to plant development.

SLMs are the product of developmentally orchestrated remodelling of DNA methylation via the RdDM pathway, but the small number of genes directly controlled by SLMs suggests that gene regulation is not the only, and perhaps not the main function of this remodelling. RdDM is known to target transposons (Kim and Zilberman, 2014; Law and Jacobsen, 2010; Matzke and Mosher, 2014), a trade-off has been characterized between methylation of transposons and gene expression (Hollister and Gaut, 2009), and transposon suppression should be particularly important in cells that contribute to the next generation (Feng et al., 2013). RdDM may be balanced more aggressively in sex cells, thus ensuring transposon silencing even at the expense of gene expression. Such a phenomenon might also occur in the shoot apical meristem, which gives rise to all above-ground plant cell types, including the gametes (Baubec et al., 2014). This could explain why we observed sexual-lineage-specific methylation but little if any soma-specific methylation. While it is unclear why specific transposons would be targeted for RdDM in the sex cells (at canonical SLHs), a more aggressive setting of the self versus non-self threshold in sex cells would explain why sexual-lineage-specific RdDM targets genes with such high frequency. Most SLMs may be functionally neutral, or even slightly deleterious, and are likely to be evolutionarily transient, but a few, such as that in MPS1, might be expected to confer a benefit and be retained through selection. While we were unable to detect a difference in meiotic expression for the chromatin genes that have SLMs in *drm1drm2* double mutants, it is possible that the regulation of these genes is more subtle in a manner similar to the finetuning of the demethylase ROS1 by RdDM (Williams et al., 2015), or else expression is influenced by these SLMs at a stage other than meiosis. Although none have been specifically implicated in sexual development (VIM2, VIM3 and

VIM4 have SLMs, but it is VIM5 which is paternally imprinted; Hsieh et al., 2016), it will be interesting to explore the expression of these chromatin genes at various stages of sexual development in RdDM mutants.

The presence of SLMs over tRNA genes is an exciting and challenging phenomenon. Although 42% of methionine tRNA genes overlap an SLM, this includes all copies of the methionine tRNA gene family responsible for the initiation of translation. Given the transcriptional and translational quiescence of pollen until pollen tube initiation (Honys and Twell, 2003; Honys and Twell, 2004), one possibility is that SLMs play a role in reducing the amount of tRNA available. Recently, it has been determined that sRNAs derived from tRNAs or mimicking tRNA sequences are capable of repressing retrotransposon expression in mice and *A. thaliana* pollen (Borges et al., 2018; Martinez, 2017; Schorn et al., 2017). It is possible that tRNA gene methylation is an unavoidable consequence of targeting transposons for repression via RdDM, i.e. sRNAs targeting transposons are also able to target tRNA genes. Alternatively, RdDM targeting of tRNA genes may increase the pool of sRNAs that can recognise transposons in pollen via a cascade – a transposon produces sRNA complementary to tRNA genes which are subsequently targeted by RdDM and produce sRNA via Pol IV to recognise other transposons in the system. Close analysis of RdDM mutant sRNA data would help to elucidate whether such a phenomenon exists.

The substantial number (253) of SLMs that overlapped genes in the genome of *A. thaliana* may elucidate a longstanding mystery regarding plant DNA methylation. The genes of flowering plants frequently exhibit CG-specific methylation of unclear origin and function (Law and Jacobsen, 2010). This methylation has been hypothesized to arise because of transient RdDM activity (Tran et al., 2005), which would have to occur in cells that contribute to the next generation—a description that fits SLMs. The remaining somatic CG methylation at SLMs which is maintained during somatic development without RdDM provides evidence supporting this hypothesis. SLMs only cover a small fraction of the more than 4,000 genes with body methylation (Takuno and Gaut, 2011; Zilberman et al., 2007), thus indicating that most body methylated genes are not presently targeted by RdDM in the male sexual lineage. However, shifting patterns of SLMs over thousands of generations may have plausibly created the existing gene-body methylation pattern owing to the strong transgenerational heritability of CG methylation (Matzke and Mosher, 2014).

How SLMs are targeted by RdDM, i.e. whether they contain remnants of transposons, are recognised by pre-existing sRNAs by another means or are targeted completely *de novo* by an unknown mechanism, is a continued focus of our lab and is examined in Chapter Three. Chapter Three also explores the genes required for specific targeting of RdDM to canonical SLHs and SLMs in *A. thaliana* male sex cells and raises questions about RdDM dynamics in terms of recognising complementary sequences of DNA within the genome.

2.10 Materials and Methods

For clarity, methods detailed below were carried out by myself or otherwise involved work from me. Other experimental procedures are published (Walker et al., 2018).

Sequencing library construction and analysis

Single-end bisulfite sequencing libraries for Illumina sequencing were constructed using the Ovation Ultralow Methyl-Seq Library Systems (Nugen, #0336) and EpiTect Fast Bisulfite Conversion (Qiagen, #59802) kits according to the kit protocols, except the incorporation of two rounds of bisulfite conversion. Bisulfite sequencing data from wild-type microspore (Calarco et al., 2012), sperm (Ibarra et al., 2012), vegetative cell (Ibarra et al., 2012), embryo (Ibarra et al., 2012), and *drm1drm2* double mutant sperm and vegetative cell (Hsieh et al., 2016) were used. Bisulfite sequencing data from 4 wild-type somatic tissues: cauline leaf (Coleman-Derr and Zilberman, 2012), rosette leaf (Coleman-Derr and Zilberman, 2012; Stroud et al., 2014), roots (Coleman-Derr and Zilberman, 2012; Zemach et al., 2013) and seedlings (Zemach et al., 2013), *rdd* mutant rosette leaf (Stroud et al., 2013). *drm*2 and *rdr2* mutant seedlings (Zemach et al., 2013) were obtained from published sources. Bisulfite sequencing data from various root cell types (Kawakatsu et al., 2016), seedlings of *met1*-1, pMET1::MET1 *met1-1* complementation lines (T-MET1a and b) and their wild-type control plants (Catoni et al., 2017) were also used in this study. DNA methylation analysis was performed as described (Ibarra et al., 2012).

Identification of differentially methylated loci between the sexual lineage and somatic tissues

Fractional methylation in 50 bp windows across the genome was compared between an average of selected sex cells (SexAV) and an average of the four somatic tissues (SomAV) (Diff = SexAV - SomAV). CG and CHG methylation averages in sex cells were calculated using meiocytes, microspores and sperm, and CHH methylation average was calculated using microspores and sperm. We first selected windows meeting the following criterion: Diff_CG > 0 & Diff_CHG > 0 & Diff_CHH > 0 & (Diff_CG + Diff_CHG + Diff_CHH) > 0.3. The selected windows were merged to generate larger SLMs if they occurred within 100 bp. Merged SLMs were retained if they covered at least 100 bp, with significantly different levels of total methylation (Fisher's exact test P-value < 0.001), having more methylation in all sex cell replicates than all somatic tissues, and met the following criterion: Diff_CG > 0 & Diff_CHG > 0.05 & Diff_CHH > 0.1 & (Diff_CG + Diff_CHG + Diff_CHH) > 0.4. This resulted in the identification of 1265 SLHs. The same criteria, except reversing the

relationship between sexual lineage and somatic methylation, were used to identify 36 loci hypomethylated in the sexual lineage. SLHs were further refined by the criterion of having significantly (Fisher's exact test P-value < 0.001) less methylation in sex cells (meiocyte, sperm and vegetative cell) of *drm1drm2* double mutant in comparison to those of wild-type, leaving 1257 loci as a refined list of SLHs for further analyses. The refined list of SLHs (1257 loci) was then separated into two groups based on the level of non-CG methylation in somatic tissues: (i) SLMs with CHH and CHG methylation lower than 0.05 and 0.1, respectively, in all 4 somatic tissues (533 loci); (ii) canonical SLHs with CHH methylation higher than 0.05 or CHG methylation higher than 0.1, in any of the 4 somatic tissues (724 loci). Both groups were analysed for overlap with published columella root cap DMRs (a merged list from reported C and CHH/G DMRs; Kawakatsu et al., 2016): 173 and 469 of the canonical SLHs and SLMs have less than 10% overlap, respectively. The non-overlapping 469 SLMs were used in subsequent analyses a refined list of SLMs.

Box plots

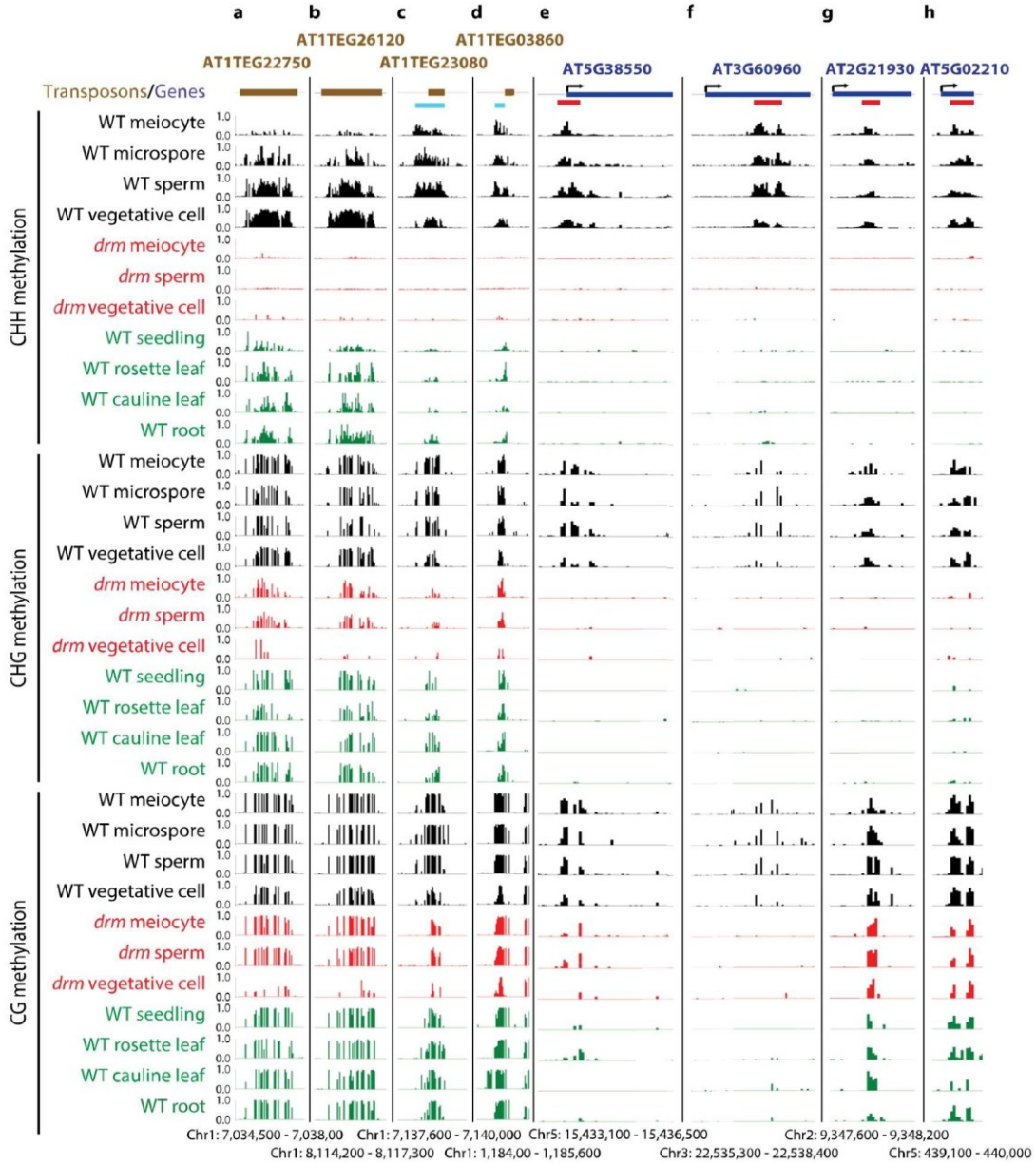
All box plots follow this format: each box encloses the middle 50% of the distribution, with the horizontal line making the median, and vertical lines marking the minimum and maximum values that fall within 1.5 times the height of the box. Figure 2.2b was generated using 50 bp windows with fractional CHH methylation larger than 0.3 in meiocytes compared to rosette leaf, and at least 20 informative sequenced cytosines in each of the 4 somatic tissues and 4 sex cells (meiocyte, microspore, sperm and vegetative cells). Figure 2.3 and Figure 2.4 were generated using 50 bp windows with substantial methylation in either wild-type sex cells or rosette leaf in the corresponding sequence context (cutoff settings: 0.3, 0.3 and 0.1 in CG, CHG and CHH contexts, respectively), and at least 10 informative sequenced cytosines in each replicate of the wild-type and *rdd* rosette leaf sample and 3 sex cells (meiocyte, microspore and sperm).

RT-PCR

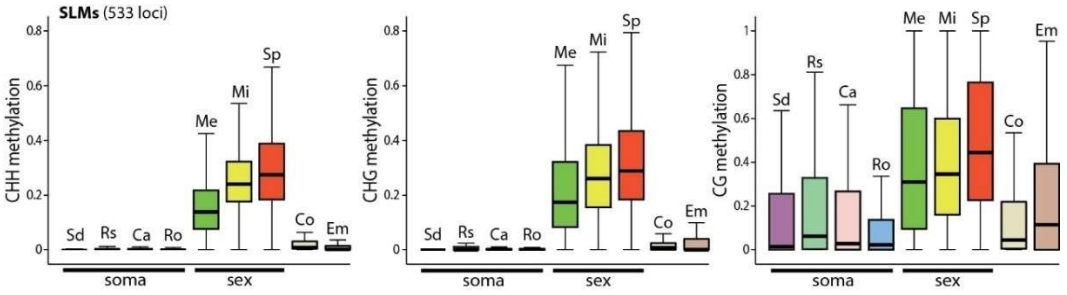
100 ng and 500 ng total RNA was reverse transcribed using RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher scientific, #K1621) for quantitative RT-PCR and RT-PCR, respectively. qRT-PCR was performed using SYBR Green (Roche, #4707516001) in triplicate on the LightCycler 480 Real-Time PCR System (Roche) and Cp values were averaged between 3 technical replicates to determine the target/reference ratio. Figure 2.6c and Figure 2.8b show the averages of at least 3

biological replicates for each genotype or tissue type. RT-PCR was performed with 46 PCR cycles using primers JW134 and JW141 for MPS1, and 30 cycles using primers PHG34 and PHG101 for the control ACTIN8. ACTIN8 was used as internal control in both RT- and qRT- PCRs, and all primers are listed in Supplementary Table 2.7.

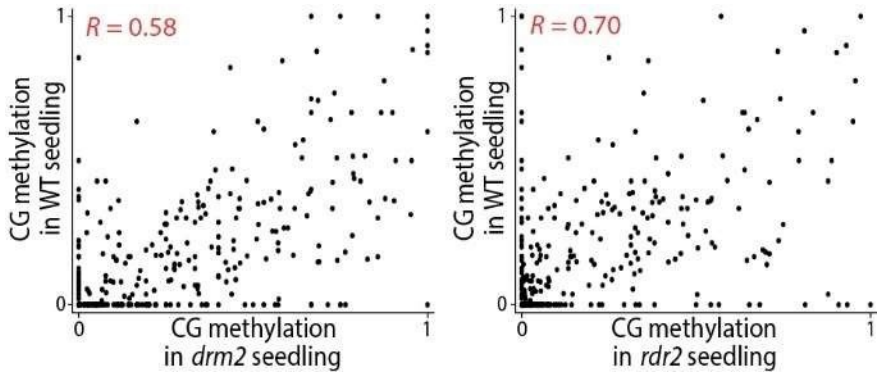
2.11 Supplementary Figures



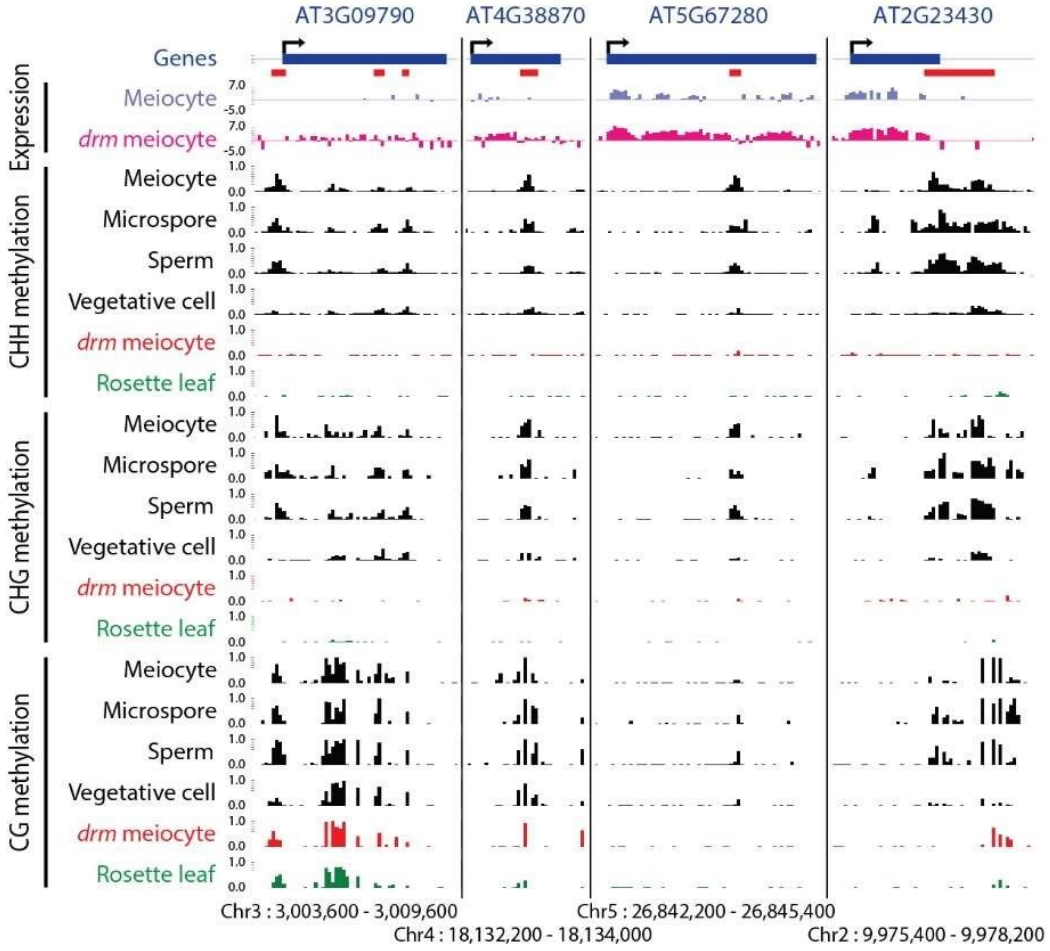
Supplementary Figure 2.1 Examples of canonical RdDM, canonical sexual-lineage-hypermethylated loci (canonical SLHs) and sexual-lineage-methylated loci (SLMs). Examples of typical RdDM (**a**, **b**), canonical SLHs (**c**, **d**, underlined in light blue), and SLMs (**e** – **h**, underlined in red) with remnant CG methylation in *drm1drm2* (*drm*) double mutant sex cells and wild-type (WT) somatic tissues (**g**, **h**). Associated transposons or genes are shown. Refer to Supplementary Table 2.2 for a full list of canonical SLHs and SLMs.



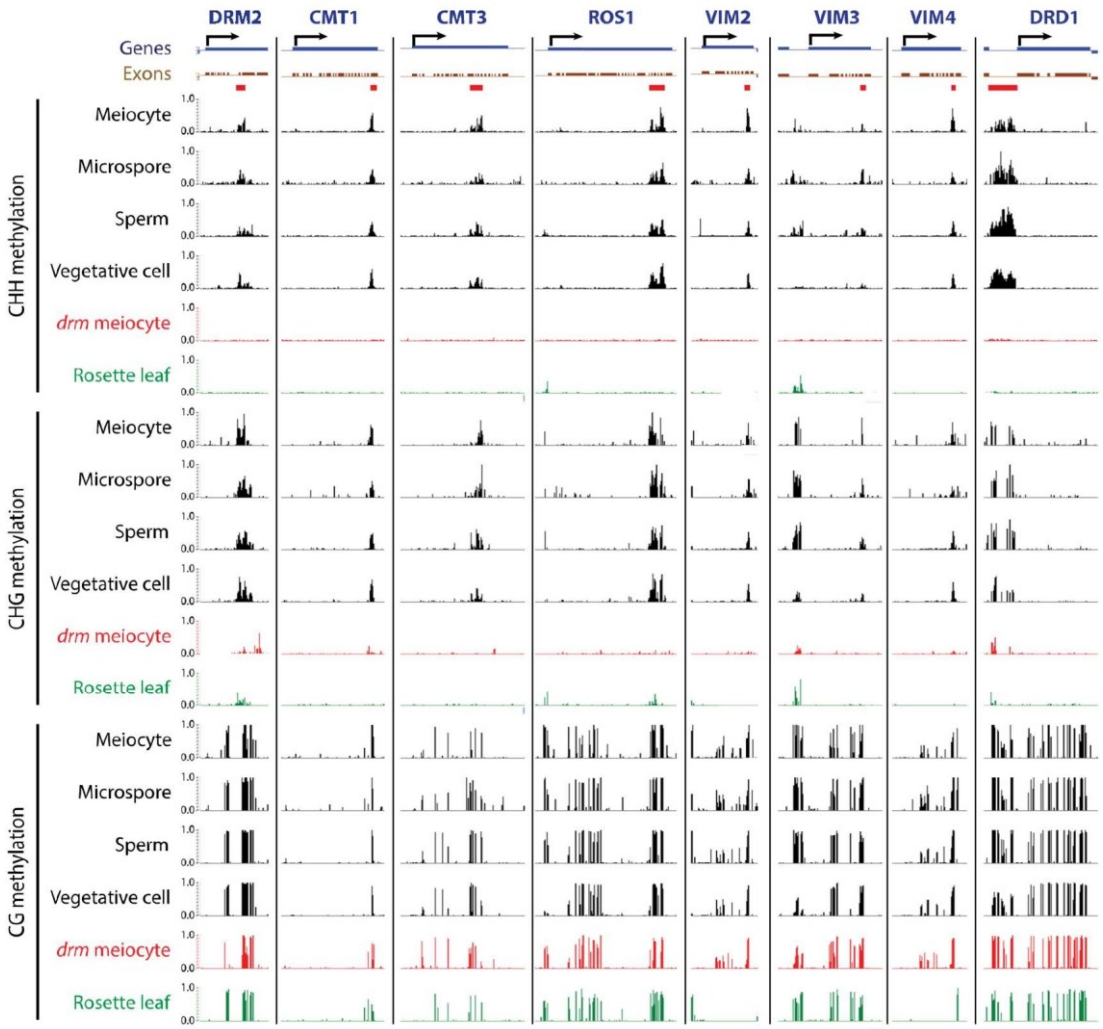
Supplementary Figure 2.2 SLMs have little CHH/G methylation in the columella and embryo. Box plots showing the absolute methylation at SLMs (533 loci before filtering out columella overlaps; see Methods) in somatic tissues (Sd, seedling; Rs, rosette leaf; Ca, cauline leaf; Ro, root), sex cells (Me, meiocyte; Mi, microspore; Sp, sperm), columella root cap (Co) and embryo (Em).



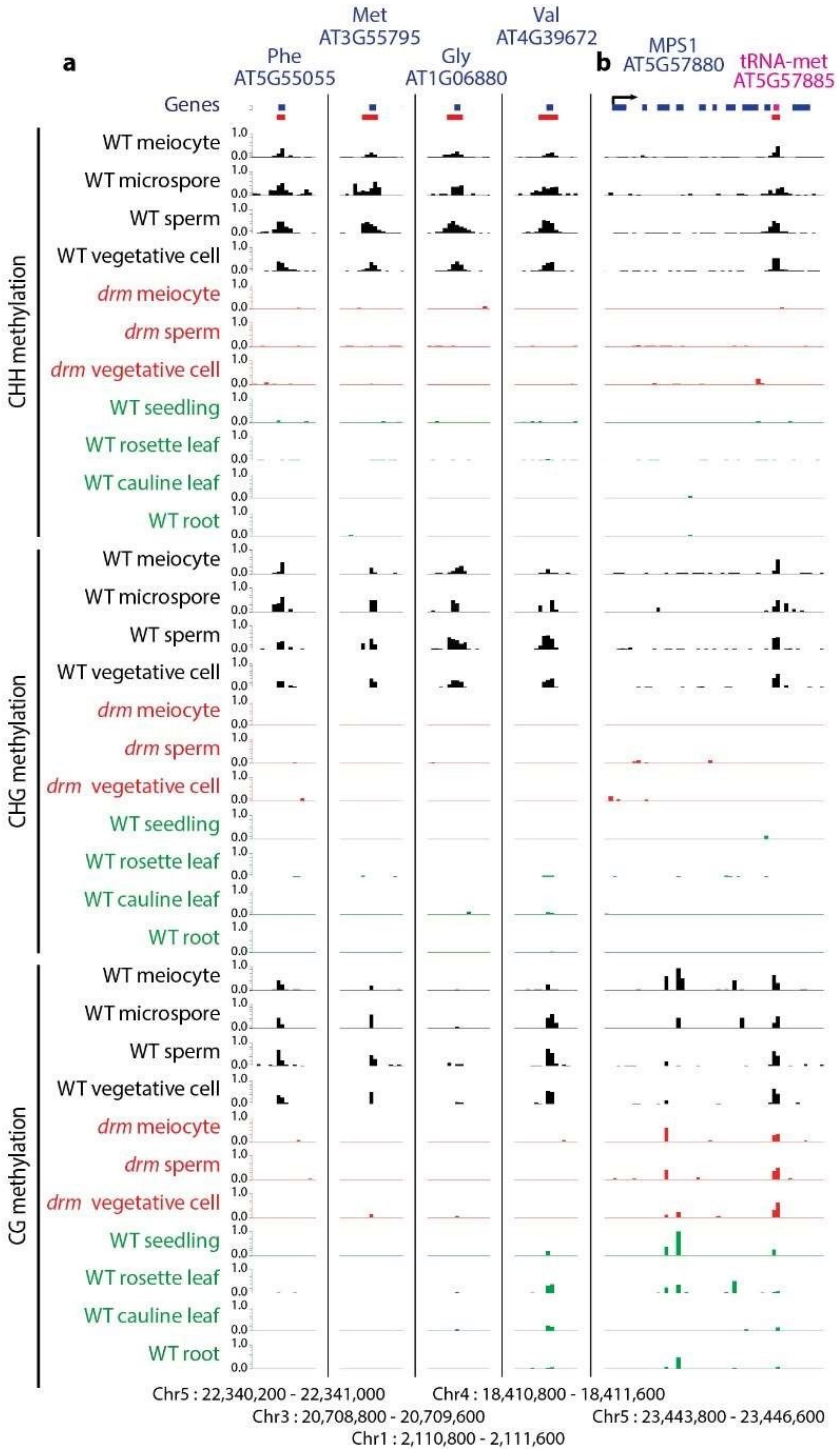
Supplementary Figure 2.3 CG methylation in seedlings of wild-type (WT) and RdDM mutants is strongly correlated. Scatter plots showing linear correlation between CG methylation at SLMs in seedlings of wild-type (WT) and *drm2* (Pearson's R = 0.58), and WT and *rdr2* (Pearson's R = 0.70).



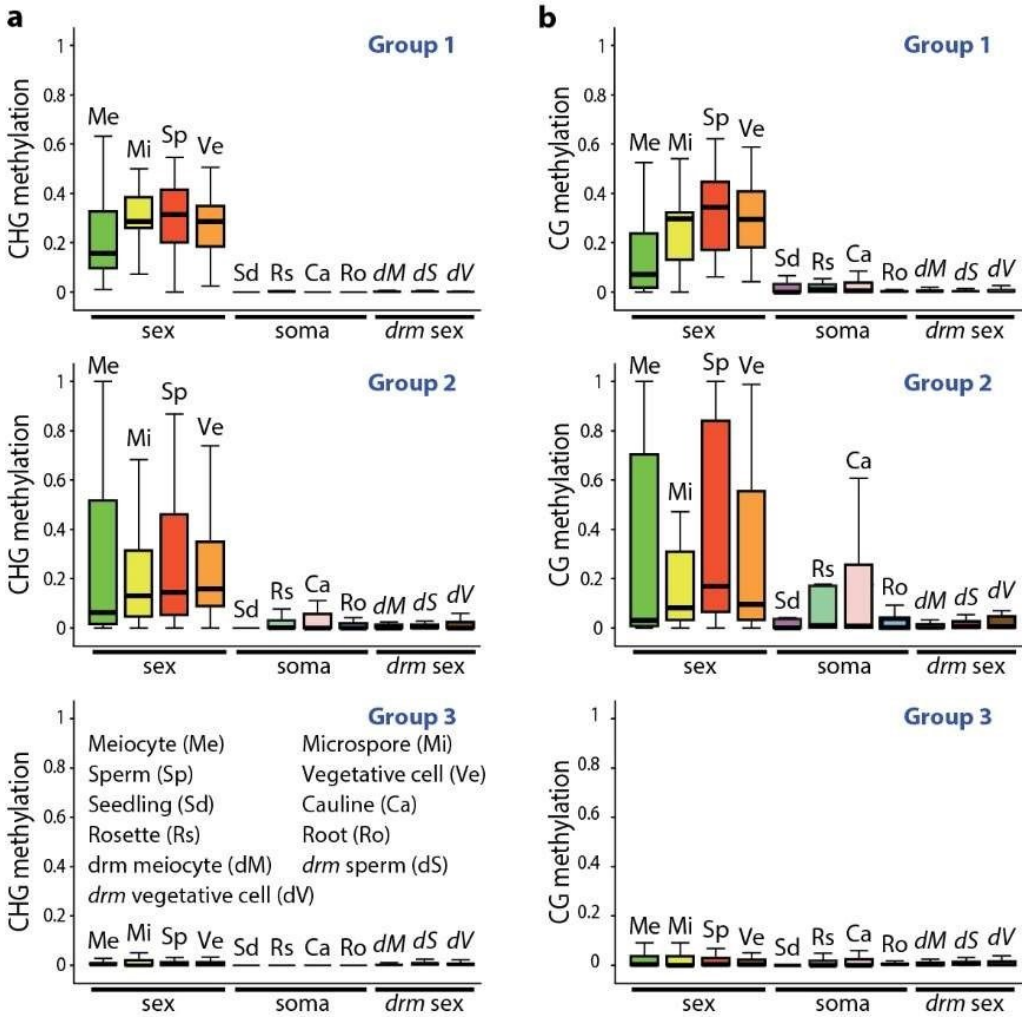
Supplementary Figure 2.4 Examples of genes suppressed by sexual-lineage-specific methylation in meiocytes. Similar to Figure 2.6b, snapshots of cytosine methylation in wild-type male sex cells, *drm1drm2* (*drm*) double mutant meiocyte, and wild-type rosette leaves, and transcriptional expression (in $\log_2$ RPKM) in wildtype and *drm* meiocyte are shown. SLMs are underlined in red.



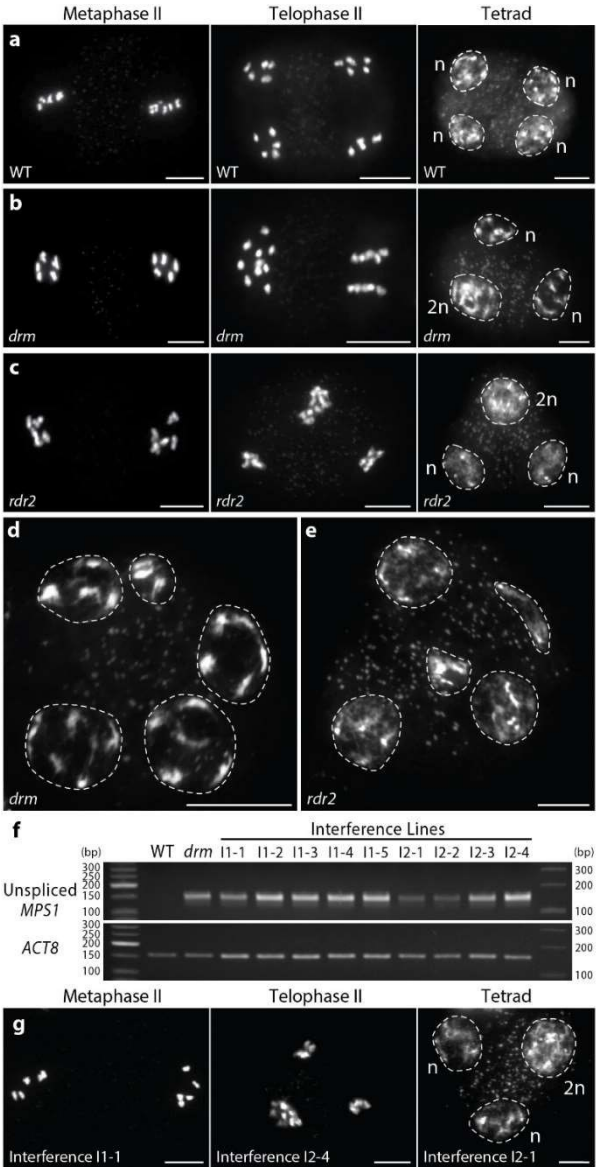
Supplementary Figure 2.5 Snapshots of sexual-lineage-specific methylation associated with genes involved in chromatin modification. Snapshots of cytosine methylation similar to Supplementary Figure 2.1, in wild-type (WT) male sex cells (black), *drm1drm2* (*drm*) double mutant sex cells (red), and four somatic tissues (green). SLMs are underlined in red.



Supplementary Figure 2.6 Examples of sexual-lineage-specific methylation at pre-tRNA genes. Snapshots of cytosine methylation, similar to Supplementary Figure 2.1, in wild-type (WT) male sex cells (black), *drm1drm2* (*drm*) double mutant sex cells (red), and four somatic tissues (green). SLMs are underlined in red. **a**, Examples of SLMs at pre-tRNA genes encoding phenylalanine, methionine, glycine or valine anticodons. **b**, SLM at the methionine pre-tRNA gene (magenta box) located in the last intron of MPS1.



Supplementary Figure 2.7 Pre-tRNA genes are hypermethylated in male sex cells. Similar to Figure 2.7b, these box plots show the absolute CHG (**a**) and CG (**b**) methylation at 3 groups of pre-tRNA genes in sex cells (Me, meiocyte; Mi, microspore; Sp, sperm; Ve, vegetative cell), somatic tissues (Sd, seedling; Rs, rosette leaf; Ca, cauline leaf; Ro, root), and *drm* (*drm1drm2*) mutant sex cells (dM, *drm* meiocyte; dS, *drm* sperm; dV, *drm* vegetative cell). Refer to Figure 2.7 legend for the 3 groups of pre-tRNA genes.



Supplementary Figure 2.8 Meiotic defects in RdDM mutants and MPS1interference lines. a - **e**, Male meiosis II in wild-type (WT; **a**), *drm1drm2* (*drm*; **b**, **d**) and *rdr2* (**c**, **e**) mutants, and the MPS1 interference lines (**g**). All instances of WT male meiosis we observed (301 observations) were normal and lead to tetrads at the end of meiosis II (**a**). However, in 7.1% (380 total observations) and 7.8% (502 total observations) instances of *drm* (**b**) and *rdr2* (**c**) male meiosis, respectively, chromosomes fail to separate, so that triads are observed at telophase II. Occasionally we also observed pentads in *drm* (**d**) and *rdr2* (**e**) mutants. **f**, RT-PCR showing the expression of MPS1 transcript retaining last intron (149 bp) in *drm* mutant and 9 T1 plants of the interference lines, but not in WT. ACT8 as control shows 156 bp bands. **g**, Interference lines exhibit even higher percentages of triads (I1, 13.4%, 463 total observations; I2, 15.1%, 292 total observations). n, the number of chromosomes in the haploid genome. Scale bars, 10 μm.

Chapter Three

Mechanisms underlying sexual-lineage-specific methylation in *Arabidopsis thaliana* male meiocytes

"Jimmy hasn't thrown away his old plates because he's branching out his research into fungi!" – Multiple sources of the Feng lab

Work from this chapter is currently being continued by Sam Deans and Dr. Jincheng Long and work detailed therein is largely considered to be preliminary.

3.1 Abstract

DNA methylation reprogramming is important for the control of gene expression and development in plants and animals. Previously, we identified regions of male sexual-lineage-specific hyper and *de novo* RdDM in the plant *Arabidopsis thaliana* and demonstrated that the *de novo* methylation controls gene expression and splicing important for cell function. Here, we provide evidence that sRNA in the male meiocyte is associated with a particularly small subset of RdDM loci that includes regions of male sexual-lineage-specific hypermethylation. We explore the involvement of RdDM components and suggest that meiotic RdDM is largely due to the chromatin remodeller CLASSY3 which is highly expressed in tapetal nurse cells. Curiously, meiotic RdDM is more associated with the helitron superfamily of DNA transposons and less associated with gypsy retrotransposons than typical RdDM. Helitrons are known to capture fragments of DNA from the genome, and we demonstrate that the gene-targeted *de novo* methylation has mismatched meiotic sRNA as a result of shared sequence complementarity within meiotic RdDM loci. Together, our findings provide an initial mechanistic understanding for the reinforcement of RdDM at a specific subset of loci and the switch to targeting genes in the male sexual lineage.

3.2 Introduction

Transposons are parasitic genetic elements that propagate within host genomes, often causing harm in development due to the disruption of host genes. While transposons can be broadly classified into retrotransposons (Class I) and DNA transposons (class II) that transpose via an RNA or DNA intermediate, respectively, both are targeted for repression via sRNAs in plants by the RdDM pathway (Kasschau et al., 2007; Mosher et al., 2008; Zhang et al., 2007). The mechanisms by which these transposons are first targeted for repression are of current interest in the field, with evidence suggesting that silencing can be initiated in *A. thaliana* through a variation of RdDM that recognises transposon transcripts (Nuthikattu et al., 2013). Recent work has also endeavoured to subdivide the existing sRNA populations associated with transposons in *A. thaliana* by examining multiple sRNA libraries (Hardcastle et al., 2018). Within the context of RdDM, it has been revealed that the four copies of CLASSY (CLSY) chromatin remodellers in *A. thaliana* are responsible for sRNA accumulation (and DNA methylation) at largely distinct loci, with CLSY1 and CLSY2 recruiting POLIV to the majority of RdDM targets (Zhou et al., 2018). However, further work is needed to understand the mechanisms and purpose for such target specificity.

During sexual reproduction, meiosis produces genetically unique gametes through recombination events between paired chromosomes. In the anthers of flowering plants, male meiocytes are connected to a somatic cell layer known as the tapetum by plasmodesmata (Steer, 1977). The tapetum undergoes endoreduplication so that these cells often have two tetraploid nuclei (Weiss-Schneeweiss and Maluszynska, 2001) and this is thought to increase metabolic activities (De Veylder et al., 2011; Shu et al., 2018). As a result, the tapetum can play a supportive role during meiosis and in pollen wall formation (Jiang et al., 2013; Quilichini et al., 2015).

During meiotic recombination, 150 – 200 double strand breaks occur across the *A. thaliana* genome (Xue et al., 2018). Despite this, genomic integrity is retained via a variety of repair mechanisms that result in about ten crossover events (Chelysheva et al., 2007; Higgins et al., 2004; Osman et al., 2011). Conflict between transposons and the host genome is also most prominent during sexual reproduction as proper inheritance of genetic information without mutation allows for healthy offspring and yet this is the avenue by which transposons can survive and propagate within a population (Bestor, 1999). Previously, we generated the methylome of male meiocytes from *A. thaliana* and showed a reinforcement of CG as well as CHG methylation alongside a reduction of CHH methylation across the genome, including

the majority of canonical RdDM loci (Chapter Two; Walker et al, 2018). However, we identified hypermethylation in male meiocytes and other male sex cells compared to somatic tissues at a specific subset of RdDM loci (canonical SLHs) as well as gene-associated *de novo* sexual-lineage-specific methylation (SLMs), both of which are the result of RdDM activity (Chapter Two; Walker et al., 2018). We were curious to explore RdDM in male meiocytes further to determine the mechanisms underlying the perceived specificity of RdDM and the *de novo* methylation in these cells.

3.3 Meiotic sRNAs are associated with a specific subset of RdDM loci

While CHH methylation at canonical SLHs are strikingly distinctive from canonical RdDM loci in meiocytes (Chapter Two; Walker et al., 2018), we were curious to compare their characteristics in somatic tissues more closely to see if they are distinguishable. For the 12,160 canonical RdDM loci (not overlapping canonical SLHs) determined by a reduction of 24 nt sRNA in poliv mutant flower buds compared to wild-type flower buds (Supplementary Table 3.1; Zhou et al., 2018), there is high CG methylation along with moderate CHG and CHH methylation in somatic tissues, while in meiocytes there is low CHH methylation despite increased CG and CHG methylation (Figure 3.1a; Chapter Two; Walker et al., 2018). By contrast, canonical SLHs have lower CG, CHG and CHH methylation than canonical RdDM loci in somatic tissues but gain methylation in all three cytosine contexts in male meiocytes (Figure 3.1b). Canonical SLHs therefore represent a specific group of RdDM targets even within somatic tissues, with a switch in CHH methylation levels between the two groups in meiocytes.

Given the low CHH methylation profile observed at canonical RdDM loci in A. thaliana meiocytes compared to other tissues and cell types, including microspores and sperm (Chapter Two; Walker et al., 2018), we were interested in examining the meiotic sRNA profile. Dr. Hongbo Gao therefore sequenced sRNA from isolated male meiotic cells and I determined 37,729 24 nt sRNA clusters using Shortstack analysis (Supplementary Table 3.2). A significant portion of these clusters overlap with genes (29,474; 78%; Supplementary Table 3.2). Recent work has also found a similar enrichment of genes for *A. thaliana* meiotic sRNA, and the authors suggest a functional role in recombination (Huang et al., 2019). However, our genic sRNA may instead be due to degradation. Indeed, we find that 32,217 (85%) of our clusters show a strand bias (Supplementary Table 3.2) and that 98% of the cluster regions with a strand bias overlapping genes share the same strandedness between the cluster and the gene, suggesting mRNA origin. Furthermore, 94% of the genic sRNAs map solely to exons rather than the entire gene. This, along with the

correlation between sRNA and meiocyte gene expression (Huang et al., 2019), indicate an issue with mRNA degradation likely due to sample preparation from these cells which are difficult to isolate. Regardless, we decided to carry out a preliminary analysis as 68% of 24 nt sRNAs (despite only comprising 8,255 24 nt clusters) in our current sample do not overlap genes.

For the complete set of 12,771 flower bud RdDM loci (Zhou et al., 2018), 11,760 (92%) have at least 2-fold less 24 nt sRNA in meiocytes than in wild-type flower buds and are hereafter referred to as typical RdDM loci (Supplementary Table 3.1). Indeed, 8672 (74%) of typical RdDM loci have no detectable meiotic 24 nt sRNA cluster (Supplementary Table 3.1). As a result, the typical RdDM 24 nt sRNA profile resembles poliv mutant flower buds and constitutes only 3% of the 24 nt sRNA pool in meiocytes compared to 69% in wild-type flower buds (Figure 3.1c; Figure 3.1d). Although this meiotic sRNA profile may reflect mRNA degradation (miRNA levels are also reduced in our meiotic dataset compared to shoot, flower bud and pollen and a smaller proportion of 21 nt sRNA map to miRNA genes in relation to all genes; Supplementary Figure 3.1), the profile does correlate with low meiotic CHH methylation at RdDM loci in male meiocytes and a similar trend of reduced RdDM sRNA in *A. thaliana* meiocytes has been independently observed (Huang et al., 2019).

Of the remaining 1,011 RdDM loci with comparable 24 nt sRNA levels in the meiocyte to wild-type flower buds (hereafter referred to as meiotic RdDM loci), we find that 437 overlap canonical SLHs (constituting 445 of the 724 canonical SLHs; 61%; Supplementary Table 3.1). Indeed, 501 (69%) out of the 724 canonical SLHs overlap meiotic 24 nt sRNA clusters (Supplementary Table 3.1). As a result, 45% of the meiotic 24 nt sRNA pool maps to canonical SLHs compared to 16% in flower buds (Figure 3.1c; Figure 3.1d). The sizes of meiotic sRNA overlapping canonical SLHs are typical for RdDM (i.e. 21 – 24 nt; Figure 3.1e), suggesting that this sRNA is not due to degradation. The 574 meiotic RdDM loci not overlapping canonical SLHs constitute a further 23% of the meiotic 24 nt sRNA pool compared to 6% in flower buds (Figure 3.1c; Supplementary Table 3.1). These loci partially resemble canonical SLHs in the soma, with reduced CHH methylation compared to typical RdDM (Figure 3.1f). While typical RdDM loci lose CHH methylation and canonical SLHs gain methylation in meiocytes, these other meiotic RdDM loci retain the same levels of CHH methylation (Figure 3.1f). Together, our data suggests a concentration of sRNA and RdDM activity in *A. thaliana* meiocytes at a specific subset of loci that includes previously identified canonical SLHs.

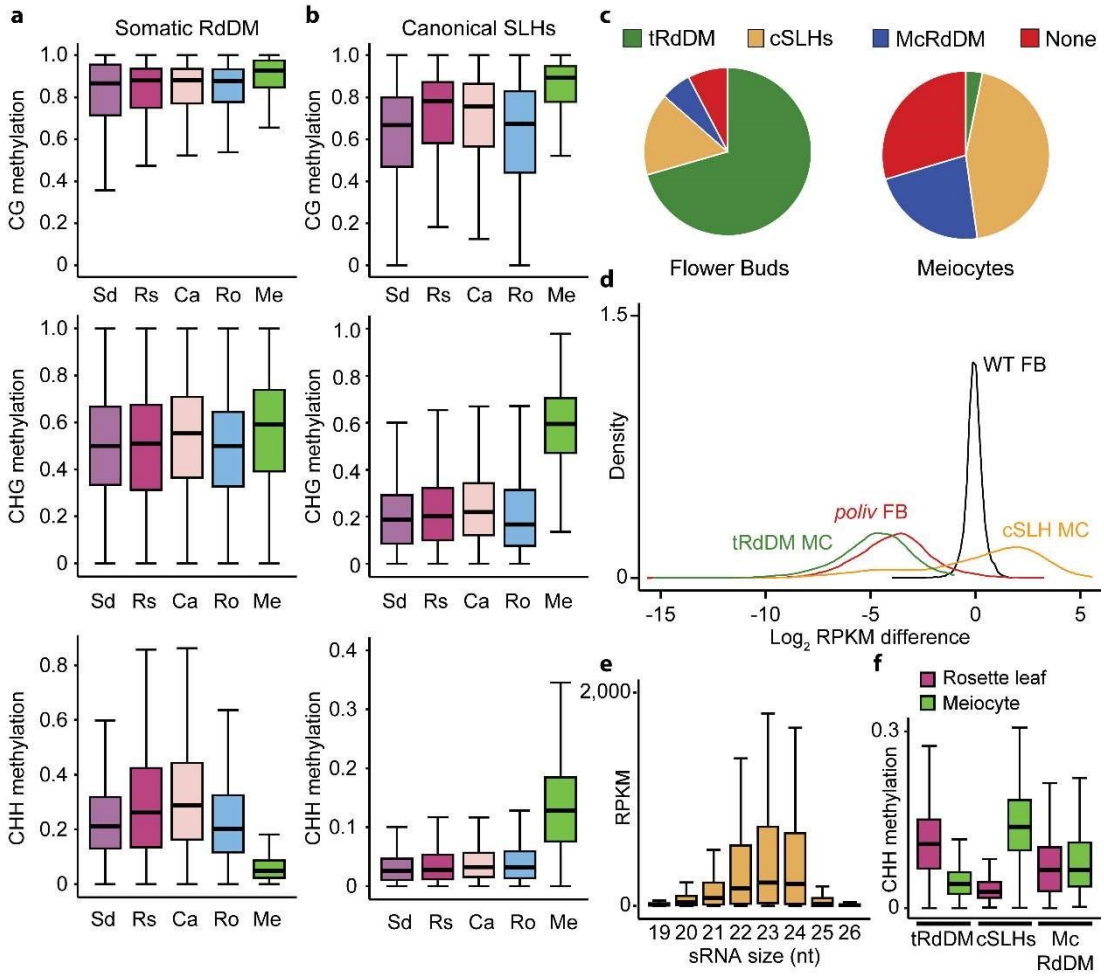


Figure 3.1 RdDM in *A. thaliana* male meiocytes is highly specific. **a**, **b**, Box plots showing absolute methylation at (**a**) typical RdDM loci and (**b**) canonical SLHs in wild-type seedling (Sd), rosette leaf (Rs), cauline leaf (Ca), root (Ro), and meiocyte (Me). **c**, Pie charts showing the association of flower bud or meiotic 24 nt sRNA with typical RdDM loci (tRdDM), canonical SLHs (cSLHs), other meiotic RdDM loci (McRdDM), or none of these. **d**, Kernel density plots trace the frequency distribution of 24 nt sRNA (log_2 FPKM) differences between two wild-type flower bud replicates at all RdDM loci (black; WT FB), wild-type flower buds and poliv flower buds at all RdDM loci (red; poliv FB), wild-type flower buds and meiocyte at typical RdDM loci (blue; tRdDM MC), and between wild-type flower buds and meiocytes at canonical SLHs (yellow; cSLHs MC). **e**, The meiotic sRNA size distribution for canonical SLHs (log_2 FPKM). **f**, CHH methylation for typical RdDM loci, canonical SLHs (as in Figure 3.1b), and other meiotic RdDM loci for rosette leaf or meiocyte.

3.4 Specific CLSY expression is likely the cause of meiotic RdDM

We wondered whether any expression changes for known RdDM components in the meiocyte (Chapter Two; Walker et al., 2018) or the surrounding tapetum (Aldridge et al., unpublished) could explain the observed specificity of RdDM in *A. thaliana* meiocytes. Several genes are expressed more highly in meiocytes compared to seedlings and rosette leaves, such as DRM1 and a DRD1 homolog (Supplementary Table 3.3). Most notable, however, is the increased expression of CLSY3 in the tapetum and CLSY4 in meiocytes (Figure 3.2a; Supplementary Table 3.1). This, and the recent demonstration that each CLSY controls largely unique subsets of RdDM loci (Zhou et al., 2018) led us to investigate the involvement of CLSYs in meiotic RdDM further.

For the 724 canonical SLHs, we find that 363 (50%) overlap clsy3-dependent flower bud sRNA clusters, while the clsy1, clsy2, and clsy4-dependent clusters combined only account for 71 (10%; Figure 3.2b; Supplementary Table 3.1). Accordingly, flower bud CHH methylation at canonical SLHs is reduced in clsy3 single mutants and not in clsy1clsy2 double mutants or clsy4 single mutants (Figure 3.2c, d and e). In line with these findings and the high expression of CLSY3, canonical SLHs are also hypermethylated in the tapetum (Aldridge et al., unpublished). A similar association of clsy3-dependent clusters with the remaining meiotic RdDM loci is also observed (Supplementary Figure 3.2), so that all meiotic RdDM loci overlap with 506 out of 804 (63%) clsy3-dependent clusters, compared to 54 out of 1,792 (3%) clsy1dependent clusters, 10 out of 45 (22%) clsy2-dependent clusters, and 75 out of 777 (9%) clsy4-dependent clusters (Supplementary Table 3.1). CHH methylation of clsy3-dependent clusters in wild-type flower buds is lower than typical RdDM loci (Zhou et al., 2018). This fits the profile of canonical SLHs which also have lower methylation levels than typical RdDM loci in the soma (Figure 3.1b). Thus, our data suggests that CLSY3 contributes towards meiotic RdDM.

As CLSY3 expression is low compared to CLSY1 in meiocytes but high in the tapetum (Figure 3.2a; Supplementary Table 3.1), there is an intriguing possibility that sRNA is produced at meiotic RdDM loci in tapetal cells and then transported into meiocytes to direct RdDM, in the same way that the supporting vegetative cell provides sRNA for the sperm cell (Calarco et al., 2012; Ibarra et al., 2012). Preliminary meiocyte methylation data from a line expressing RDR2 under a tapetum specific promoter in a *rdr2* mutant suggests that this may be the case (Aldridge, unpublished), and others in the lab are working to establish the importance of tapetal sRNA in sexual development.

Despite the increase of expression in the meiocyte (Figure 3.2a), CLSY4 does not seem to be solely responsible for many meiotic RdDM loci (Supplementary Table 3.1). However, clsy3clsy4-dependent clusters do overlap an additional 42 canonical SLHs, 56 of the other meiotic RdDM loci, and lose more CHH methylation at canonical SLHs in flower buds than *clsy3* single mutants so that CHH methylation in *clsy3clsy4* double mutants resemble poliv mutants at canonical SLHs (Figure 3.2f, g; Supplementary Table 3.1), suggesting that CLSY4 (which is highly expressed in meiocytes) may contribute towards the meiotic RdDM profile. A remaining 207 canonical SLHs overlap with clusters determined from quadruple *clsy* mutants (Supplementary Table 3.3), indicating that CLY3 and/or CLSY4 also target RdDM to these additional loci. Examining CLSY protein levels in male sexual development and exploring the methylation as well as sRNA profiles of single and double *clsy* mutant meiocytes will be instrumental in clarifying this situation.

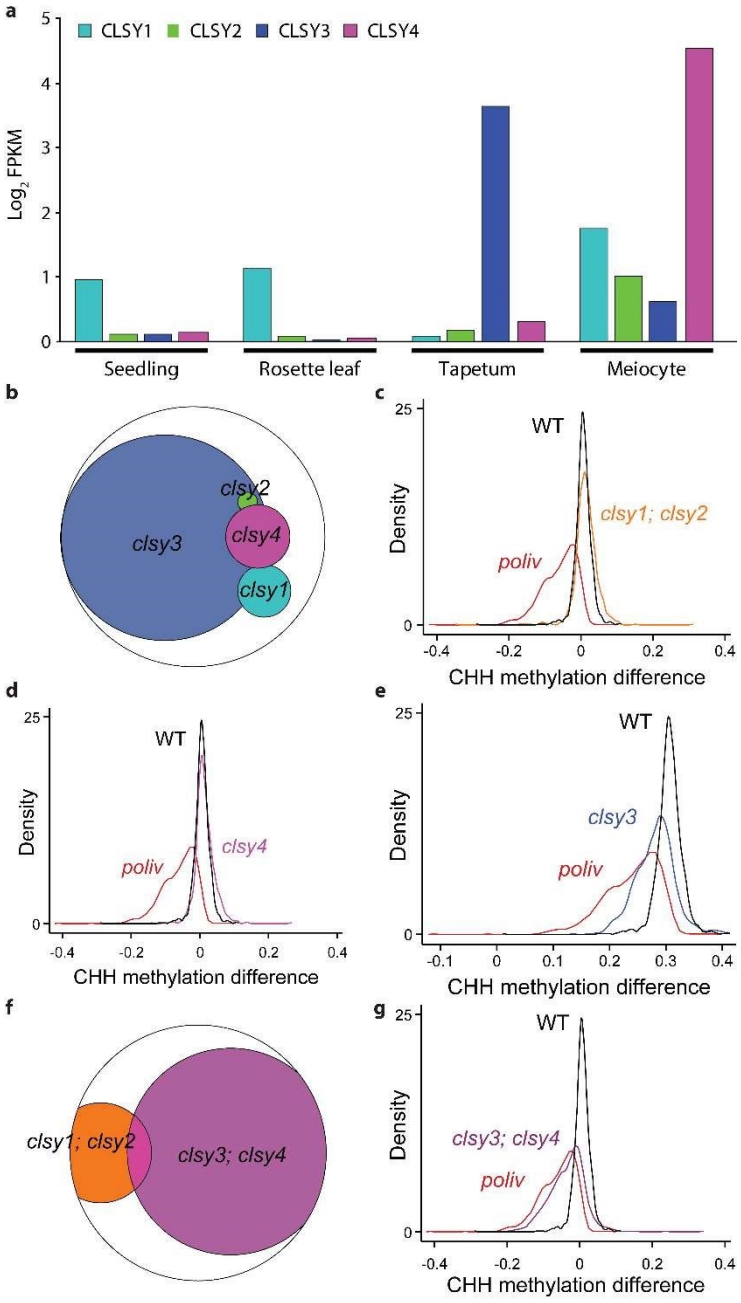


Figure 3.2 CLSY involvement in meiotic RdDM. a, Expression of the four *A. thaliana* CLSY genes in seedlings, rosette leaf, tapetum, and meiocyte (log_2 FPKM). **b** and **f**, Scaled Venn diagrams showing the association of canonical SLHs with reduced 24 nt sRNA clusters for *clsy1*, *clsy2*, *clsy3* and *clsy4* single mutants, *clsy1clsy2* and *clsy3clsy4* double mutants and *clsy* quadruple mutants. **c**, **d**, **e**, **g**, Kernel density plots trace the frequency distribution of flower bud CHH methylation differences at canonical SLHs between two wild-type replicates (black; **c**, **d**, **e**, **g**), or between wild-type and *clsy1clsy2* double mutants (orange, **c**), *clsy4* single mutants (magenta, **d**), *clsy3* single mutants (blue, **e**), *clsy3clsy4* double mutants (purple, **g**) and *poliv* quadruple mutants (red, **c**, **d**, **e**, **g**).

3.5 Meiotic RdDM is associated with helitrons but not gypsy elements

RdDM is typically associated with all classes of transposons in *A. thaliana* (Figure 3.3; Mosher et al., 2008; Zhang et al., 2007). Although canonical SLHs are not as strongly associated with transposons as canonical RdDM (Chapter Two; Walker et al., 2018), I wondered whether canonical SLHs and other meiotic RdDM loci showed any transposon preference. We found that canonical SLHs are enriched for helitrons, with these elements comprising 66% (515 out of 779) of the canonical SLH associated transposons compared to 33% (5677 out of 17,408) of typical RdDM loci and 42% (12,966 out of 31,225) of all transposons in the genome ($P < 0.0001$ for both comparisons; Fisher's exact test; Figure 3.3; Supplementary Table 3.4). In addition to the enrichment for helitrons, gypsy elements (retrotransposons) are absent from canonical SLHs, comprising only 1% of associated transposons compared to 21% for typical RdDM loci ($P < 0.0001$; Fisher's exact test; Figure 3.3; Supplementary Table 3.4). Other meiotic RdDM loci are also partially enriched for helitrons and have proportionally fewer gypsy elements ($P < 0.0001$ for both comparisons; Fisher's exact test; Figure 3.3; Supplementary Table 3.4). However, there is no enrichment of any specific helitron families for meiotic RdDM (Supplementary Figure 3.3; Supplementary Table 3.5). Only 14% of RdDM-associated helitrons are meiotically targeted (Supplementary Table 3.4) and so further work is currently underway to determine the age of transposons associated with meiotic RdDM to further understand the specificity observed.

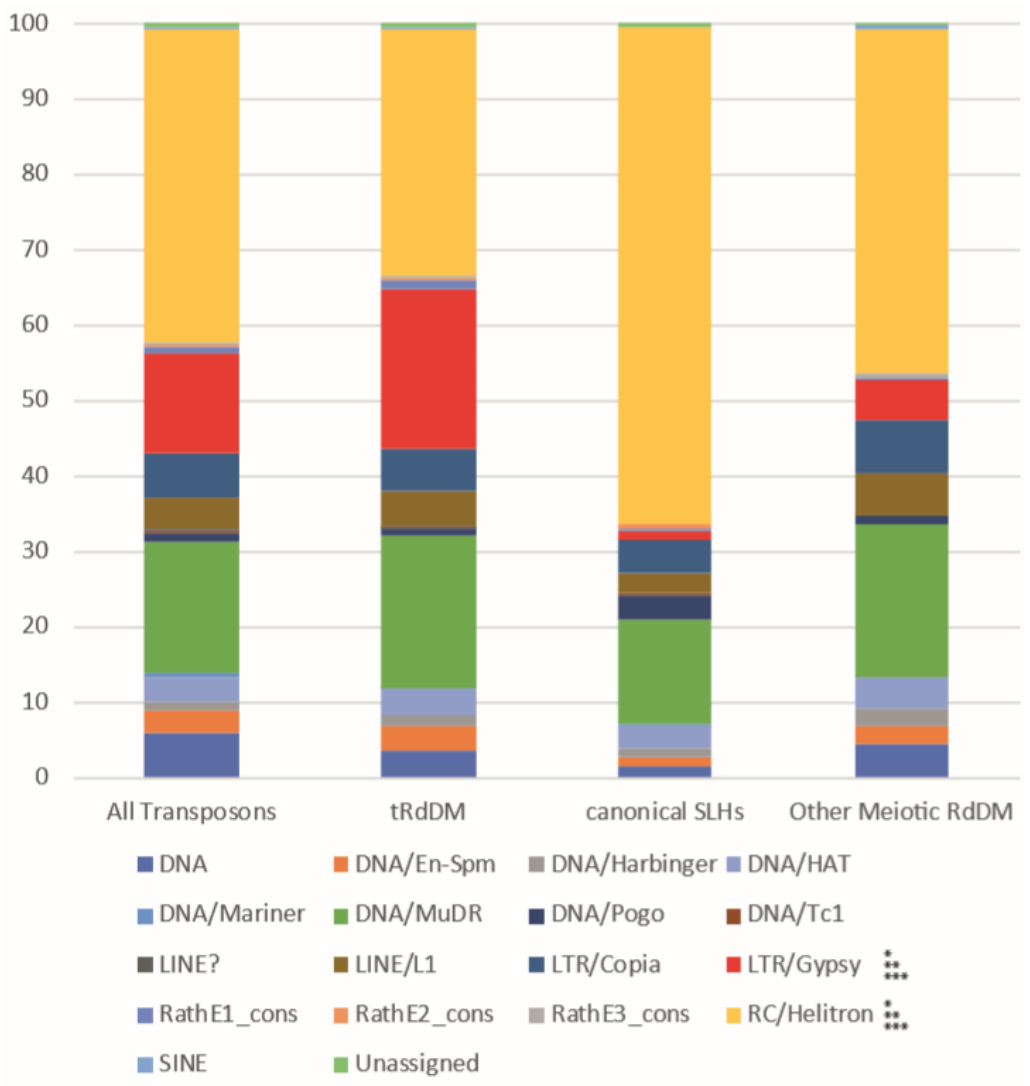
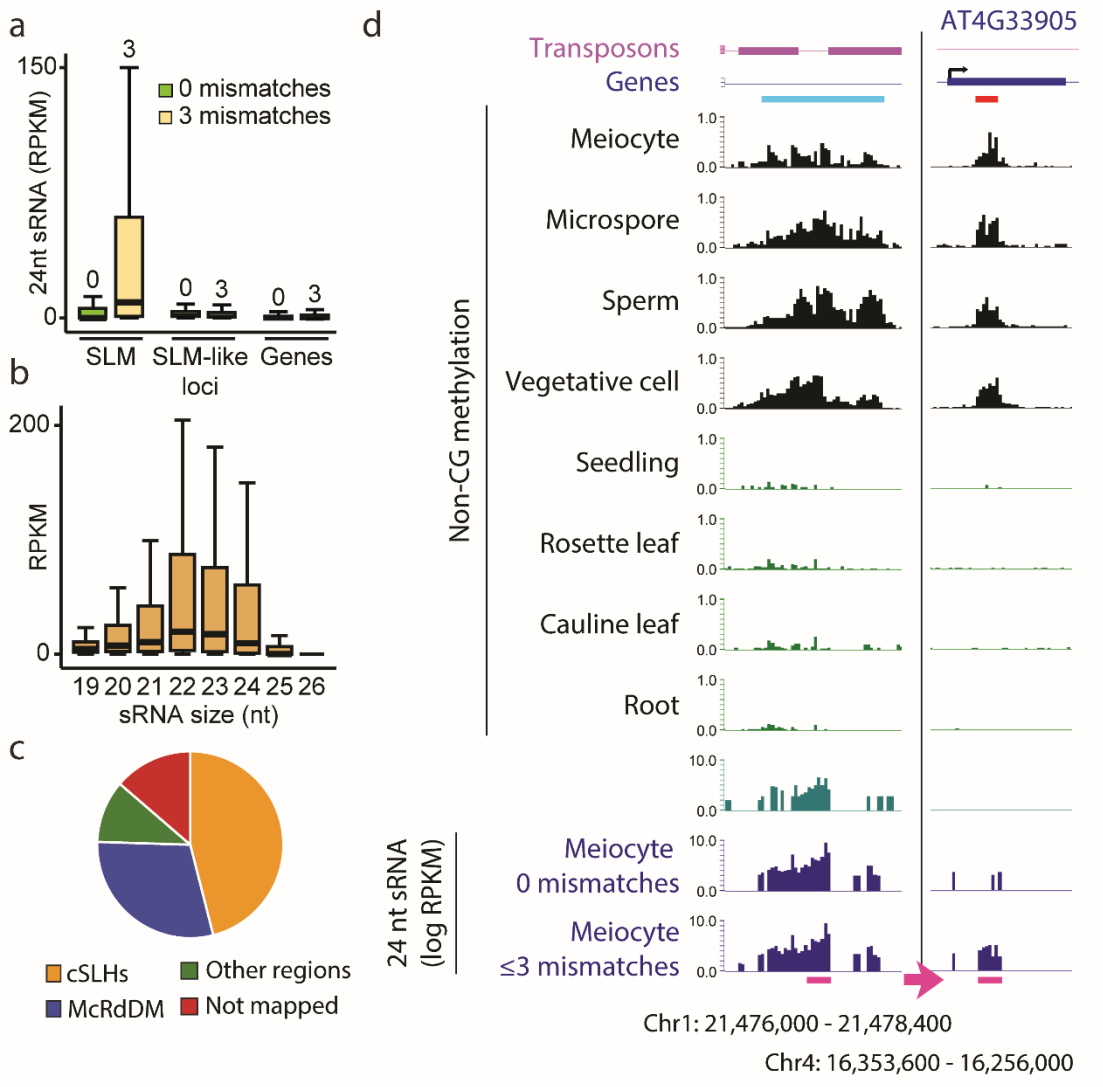


Figure 3.3 Meiotic RdDM association with transposons. Bar graphs showing the proportions of transposon families across the TAIR10 genome, and transposon families associated with typical RdDM loci (tRdDM), canonical SLHs, and other meiotic RdDM loci. * Significant difference between tRdDM and canonical SLHs, ** significant difference between all transposons and canonical SLHs, *** significant difference between tRdDM and other meiotic RdDM ($P < 0.0001$ for all comparisons; Fisher's exact test).

3.6 SLM sRNAs originate from meiotic RdDM loci

While meiotic sRNA map well to canonical SLHs under typical parameters, recording all the best hits for sRNA against the genome (with weighting) and allowing zero mismatches, sRNAs at the 469 SLMs are barely detectable (Figure 3.4a). Once the parameters were relaxed to allow up to three mismatches, however, SLM loci gain meiotic 24 nt sRNA compared to SLM-like loci and genes (Figure 3.4a), with sizes ranging from 21 - 24 nt as with other meiotic RdDM (Figure 3.4b). The presence of mismatches indicates either that these sRNA are artefacts within the data or that they originate from elsewhere in the genome where they map perfectly. In agreement with the latter, meiotic 24 nt sRNA fail to map against SLMs when parameters allow for up to three mismatches but only the best hits in the genome are recorded (data not shown). Indeed, 86% of mismatched 21 – 24 nt SLM sRNA map perfectly elsewhere in the genome, with 46% mapping to canonical SLHs and 30% mapping to other meiotic RdDM loci (Figure 3.4c). I created a network detailing the association between SLMs and meiotic 24 nt sRNA clusters, constituting 295 out of the 469 SLMs, with an average of one meiotic 24 nt sRNA cluster contributing to one SLM (Figure 3.4c-e; Supplementary Table 3.5). Interestingly, we found that the regions sharing sRNA between meiotic 24 nt sRNA clusters and SLMs could get as large as 340 bp, indicating a shared origin of genomic sequence. It has recently been reported that sRNA involved in RdDM can act with mismatches in a similar manner to miRNA (Burghgraeve et al., 2018). Here, we suggest that SLMs are targeted for methylation in meiocytes in trans from meiotic RdDM loci and then SLMs produce their own sRNA in pollen (Walker et al., 2018). Work is currently underway to delete specific meiotic RdDM sources and examine male sex cell methylation at the associated SLMs to provide further evidence for this link.



e

SLM_#	Gene_association	Distance_to_gene	SLM_location	SLM_sRNA_location			SLM_sRNA_location_size
SLM352	AT4G33905	0	chr4:16254846-16255195	chr4	16254862	16255141	279

SLM_sRNA_origin			origin_sRNA_size	origin_sRNA_abundance	origin_cluster	origin_cluster_location	origin_SLH	origin_SLH_location
chr1	21477121	21477384	263	122	clustermerge6499	chr1:21476263-21477465	SLH136	CHR1:21476286-21478235

Figure 3.4 Mismatched SLM sRNA originate from meiotic RdDM loci. a, 24 nt sRNA from meiocytes do not map well to SLMs with zero mismatches but do with up to three mismatches. A similar gain of sRNA is not seen for randomly generated SLM-like loci or genes. **b**, The size distribution for meiotic sRNA with up to three mismatches mapped to SLMs (log_2 FPKM). **c**, Pie chart showing the associated features for SLM mismatched 21 – 24 nt sRNA that have been mapped to an SLM-masked TAIR10 genome with zero mismatches. **d**, Example of a canonical SLH (underlined in light blue) that has zero mismatched 24 nt sRNA with an overlapping region (underlined in magenta) to an SLM that has corresponding mismatched sRNA. **e**, Excerpt from Supplementary Table 3.5 showing an SLM and the corresponding sRNA origin location and associated canonical SLH.

3.7 Discussion

The genome of the sexual-lineage is passed on to the next generation and thus protection from mutation is critical. Transposons are also more likely to propagate throughout a host population by transposing in the sexual-lineage, and so a conflict exists and drives the evolution of more aggressive transposons and host defence mechanisms during this period of development (Bestor, 1999). Perhaps paradoxically, RdDM (a mechanism important for transposon repression; He et al., 2011; Law and Jacobsen, 2010) appeared to be reduced at the majority of targets in the male meiocytes of *A. thaliana*, as indicated by a loss of CHH methylation (Chapter Two; Walker et al., 2018). Although we require replicates of meiotic sRNA libraries, and a clearer definition of meiotic RdDM needs to be determined by measuring reduced sRNA in poliv mutant meiocytes, my preliminary analysis provides interesting insights into meiotic RdDM. I show that the male meiocyte 24 nt sRNA profile correlates with meiotic methylation, confirming a relative reduction of typical RdDM activity in these cells. A small subset of RdDM loci retain or show increased levels of 24 nt sRNA in the meiocyte, however, and these loci are enriched for the previously identified loci that are hypermethylated in male sex cells due to RdDM activity (canonical SLHs; Chapter Two; Walker et al., 2018). I provide evidence that the specificity of meiotic RdDM is due at least in part to CLSY3 which is highly expressed in the supporting tapetal cells, indicating these somatic cells are a potential source of sRNA for sexual-lineage-specific methylation in a similar manner to the vegetative cell supporting transposon repression in sperm cells (Calarco et al., 2012; Ibarra et al., 2012). I also demonstrate that canonical SLHs and meiotic RdDM are more associated with helitrons than typical RdDM. Helitrons are known to capture fragments of DNA including genes (Barbaglia et al., 2012; Hollister and Gaut, 2007), and I show that sRNA from meiotic RdDM loci can map to gene-targeted *de novo* methylated loci with a small number of mismatches due to shared sequence complementarity. Together, this work highlights the specificity of RdDM in meiotic cells and provides a mechanism by which *de novo* methylation can be reliably achieved by RdDM at specific loci during sexual development.

While the reinforcement of CG and CHG methylation by other mechanisms could be expected to ensure silencing of transposons in the absence of CHH methylation, the purpose of specialised RdDM in meiocytes remains to be seen. It could play a role in meiosis such as in meiotic recombination, where other DNA methylation mechanisms have been implicated to play a role (Underwood et al., 2018; Yelina et al., 2015). Indeed, AGO mutants have been shown to have defective recombination

(Oliver et al., 2016). However, it is difficult to determine the underlying cause of these issues due to the requirement of RdDM for MPS1 activity (Chapter Two; Walker et al., 2018). The enrichment of canonical SLHs and other meiotic RdDM loci for helitrons may be in some way linked to the double strand breaks that occur during the recombination process (Xue et al., 2018), as helitrons are thought to utilise double strand breaks for transposition (Kapitonov and Jurka, 2007). Exploring transposon expression in meiocytes and tapetum, as well as transposition in pollen, for both wild-type and various RdDM mutants will hopefully allow further conclusions to be made.

The link between the association of CLSY3 with meiotic RdDM and the high expression of CLSY3 in the tapetum is yet to be determined. Indeed, meiotic RdDM may be an indirect consequence of tapetal RdDM activity essential for tapetal development, and evidence of AGO-driven tapetal development has been reported in rice (Zheng et al., 2019). Tapetal cells undergo endoreduplication and are metabolically active (De Veylder et al., 2011; Shu et al., 2018). In turn, these cells are particularly sensitive to abiotic stress (Parish et al., 2012) and RdDM activity may help to repress particularly aggressive transposons that could disrupt tapetal development or else be passed on to the meiocytes through plasmodesmata. Alternatively, endoreduplication in the tapetum allows for more sRNA to be produced at the limited number of meiotic RdDM loci by producing more DNA templates for transcription (as is thought to occur with typical transcription in endoreduplicating cells; Larkins et al., 2001), and thus the tapetum may act as the sRNA powerhouse for RdDM targeting in meiocytes. Exploring these possibilities (for example in endoreduplication mutants) will add to our understanding of chromatin regulation during sexual reproduction.

The presence of SLMs as a by-product of RdDM at canonical loci is an attractive model to explain the establishment of *de novo* methylation for a typically self-reinforcing pathway. The often-large size of sequence complementarity between SLMs and meiotic RdDM loci would suggest that at least some SLMs are not targeted for RdDM by chance from the existing sRNA pool (i.e. by a single sRNA that attains sequence complementarity through SNP mutation), but rather suggests that these sequences share a genomic origin. Given the enrichment of SLHs for helitrons, and the known ability for helitrons to capture fragments of genes (Barbaglia et al., 2012; Hollister and Gaut, 2007), we hypothesise that gene-capture by transposons targeted by meiotic RdDM results in the production of sRNA from the captured fragment that is then able to target the endogenous location for RdDM.

Confirmation of the link between meiotic RdDM and SLMs by examining the methylation of specific SLMs in the respective meiotic RdDM CRISPR knockout mutants, as well as the examination of the gene-fragment capture phenomenon by exploring the genomes of *A. thaliana* accessions and other Brassica species, are currently being carried out.

While SLMs are observed in the male sexual lineage, the sRNA of meiotic RdDM loci are present in other tissues, and hypermethylation of canonical SLHs is seen in specific tissues such as the columella and embryo (Chapter Two; Walker et al., 2018). While SLM existence in meiocytes may in part be explained by the sheer abundance of sRNA from meiotic RdDM loci (which may meet a threshold that allows mismatch targeting), SLMs are not present in the tapetum despite canonical SLH hypermethylation and CLSY3 expression (Aldridge, unpublished). The existence of SLMs in meiocytes could also be explained by the involvement of other RdDM components that allow relaxed targeting in meiocytes, which in turn may be important in increasing detection of deleterious transposons. Examining the involvement of other RdDM components that are upregulated in meiocytes may reveal the underlying cause of mismatch-targeted RdDM. However, it may be difficult to distinguish between components required for SLM establishment and components required for canonical meiotic RdDM.

Relaxed targeting of RdDM in meiocytes may explain the specific sRNA pool observed, as increased sensitivity with a large variety of sRNA may cause too much mis-targeting of RdDM and wide-scale repression. Examining whether other RdDM targets are possible through bioinformatic analysis and analysis of methylation in clsy3 knockout meiocytes or meiocytes from other CLSY overexpression lines may reveal the underlying mechanics involved. Analysis of other potential targets in the genome from meiotic RdDM may also indicate whether features such as chromatin accessibility are important factors for *de novo* methylation that we observe with SLMs.

Given the diverse sRNA biology across plants and animals, particularly during sexual reproduction, it is possible that sexual-lineage-specific methylation in *A. thaliana* is a recently acquired trait (*A. thaliana* appears to be among the minority of species that lack phasiRNAs within eudicots and monocots; Xia et al., 2019). CLSY3 and CLSY4 can be distinguished from other CLSYs across flowering plants (Bargsten et al., 2013; Hale et al., 2007). However, the orthologue of CLSY3 and CLSY4 in maize, REQUIRED TO MAINTAIN REPRESSION (RMR1), plays a more

major role in sRNA accumulation, DNA methylation, and paramutation (Hale et al., 2007). Thus, whether gene-targeted methylation exists in the sex cells of other plant species and whether this methylation is produced through similar mechanisms is of current interest. In Chapter Four, gene-targeted methylation and other methylation reprogramming phenomena during sexual reproduction is examined in the basal land plant *Marchantia polymorpha*.

3.8 Materials and Methods

Small RNA library preparation, sequencing, and analysis

Meiocytes were extracted as previously described (Walker et al., 2018). 1 microgram of total RNA was extracted from ~100,000 meiocytes with the mirVANA RNA preparation kit (Promega, UK). This RNA was then separated by PAGE gel and the appropriate sRNA fraction was purified using Zymo sRNA PAGE recovery kit (ZymoResearch, USA) according to the manufacturer's instructions. The sRNA seq library was created using the NEBNext® Multiplex Small RNA kit (NEB, USA) according to manufacturer's instructions. The library was multiplexed with RNA-seq and BS-seq libraries and sequenced on Nextseq 500 and Hiseq 2500 platforms.

All sRNA was mapped using bowtie (Langmead, 2010) to the TAIR10 genome with zero mismatches unless otherwise stated. Meiotic 24 nt sRNA clusters were determined by shortstack (Johnson et al., 2016), and strandedness was determined using a cut-off of 0.8. Overlaps of clusters, genomic features and other regions of interest were determined using BEDtools intersect with the appropriate parameters (Quinlan and Hall, 2010). SLM-like loci used for mismatch analysis are previously detailed (Walker et al., 2018).

Downloaded methyl-seq data used are floral bud wild-type rep1 (GSM2650205), rep2 (GSM2650206), *clsy3-1* (GSM2650210), *clsy4-1*, (GSM2650211), *clsy1-7,2-2* (GSM2650212), *clsy3-1,4-1* (GSM2650213), and *poliv* (*nrpd1-4*; GSM2650216), as well as microspore, sperm, vegetative cell, seedling, rosette leaf, cauline leaf, and rosette leaf detailed in Chapter Two. Downloaded sRNA-seq data are floral bud wild-type rep1 (GSM2650248), rep2 (GSM2650249), and *poliv* (*nrpd1-4*; GSM2650267). poliv- and clsy-dependent sRNA clusters were previously described (Zhou et al., 2018).

SLM sRNA origin locations

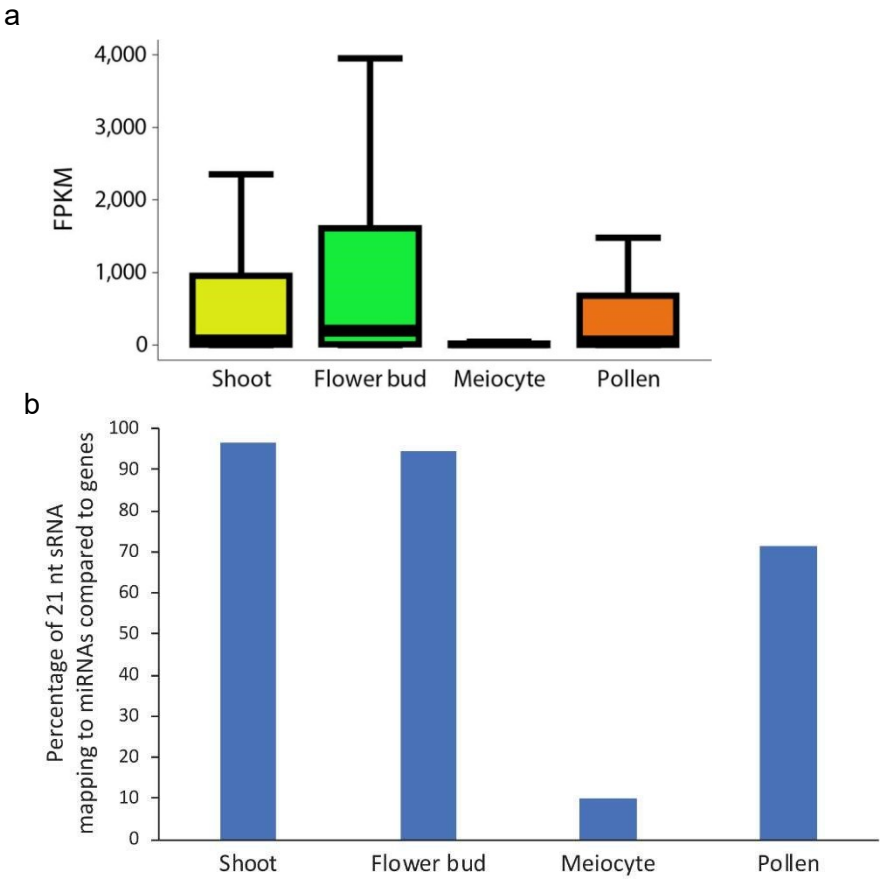
To determine the origin of mismatched SLM sRNA, 21 – 24 nt sRNA were first mapped to the 469 SLM sequences (Walker et al., 2018) with up to three mismatches. Mapped sRNA with at least one mismatch were collected (247,486 reads), remapped to an SLM masked TAIR10 genome allowing zero mismatches, and the sRNA abundance of specific features were determined. To create the sRNA network, the above steps were carried out, except that all SLM sRNAs were used to map against the SLM masked TAIR10 genome, keeping all the best hits with fewer or equal number of mismatches as when the sRNA was mapped to SLMs. A record

of the abundance of each sRNA sequence, the SLM target for each sRNA and the number of mismatches, the associated gene, the sRNA origin location and number of mismatches, as well as the associated meiotic 24 nt sRNA cluster and canonical SLHs were recorded. sRNA were merged into clusters if they were immediately adjacent.

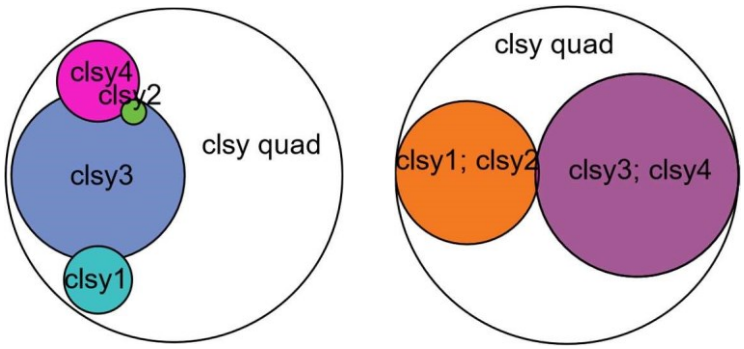
Box plots

All box plots follow this format: each box encloses the middle 50% of the distribution, with the horizontal line making the median, and vertical lines marking the minimum and maximum values that fall within 1.5 times the height of the box.

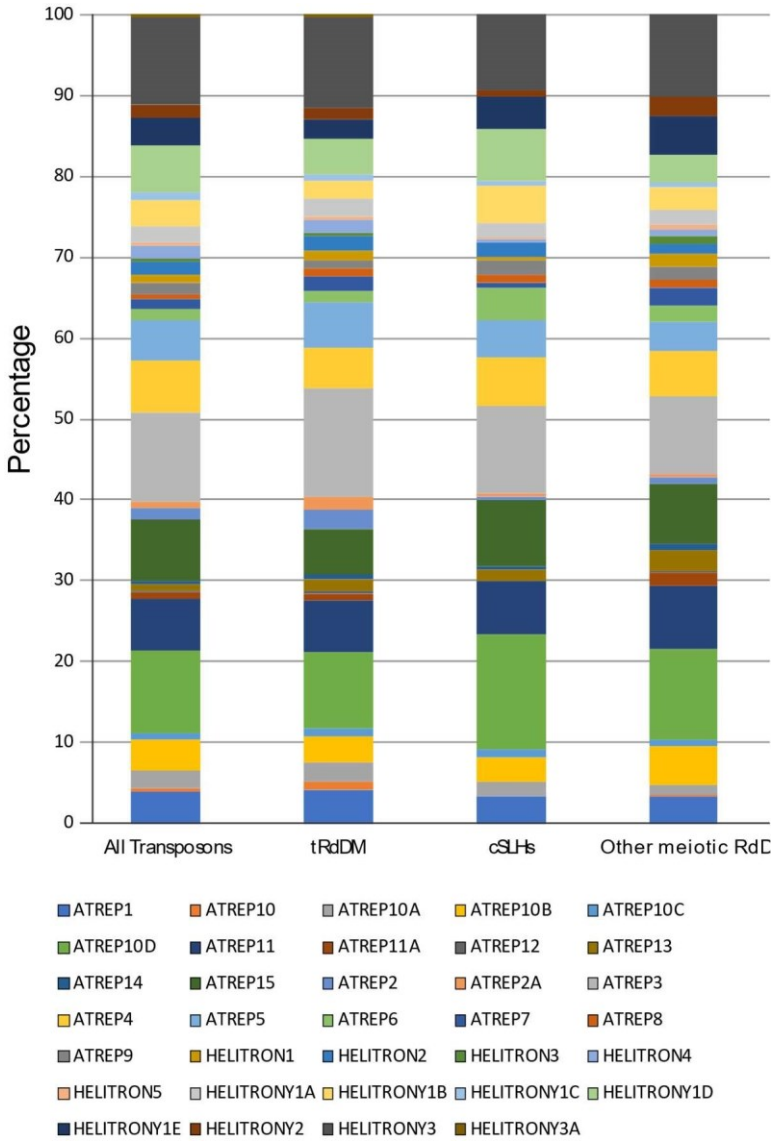
3.9 Supplementary Figures



Supplementary Figure 3.1 21nt association with miRNA genes. a, 21nt sRNA levels over miRNA genes in shoot, flower bud, meiocyte, and pollen. b, Proportion of 21nt sRNA that map to miRNA genes compared to all genes in the TAIR10 genome in shoot, flower bud, meiocyte and pollen.



Supplementary Figure 3.2 Association of meiotic RdDM clusters not overlapping canonical SLHs with clsy-dependent clusters.



Supplementary Figure 3.3 Association of helitron subfamilies with meiotic RdDM.

Chapter Four

Dynamic DNA methylation reprogramming during sexual development in the basal land plant *Marchantia polymorpha*

"It took me 6 months and some special lighting, but I finally have some sex organs!"
– James Walker, 2017

4.1 Abstract

DNA methylation is an epigenetic mark often associated with transposon repression in eukaryotes. While it was previously thought that DNA methylation is static throughout development in the plant cells that are inherited across generations, we recently showed that gene-targeted *de novo* methylation occurs in the male sexual lineage of the plant *Arabidopsis thaliana* and demonstrated that this methylation contributes to both gene expression and cellular function. Whether such changes in DNA methylation occur in the sex cells of other plant species remains to be seen, however. Here, we confirm reports that the sporophyte of *Marchantia polymorpha* (the stage undergoing meiosis) exhibits reinforcement of methylation at repetitive elements and we also confirm a global cytosine methylation phenomenon in *M. polymorpha* sperm cells. We show that like mammals, *M. polymorpha* has two hypermethylation events during sperm maturation. In the early stages methylation reinforcement occurs at repetitive elements, particularly in the CAH context, while at a later stage global methylation excluding the CAH context appears. We also find specific, gene-associated methylation in both the sperm and sporophyte of *M. polymorpha*, with the latter resembling patterns of *de novo* methylation in *A. thaliana* sex cells. In the sperm genome, regions of methylation exclusion are predominantly around genic transcription start sites and we demonstrate that this hypomethylation strongly correlates with antheridia gene expression. Finally, we identify the candidate methylases responsible for sperm methylation that are expressed during sperm development. Together, these results establish the dynamic DNA methylation reprogramming events during sexual reproduction in an early land plant species and highlights potential new roles for this chromatin mark in development.

4.2 Introduction

The four bases within DNA act as the genetic code for RNA and proteins within a cell. Base modifications have also been observed, with a methyl group added at the

fourth position of the cytosine purine ring (4mC), the fifth position of the cytosine purine ring (5mC), and at the sixth position of the adenine pyrimidine ring (6mA). Bacteria and archaea often use these base modifications in the restriction-modification system, whereby methylases mark the host genome at specific sequences and other enzymes recognise the same sequences but only cut unmethylated invading DNA (Oliveira et al., 2014; Wilson and Murray, 1991). 4mC and 6mA methylases have highly similar motifs within their catalytic domains (TSPPY and (D,N)PPY respectively; Klimasauskas et al., 1989) and there are indications that a 4mC or 6mA methylase can modify both cytosine and adenine (Jeltsch, 2001). 5mC methylases, on the other hand, have an invariable Pro-Cys dipeptide which forms a transient covalent bond to the carbon-6 that activates carbon-5 as a methyl acceptor (Malone et al., 1995).

As was described in Chapter one, many eukaryotes utilise 5mC methylases evolutionarily acquired from bacteria (Aravind et al., 2014). The context in which methylation occurs in eukaryotes is less specific than the restriction-modification system in prokaryotes, with enzymes recognising features such as CG in mammals (although non-CG methylation has also been reported in neurons and embryonic stem cells; He and Ecker, 2015; Ziller et al., 2011) or CG, CHG, and/or CHH in plants. The methylation of the eukaryotic genome is typically mosaic, i.e. with regions of low and high methylation. Regions of high methylation typically associate with specific genomic features such as gene bodies and transposons, which is the case with most plants (Figure 4.1; Suzuki and Bird, 2008). The role of gene body methylation is still unclear - it is typically associated with constitutively expressed genes - although there are suggestions that the methylation prevents cryptic transcription (Zemach and Zilberman, 2010; Zhou et al., 2019). Methylation of transposons, however, is known to repress expression of these elements and prevent transposition that would otherwise harm the host genome (Kato et al., 2003). Despite being an important signal for repression, some organisms such as fission and budding yeast, *Drosophila*, and *C. elegans* have lost 5mC altogether, showing that in some circumstances it is dispensable (Zemach and Zilberman, 2010). Vertebrates have perhaps one of the most unique methylation patterns, as most cytosines in the CG context are methylated in a blanket manner across the genome (Figure 4.1), though even here methylation is also thought to be especially associated with transposons, particularly in the male germline (Barau et al., 2016; Lehnertz et al., 2003). The only regions of low methylation in vertebrates occur over promoters and transcription start sites of genes or in enhancer elements (Figure 4.1;

Ambrosi et al., 2017). The purpose of blanket methylation in vertebrates is unclear, although some evidence suggests a similar host-recognition system as in prokaryotes (Krieg, 2006; Suzuki and Bird, 2008). How such a system evolved from organisms with a mosaic methylation pattern without drastic chromatin and phenotypic consequences is unknown.

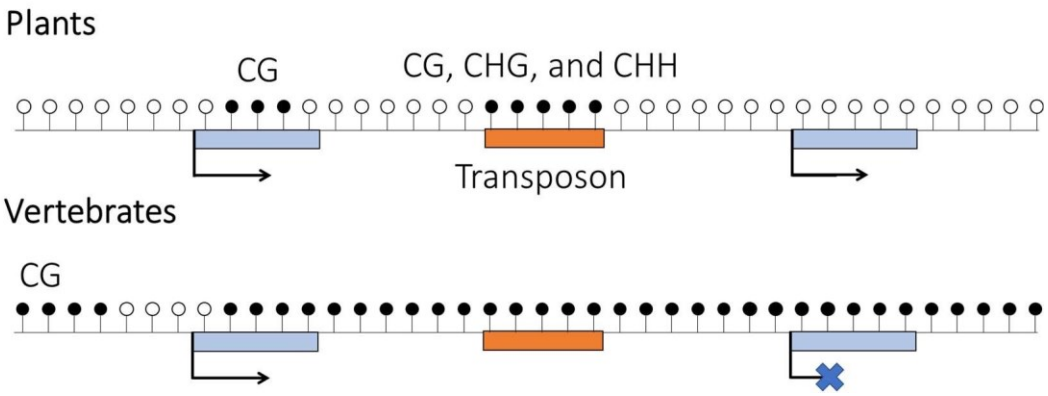


Figure 4.1 Typical methylation patterns of plants and vertebrates. Methylation patterns observed in plants and vertebrates. Blue bars represent genes which are either expressed or repressed (blue cross). Lollipops represent cytosines that are either methylated (black) or unmethylated (white). H = A, C, or T.

The eukaryotic 5mC DNA methylase families consist of DNA (cytosine-5) methyltransferase 1 (DNMT1), Defective in methylation 2 (Dim-2), Chromomethylase (CMT), DNMT3, DNMT5 and DRM, some of which have been previously introduced. DNMT1 (MET1 in plants) maintains CG methylation across cell divisions via Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1, or VIM proteins in plants), which recognises hemi-methylated DNA during DNA replication. Dim-2 is exclusively found in fungi and is thought to have evolved from DNMT1 (Zemach and Zilberman, 2010). Dim-2 (also known as DNMT4) recognises H3K9 methylation and methylates the DNA of fungal transposons. CMTs are the plant equivalent of Dim-2 and are also thought to have evolved from DNMT1 (Zemach and Zilberman, 2010). Unlike other methylases, CMTs have a chromodomain allowing them to directly recognise H3K9 methylation in heterochromatic regions and methylate DNA. Flowering plants utilise CMT2 to methylate CHH contexts and CMT3 to methylate CHG contexts (Stroud et al., 2013), although more specific preferences have been observed for CMT2: CAA and CCA in maize and *A. thaliana*, and CAA as well as CAT in tomato (Gouil and Baulcombe, 2016). CMT3 is also proposed to be important for gene body

methylation (which occurs primarily in the CG context; Figure 4.1) through some undetermined mechanism, with CMT3 loss in *Conringia planisiliqua* and *Eutrema salsugineum*, two Brassicaceae species, being associated with a lack of gene body methylation (Bewick et al., 2016). Non-flowering plant CMTs form separate clades to CMT2 and CMT3, although CMT in the moss *Physcomitrella patens* has been shown to be responsible for CHG methylation of transposons in a similar manner to CMT3 in flowering plants (Noy-Malka et al., 2014). DNMT3 (DNMT3A and DNMT3B in mammals) methylates the CG context across eukaryotes *de novo* as well as CH contexts in mammals in pluripotent cells and the brain, with a preference for CA (He and Ecker, 2015). DNMT5 is a structurally distinct orthologue of DNMT3 but also methylates the CG context in algae and fungi (Huff and Zilberman, 2014). DNMT3 and DNMT5 are absent in flowering plants but another orthologue, DRM, methylates cytosines in all contexts (CG, CHG, and CHH) in the previously described RdDM pathway with no specific context preference (Gouil and Baulcombe, 2016). The basal land plants *P. patens* and *M. polymorpha* have both DNMT3A/B and DRM genes (Bowman et al., 2017; Malik et al., 2012), and recent work has established that PpDNMT3B can methylate CG and CHH contexts *de novo* but also acts like CMT2 in flowering plants by maintaining the methylation of transposons in the CHH context across the moss genome (Yaari et al., 2019).

As we've seen, epigenetic reprogramming during reproductive development is commonly observed and issues with this reprogramming often results in sterility or failed embryonic development. In the sperm of numerous animals including mammals, histones are replaced by arginine-rich proteins known as protamines as part of the DNA compaction process (Balhorn, 2007), and this compaction is accompanied by a loss of transcriptional activity in the mature sperm cell (Goodrich et al., 2013; Grunewald et al., 2005). Protamine-like proteins are also used in plant species with motile sperm such as *P. patens* and *M. polymorpha* (Higo et al., 2016) and RNA extraction from *M. polymorpha* sperm has been difficult in our hands and others (Schmid et al., 2018). Global methylation is lost and regained twice during mammalian development, first in primordial germ cells (both sperm and egg), and following fertilisation (Zeng and Chen, 2019). Blanket methylation is regained in developing mammalian sperm prior to meiosis while mosaic methylation is established in the oocyte, with indications that this is related to imprinting and germ cell function (Stewart et al., 2016; Zeng and Chen, 2019). Plant methylation was thought to be generally static and not drastically reprogrammed in the cells that pass on information to the next generation as there is a strong association of methylation

with transposons and methylation is faithfully maintained and inherited (Feng et al., 2010; Pikaard and Mittelsten Scheid, 2014). However, it is now clear that DNA methylation reprogramming does occur in flowering plants, at least on a local level at distinct loci (Chapter Two; Walker et al., 2018). We were thus curious to explore the methylation changes during sexual development in a basal land plant, *M. polymorpha* (an emerging model system for plants due to the ease of transformation, resource availability, and relatively fast life cycle of 1-2 months; Ishizaki et al., 2016) to compare any similarities or differences in methylation patterns and mechanisms with *A. thaliana* and other organisms.

The haploid gametophyte generation is the dominant stage of life in *M. polymorpha*, other liverworts, mosses, and hornworts (Ishizaki et al., 2016) *M. polymorpha* is dioecious, having male and female plants, each with eight autosomes and either an X chromosome (female) or Y chromosome (male; Bowman et al., 2016). The male and female spores develop mitotically to produce the thallus (relatively small leafy tissue – synonymous but not orthologous with leaf tissue) and rhizoids (root-like tissue). *M. polymorpha* can reproduce asexually by producing clones known as gemmae inside gemmae cups on the topside of thalli. Each gemma comes from a single cell, allowing homogenous mutants to be created from chimeric parents (Ishizaki et al., 2016). Gemmae are dispersed by rain and develop into more thalli. Sexual organ induction occurs with the addition of far-red light, producing antheridiophores (male) and archegoniophores (female) from meristematic tissue. Antheridiophores produce motile sperm (antherozoids) inside capsules known as antheridia, and the sperm are released from the surface of the antheridiophores upon the addition of water. The sperm swims to the archegoniophores and fertilises the egg that develops inside archegonia (with multiple archegonia per archegoniophore). The resulting zygote develops into an embryo and then a relatively small mature sporophyte while attached to the female gametophyte (it is therefore dependent of the gametophyte for resources). Numerous cells inside the sporophyte subsequently undergoes meiosis to produce male and female spores. Thus, while *A. thaliana* sperm are produced after two mitotic events from meiosis, *M. polymorpha* sperm are produced after spores have undergone the vegetative phase of development. As a result, the presence of gene-associated *de novo* methylation in either the sperm or the sporophyte (the tissue undergoing meiosis) may help illuminate the role of this mechanism in development.

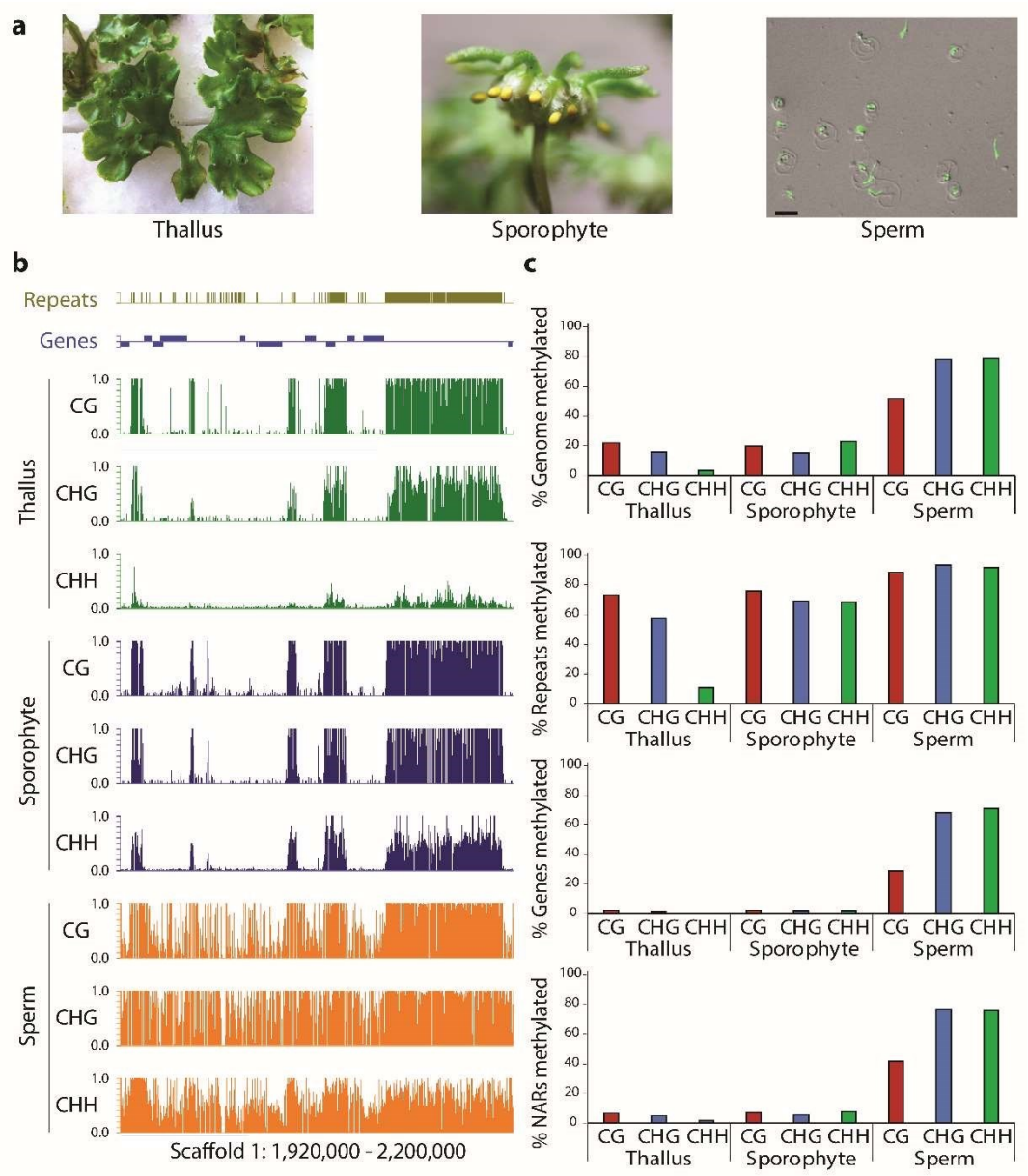
Repetitive elements in *M. polymorpha* represent 22% of the autosomal genome compared to 48% in *P. patens* (Bowman et al., 2017; Rensing et al., 2008). Methylation is observed in CG, CHG, and CHH contexts across repetitive elements in the gametophyte and sporophyte but not notably in gene-bodies (Ikeda et al., 2018; Schmid et al., 2018; Takuno et al., 2016). CHH methylation is lower in the gametophyte than CG and CHG methylation and preferentially occurs in the CCH context, while in the sporophyte CHH methylation is increased (Ikeda et al., 2018; Schmid et al., 2018). *M. polymorpha* has MET1, DRMa, DRMb, CMTa, CMTb, DNMT3A and DNMT3B. MpDNMT3A and MpDNMT3B are not orthologous to mammalian DNMT3A and DNMT3B, respectively, but were independently duplicated (Yaari et al., 2019). MpMET1 mutants lose CG methylation, as expected, but also gain non-CG methylation specifically in the CTA context likely due to a shift in chromatin accessibility and/or methylase activity (Ikeda et al., 2018). Mutations in MpMET1 also result in undifferentiated tissue, and thus these plants are unable to undergo sexual reproduction (Ikeda et al., 2018). MpDRMa is constitutively expressed, while MpDRMb is expressed in the diploid sporophyte (Bowman et al., 2017). Similarly, MpCMTa is expressed constitutively (although it is enriched in antheridiophores), while MpCMTb is exclusively expressed in the sporophyte (Bowman et al., 2017). MpDNMT3A and MpDNMT3B are generally expressed but are also upregulated in antheridiophores (Bowman et al., 2017). Other genes of note that are upregulated in the antheridiophore are MpHEN1, an orthologue of a methyltransferase that methylates sRNAs on the ribose of the last nucleotide to protect the 3' ends of the sRNAs from degradation, and MpPiwia as well as MpPiwib, tandemly duplicated genes that are typically involved in sRNA binding. However, these Piwis lack the standard PAZ domain known to interact with the 3’ end of sRNA (Bowman et al., 2017; Ma et al., 2004). These expression patterns suggest that methylation reprogramming or at least methylation reinforcement during sex development is prominent in *M. polymorpha*. Work done in parallel to my PhD has shown this to be the case, with *M. polymorpha* showing some of the most drastic methylation changes seen in any plant (Schmid et al., 2018). This data is thus analysed alongside this work.

4.3 The genome of *M. polymorpha* is heavily methylated in sperm

To compare the DNA methylation at key stages of reproductive development in *M. polymorpha*, I generated and sequenced bisulphite libraries of the male haploid gametophyte thallus, haploid sperm cells, and the mature diploid sporophyte containing spores (Figure 4.2a). As expected, CG methylation in the thallus covers approximately 20% of the genome (based on scaffold sequence data) and this methylation is largely restricted to repetitive regions (Figure 4.2b; Figure 4.2c; Figure 4.2d; Supplementary Figure 4.1; Bowman et al., 2017; Ikeda et al., 2018; Schmid et al., 2018). A similar proportion of the genome has appreciable amounts of CHG methylation (16%), while the proportion of the genome that shows CHH methylation is only 3% (Figure 4.2c). Methylation in these contexts is also primarily associated with repetitive regions (Figure 4.2b; Figure 4.2d). As a result, only repetitive elements show detectable levels of methylation in any context, with high levels of CG methylation, moderate levels of CHG methylation, and low levels of CHH methylation (Figure 4.2b; Figure 4.2e; Supplementary Figure 4.1; Ikeda et al., 2018; Schmid et al., 2018). In the sporophyte, the proportion of the genome showing CG and CHG methylation is comparable to the thallus (19% and 15%, respectively), while a considerably larger portion of the genome shows CHH methylation (23%; Figure 4.2c; Figure 4.2e). This again is primarily associated with repetitive elements so that an increase in the levels of CHH (and CHG) methylation are observed over these regions (Figure 4.2b; Figure 4.2d; Figure 4.2e; Supplementary Figure 4.1).

In contrast to the thallus and sporophyte, a remarkably high proportion of the genome has CHG and CHH methylation in sperm (78% and 79%, respectively; Figure 4.2c). This methylation covers genes, transposons and non-annotated regions so that increased levels of CHG and CHH methylation can be seen for these three genomic features in an indiscriminate manner compared to the thallus and sporophyte (Figure 4.2b; Figure 4.2c; Figure 4.2d; Figure 4.2e). Published sperm data shows the same indiscriminate methylation patterns (Supplementary Figure 4.2; Schmid et al., 2018). However, while the published data exhibits blanket CG methylation in a similar manner to CHG and CHH, our observations show that sperm CG methylation is still partially restricted to repetitive elements (covering 52% of the genome and showing lower levels of methylation than CHG and CHH for genes and other regions; Figure 4.2b; Figure 4.2c; Figure 4.2d; Figure 4.2e; Supplementary Figure 4.1). In all contexts the published data shows a greater proportion of methylation in sperm, covering a remarkable 99%, 95%, and 90% of the genome for CG, CHG, and CHH, respectively (Figure 4.2c, Figure 4.2e;

Supplementary Figure 4.2; Schmid et al., 2018; Takuno et al., 2016). This indicates either an issue with bisulphite conversion or that there is some variation of methylation in *M. polymorpha* sperm, as is also known to occur for flowering plants (Hsieh et al., 2016; Ibarra et al., 2012). Conversion rate is difficult to check from the available data, as in our hands for example the thallus and gametophyte show chloroplast CHH methylation of <0.3% but sperm shows about 4% CHH methylation which could be potentially biological given the strange global methylation observed in sperm– a spike in control of unmethylated DNA is thus required Together our results confirm that, while the gametophyte and sporophyte of *M. polymorpha* show typical plant methylation profiles, *M. polymorpha* sperm cells exhibit a drastic blanket DNA methylation reprogramming event that is unseen in other plants and is largely indiscriminate in targeting the genome.



See following page for legend.

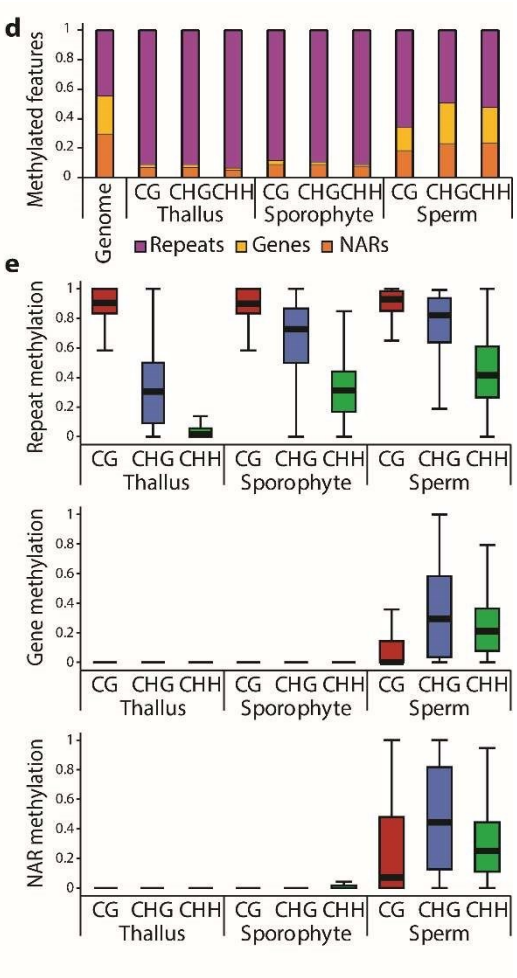


Figure 4.2 The sperm of *M. polymorpha* has a radically different methylation profile to other plants. a, The male thallus with gemmae cups, an archegoniophore with mature sporophytes (yellow), and sperm (nuclei stained with SYBR-Green) of *M. polymorpha*. Thallus and sporophyte images were obtained from Jim Haseloff (https://haseloff.plantsci.cam.ac.uk/). Scale bar, 10 µm. **b**, Snapshot of DNA methylation patterns in the male thallus (green), mature sporophyte (blue), and sperm (orange) in CG, CHG, and CHH contexts. **c**, Bar charts showing the percentage of 50 bp windows across the *M. polymorpha* genome with more than 10% CG, CHG, or CHH methylation in the three samples across the genome or associated with repetitive elements, genes, or non-annotated regions (NARs). **d**, Bar chart showing the proportions of the genome exclusively associated with repetitive elements, genes, or NARs and the proportions of 50 bp windows associated with these features that have more than 10% CG, CHG, or CHH methylation. **e**, Methylation levels in the thallus, sporophyte, and sperm for CG, CHG and CHH contexts for repetitive elements, genes, and NARs.

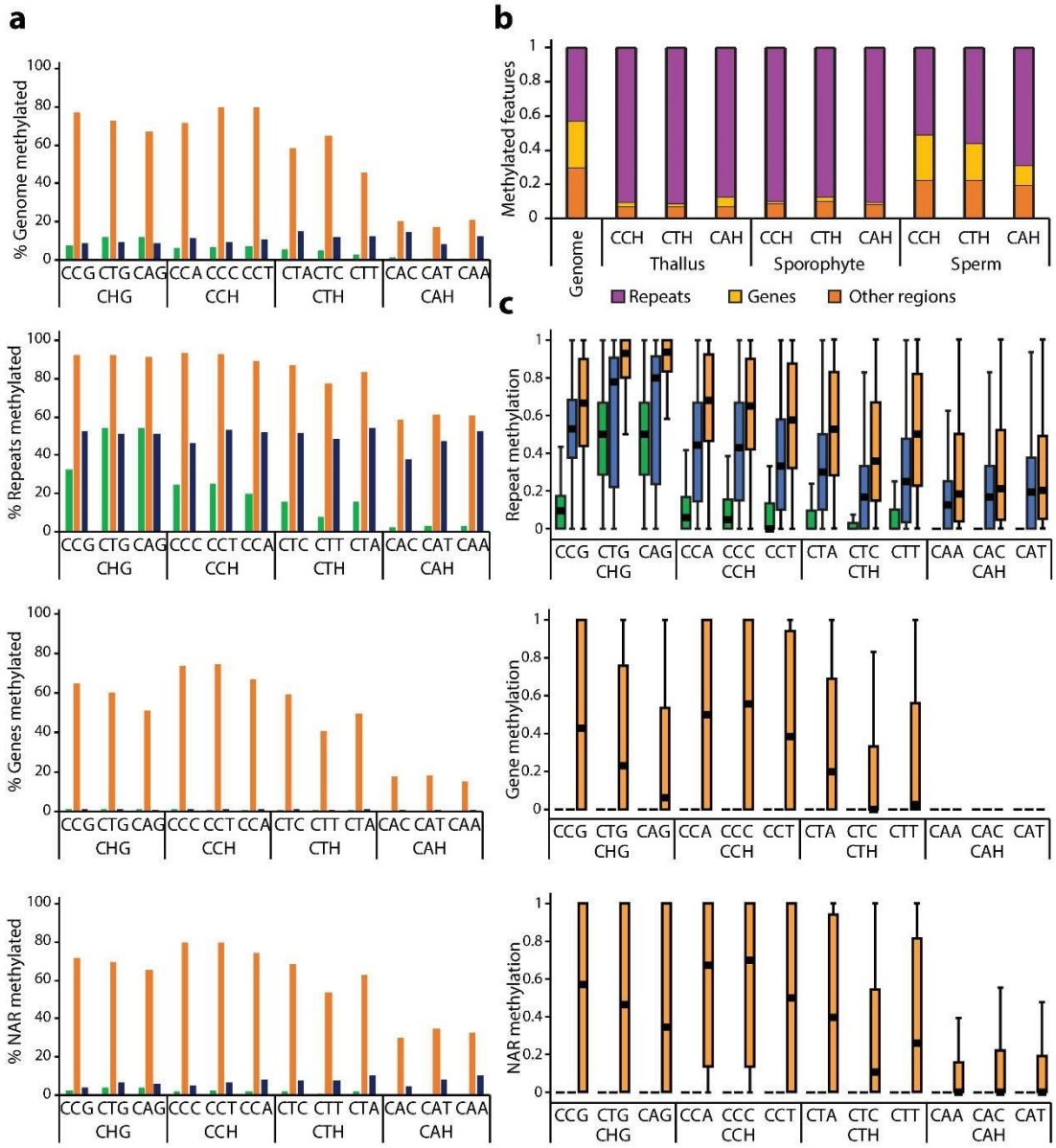
4.4 Transposons in sperm have distinguishable CAH methylation

Given the known differences in methylation between specific CHH contexts (Gouil and Baulcombe, 2016; Ikeda et al., 2018), I decided to examine the unique non-CG methylation patterns of *M. polymorpha* in more detail. As was previously reported (Ikeda et al., 2018), the thallus shows a preference for CHG, CCH, and CTH methylation over CAH methylation, with 8-12%, 6-7%, 3-6%, and 1% of the genome showing methylation in these contexts, respectively (Figure 4.3a). In line with the overall CHH context, CCH, CTH, and CAH methylation in the thallus is restricted to repetitive elements (Figure 4.3a; Figure 4.3b; Figure 4.3d; Supplementary Figure 4.4; Supplementary Figure 4.5). Similarly, only repetitive elements have measurable levels of methylation in the thallus and these show similar patterns to the extent of methylation, along with low CCG methylation as is common in plants (Figure 4.3c; Figure 4.3d). In the sporophyte, the proportion of the genome with CHG methylation is similar to the thallus (9%) while the proportion of the genome with CCH and CTH methylation is larger (9-11% and 12-15%), and the proportion of the genome with CAH methylation is also considerable (8-15%; Figure 4.3a). As with CHH methylation overall, methylation in these contexts is restricted to repetitive elements (Figure 4.3a; Figure 4.3b; Figure 4.3d; Supplementary Figure 4.4; Supplementary Figure 4.5). The methylation levels of repetitive elements are higher in the sporophyte than the thallus in all contexts, although the patterns remain the same (such as the lower CCG methylation and lower CAH methylation; Figure 4.3c; Figure 4.3d; Supplementary Figure 4.4; Supplementary Figure 4.5).

In the sperm genome, there is a slight preference for CHG and CCH methylation over CTH methylation, although CTH methylation still covers a substantial proportion of the genome (67-77%, 72-79%, and 46-65%, respectively; Figure 4.3a). This pattern is reflected in the genic regions and non-annotated regions, while the extent of methylation in these contexts over repetitive elements is generally saturated (Figure 4.3a). Similarly, the levels of CHG and CCH methylation are higher in repetitive elements, genic regions, and non-annotated regions than CTH methylation (Figure 4.3c; Figure 4.3d).

The proportion of the sperm genome with CAH methylation is remarkably less than the other contexts, covering only 17-21% of the genome (Figure 4.3a). These patterns are also reflected in the published sperm data (Supplementary Figure 4.3). However, the proportion of the genome with CAH methylation is still more than both the thallus and sporophyte (Figure 4.3a). This may be due to hypermethylation over repetitive elements or else blanket methylation as with the other CHH contexts.

Unlike the sporophyte, there is detectable CAH methylation in genic regions and non-annotated regions in sperm (Figure 4.3a; Figure 4.3b; Figure 4.3d), showing that there is at least some blanket CAH methylation. However, a significantly larger proportion of CAH methylation is seen in repetitive elements than genic regions and non-annotated regions (Figure 4.3a; Figure 4.3b; Supplementary Figure 4.4 Supplementary Figure 4.5). CAH methylation levels for repetitive elements are also notably higher than non-annotated regions, while CAH methylation levels over genes cannot be detected (Figure 4.3c; Supplementary Figure 4.4; Supplementary Figure 4.5). In addition, methylation levels for CAH and other contexts are higher over repetitive elements in sperm compared to both the thallus and sporophyte, particularly at the edges of repetitive elements and/or in shorter repetitive regions (Figure 4.3a; Figure 4.3b; Supplementary Figure 4.4; Supplementary Figure 4.5). I identified 8,504 CAH methylated regions in sperm, the majority of which overlapped with CG methylated regions in the thallus or CAH methylated regions in the sporophyte (7,576 and 7,470, or 94% and 93%, respectively) and with repetitive regions (6,478; 76%; Figure 4.2c; Supplementary Table 4.1). Together, these results show that the global methylation pattern in *M. polymorpha* sperm generally excludes the CAH context. However, reinforcement of methylation at repetitive elements including the CAH context in sperm allows these elements to be distinguished from the global methylated regions. There are more subtle differences between the CHH contexts in both the extent and level of methylation for all features examined, which likely reflects the biochemistry of the methylase(s) involved but remains to be determined.



See following page for legend.

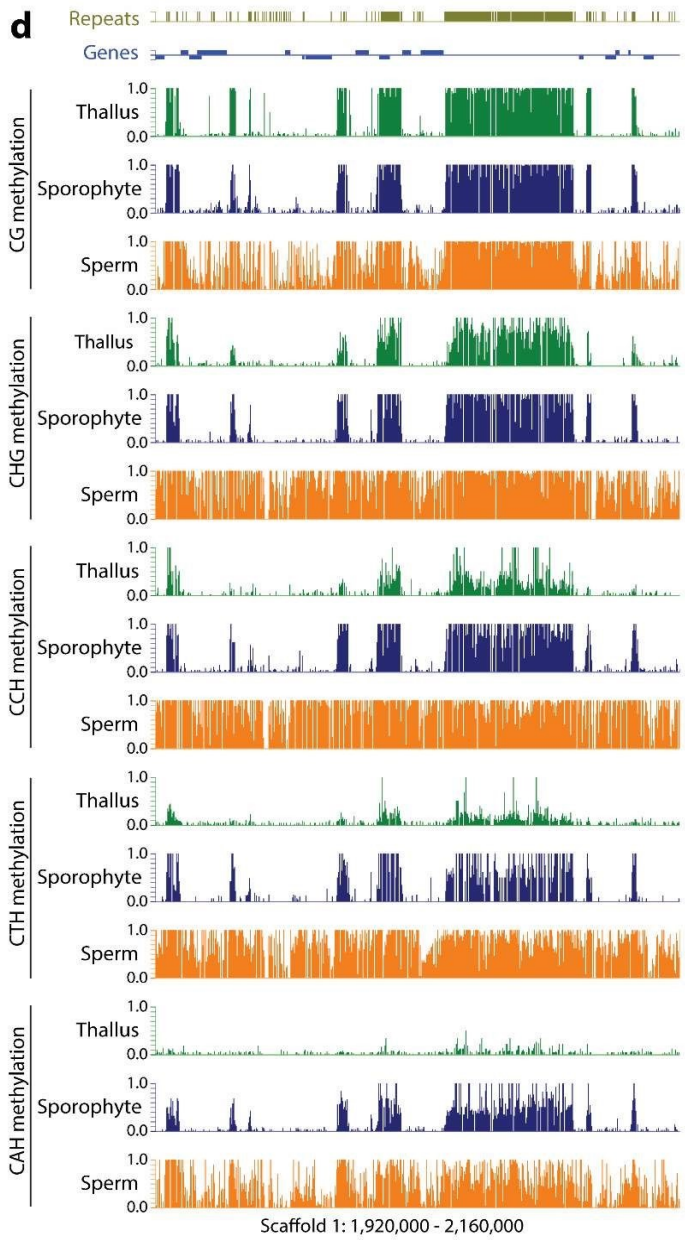


Figure 4.3 Blanket and targeted CAH methylation in *M. polymorpha* sperm. a, Bar charts showing the percentage of 50 bp windows across the *M. polymorpha* genome with more than 10% CHG, CCH, CTH, or CAH methylation in the three samples across the genome or associated with repetitive elements, genes, or nonannotated regions (NARs). **b**, Bar chart showing the proportions of the genome exclusively associated with repetitive elements, genes, or NARs and the proportions of 50 bp windows associated with these features that have more than 10% CCH, CTH, or CAH methylation. **c**, Methylation levels in the thallus, sporophyte, and sperm for CHG, CCH, CTH, and CAH contexts. **d**, Snapshot of DNA methylation patterns as in Figure 4.2a in the male thallus (green), mature sporophyte (blue), and sperm (orange) in CG, CHG, CCH, CTH, and CAH contexts.

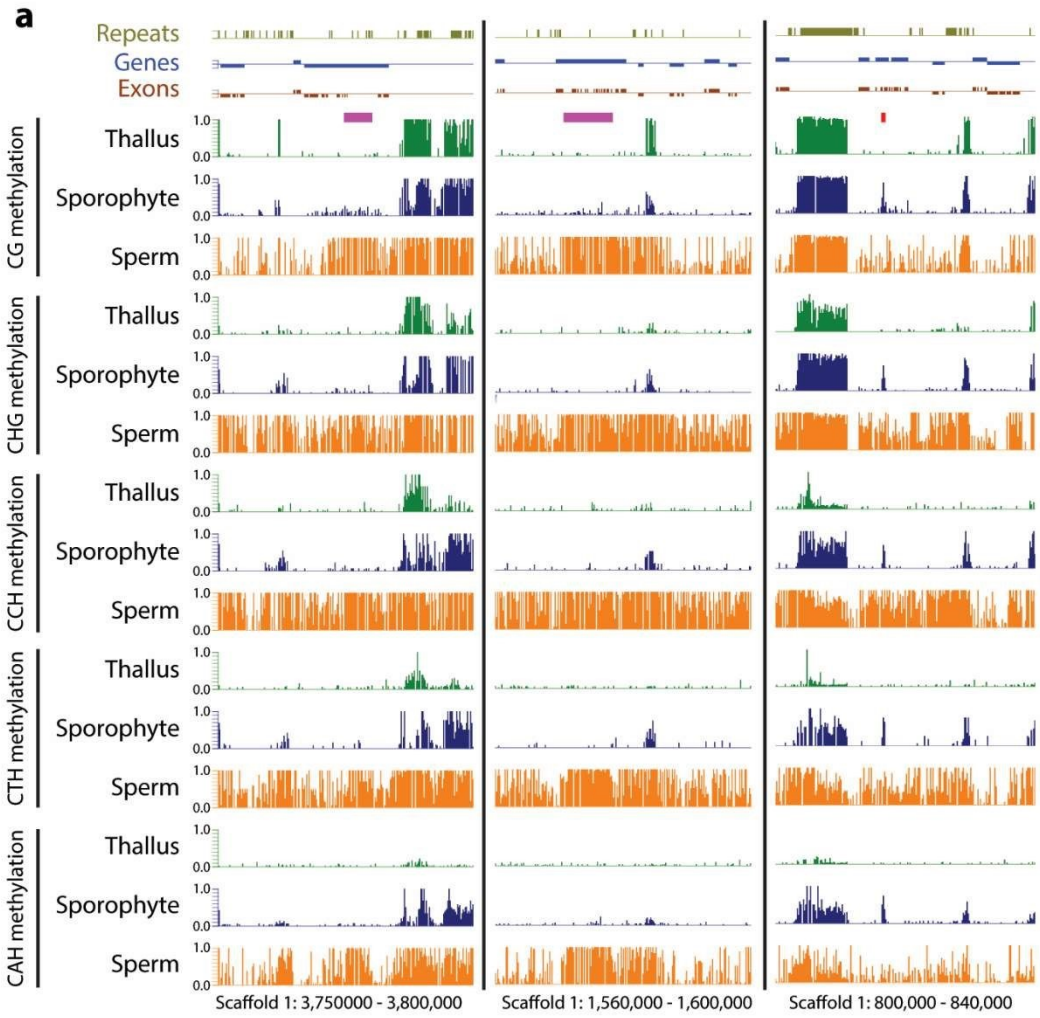
4.5 Gene-associated methylation exists in sperm and sporophytes

Of the 511 CAH methylated regions in sperm that don't overlap CG methylated regions in the thallus, CAH methylated regions in the sporophyte, or repetitive elements, I noted that 241 (47%) overlapped genes (hereafter referred to as genic CAH islands; Figure 4.4a; Supplementary Table 4.1). These 241 genic CAH islands regions also displayed CG methylation in the sperm so that they are methylated in a similar manner to repetitive elements (Figure 4.4a; Figure 4.4b). However, while genic CAH islands have no methylation in the thallus and only residual (if any) methylation in the sporophyte, they are methylated even more heavily than repetitive elements in sperm (Figure 4.4a; Figure 4.4b). Genic CAH islands have a median size of 3,250 bp. Although 11 are solely intronic and a further 25 overlap with an intron by at least 50% of their length, the majority occur over both introns and exons (Figure 4.4a). The expression of genes associated with CAH islands in different tissues did not reveal any obvious link to this sperm-specific methylation pattern (data not shown). However, a GO analysis showed an enrichment for epigenetic factors and development, similar to an analysis carried out previously on genes with detectably more methylated cytosines overall in sperm (Schmid et al., 2018). These genes include two AGOs, SPLAYED, Flowering Locus A, and a SUVH2 gene, critical genes for epigenetics or meristem identity (Supplementary Table 4.1). Thus, in addition to the blanket methylation and repetitive element hypermethylation, there is a subset of gene-associated regions that are targeted for methylation in *M. polymorpha* sperm (probably utilising the same mechanism for repetitive element methylation) with implications on gene expression. Together, these results show dynamic DNA methylation reprogramming within *M. polymorpha* sperm beyond global methylation, with potentially critical roles for development.

We were curious to identify potential *de novo* methylated regions in the sporophyte where MpDRMb is expressed and repetitive element hypermethylation also occurs. Among the 7,039 loci that were differentially methylated between thallus and sporophyte in the non-CG context, almost all (7,037; 99.9%) were hypermethylated in the sporophyte (Supplementary Table 4.2). The majority overlap with thallus CG methylated islands (6,144, 87%), as expected due to the observed hypermethylation of repetitive elements (Figure 4.1c, d and e; Supplementary Figure 4.1; Schmid et al., 2018). However, 443 non-CG hyper DMRs show marginal CG methylation in the thallus (Figure 4.4b; Supplementary Table 4.2). These sporophyte-specific methylated regions are relatively small (a median size of 800 bp), together encompassing 0.2% of the nuclear genome. These regions display minimal levels of

methylation in the thallus but moderate levels of methylation in all cytosine contexts in the sporophyte (Figure 4.4b). Notably, higher levels of CG and CAH methylation is found over these regions compared to non-repetitive regions in sperm, suggesting that these sporophytic DMRs are also targeted for methylation in male germ cells (and indeed may be established in sperm; Figure 4.4b). However, only 58 (14%) of the sporophyte-specific DMRs overlap with sperm CAH islands, showing that there is minimal overlap of these features based on our current analysis (Supplementary Table 4.2). While the other non-CG sporophytic DMRs showing CG methylation in the thallus are more likely to overlap repetitive elements, 111 (26%) of the sporophyte-specific methylated loci overlap genes (Supplementary Table 4.2). Further analysis is currently underway to determine the significance of this association.

Together, these results show a *de novo* methylation pattern in the sporophyte as well as sperm in *M. polymorpha*. The former more closely resembles SLMs in *A. thaliana* due to the smaller size. However, the dependency on RdDM activity remains to be seen.



See following page for legend.

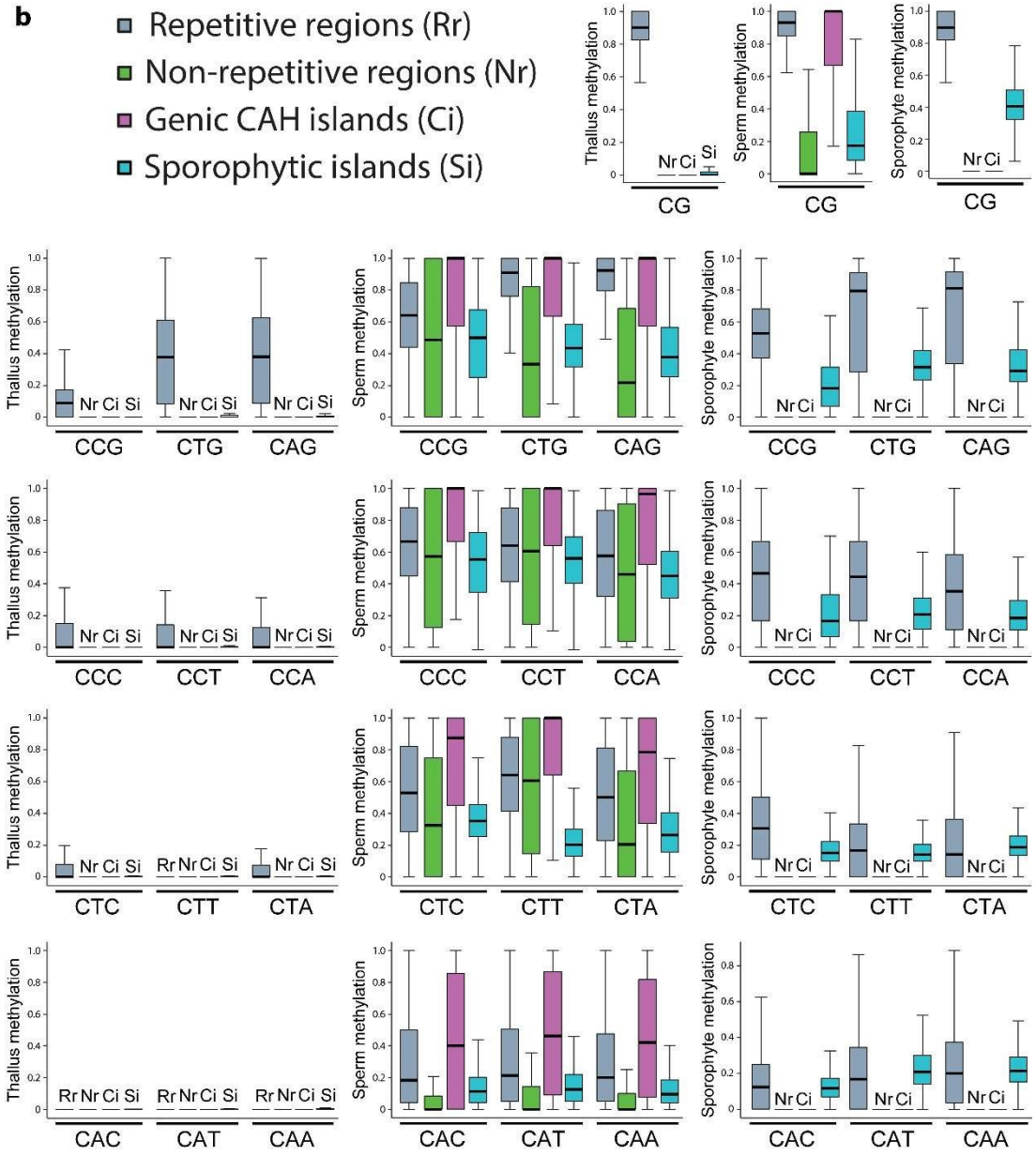


Figure 4.4 Specific methylated regions in the sperm and sporophyte. a, Snapshots of DNA methylation in thallus, sperm, and sporophyte for CG, CHG, CCH, CTH, and CAH contexts showing examples of genic CAH islands and a sporophytic island. Example sperm genic CAH islands depicted with magenta bars. The example sporophytic island is depicted with a red bar. **b**, Methylation levels of 50 bp windows in CG and specific non-CG contexts in the thallus, sperm and sporophyte for repetitive regions (Rr), non-repetitive regions (Nr), genic CAH islands (Ci), and sporophytic islands (Si).

4.6 Low-methylated regions are associated with genes expressed during sperm development

In vertebrates, global methylation is punctuated by low methylation at promoters and enhancers related to transcription factor binding and gene expression (Ambrosi et al., 2017). I was therefore curious whether similar patterns could be observed in *M. polymorpha* sperm. Due to the higher proportion of the sperm genome with CHG, CCH, and CTH methylation compared to CG and CAH methylation, I only considered these three contexts for further analysis. I identified 8,060 low-methylated regions (LMRs) which were significantly hypomethylated in two of either CHG, CCH, and CTH contexts in comparison to the whole genome (Supplementary Table 4.3). The median size of these regions is 2,700 bp, together making up 10.8% of the genome. LMRs primarily overlap genes or non-annotated regions rather than repetitive regions (66%, 27%, and 6% respectively; Figure 4.5a), with a total of 5,922 out of 19,287 genes (31%) having at least one LMR. Genic LMRs typically overlap transcriptional start sites or are within the gene body (Figure 4.5a; Figure 4.5b; Figure 4.5c). The same profile is also seen in the published sperm data, although CG hypomethylation is less pronounced (again perhaps due to conversion or variation; Supplementary Figure 4.6; Schmid et al., 2018). As low-methylated regions in mammalian sperm have been linked to expression of housekeeping and sperm-specific genes (Molaro et al., 2011), I was interested to see whether gene expression in *M. polymorpha* sperm correlates with transcription start site hypomethylation. As stated, however, RNA extraction from *M. polymorpha* sperm has been difficult in both our hands and others (Schmid et al., 2018). We therefore decided to analyse published RNA-seq data of pooled antheridia containing sperm at different developmental stages including the late phase of development (Higo et al., 2016). As expected, there is a correlation between antheridia gene expression and hypomethylation (Figure 4.5d; $p < 0.001$ K-smirnov test). Thus, the blanket methylation observed in sperm behaves similarly to mammals, with exclusion at the start of genes that are expressed during sperm development. Given the well-established role of DNA methylation for gene suppression at transcriptional start sites (Walker et al., 2018; Schmitz et al., 2019), it is possible that methylation exclusion allows gene expression during antheridia development. Alternatively, transcription factors and polymerase II activity may exclude methylase activity in these regions.

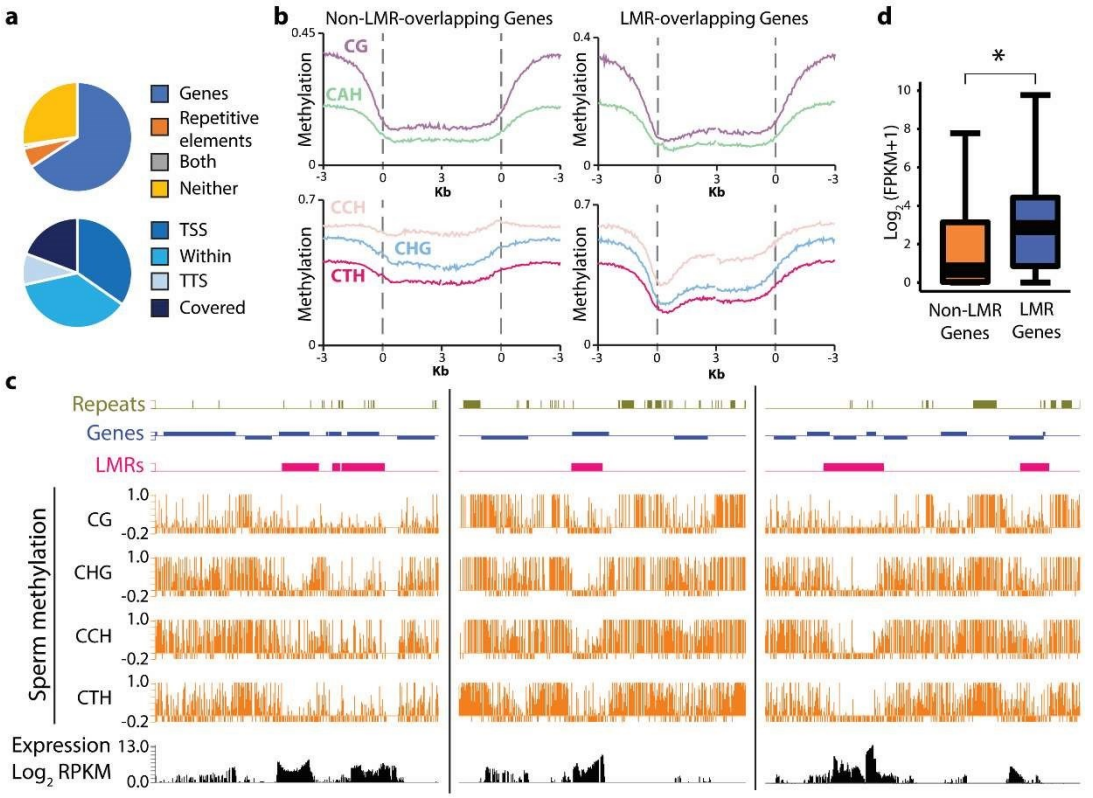


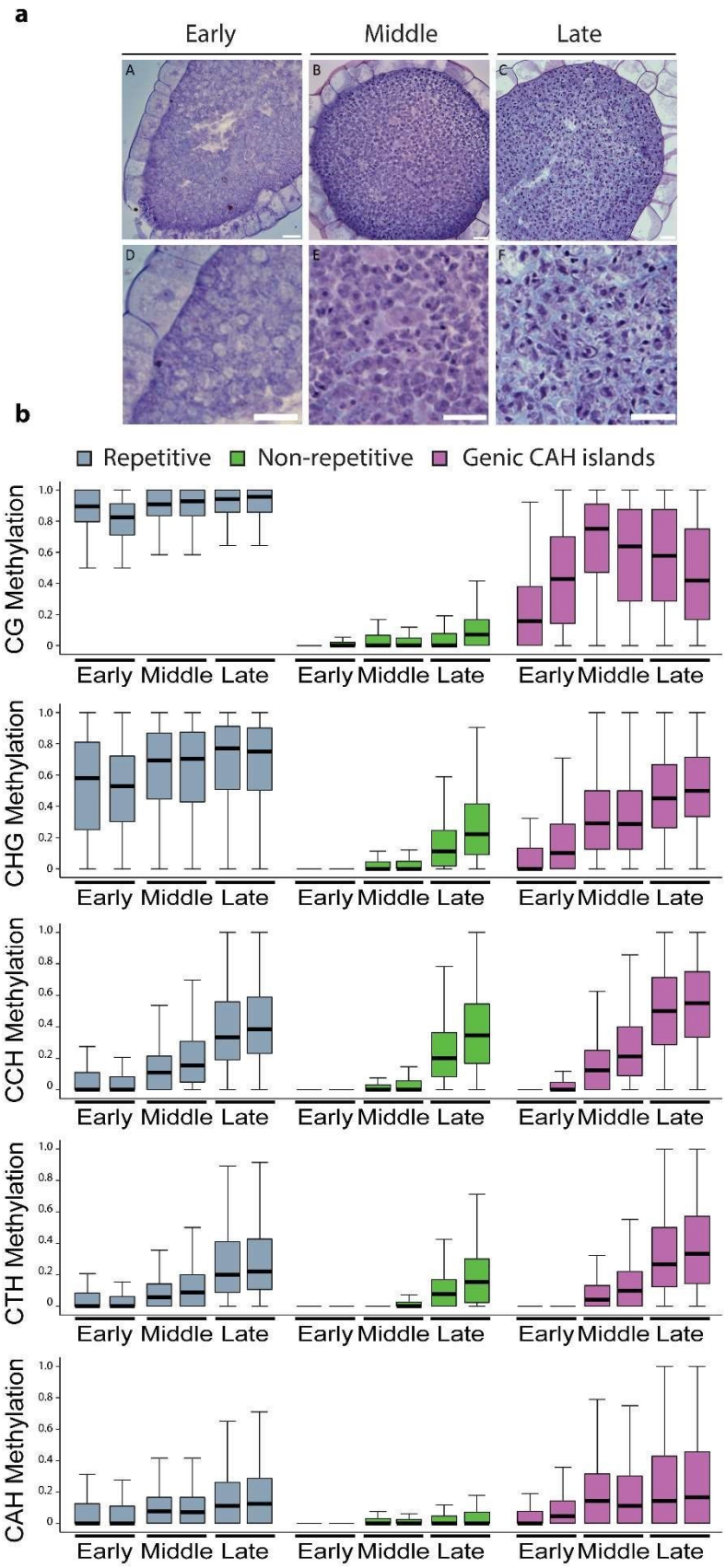
Figure 4.5 Low methylated regions are associated with genes expressed during sperm development. a, Pie charts showing LMR association with genes, transposons, both, or neither, and genic LMR association with the transcriptional start site (TSS), gene body, transcriptional termination site (TTS), or the whole gene (covered). **b**, Ends analysis of genes lacking LMRs or with LMRs for CG and CAH methylation or CHG, CTH and CCH methylation. **c**, Snapshots of genes associated with low-methylated regions (LMRs) in *M. polymorpha* sperm (indicated by magenta bars) in CHG, CCH, and CTH contexts along with antheridia gene expression. **d**, Gene expression for genes lacking or associated with LMRs. * p <0.001 K-smirnov test.

4.7 Methylation patterns are established during sperm development and are associated with the expression of specific methylases

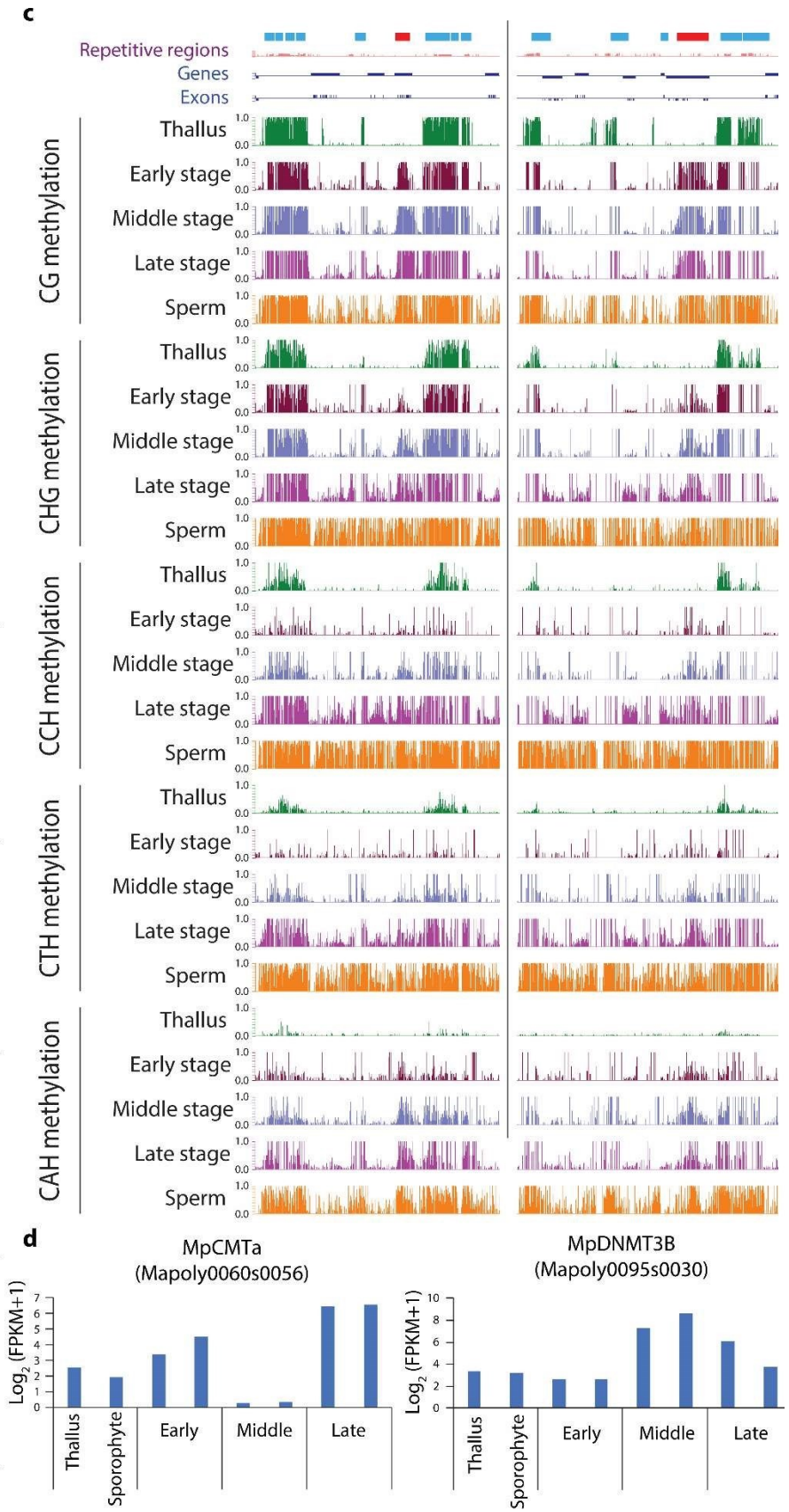
Both the blanket and targeted methylation patterns observed in *M. polymorpha* sperm are substantially different from the thallus. However, it is uncertain whether these patterns are established during sperm development or are present at an earlier stage such as in meristematic cells and perhaps are lost in thallus tissue (although it is reported that such methylation is absent from apical notches containing vegetative meristems; Schmid et al., 2018). To determine methylation changes during antheridium development, Dr. Jingyi Zhang collected individual antheridia containing developing sperm and obtained BS-seq as well as RNA-seq data from half of each antheridium. Staging was determined by sectioning and imaging the other half, and antheridia were separated into early (before spermiogenesis), middle (during early spermiogenesis) and late (during DNA compaction and tail formation) stages based on the overall phase of development, although it should be noted that some rare cells were asynchronised (Figure 4.6a). These stages were confirmed by comparing the expression of genes known to be differentially expressed during sperm development (supplementary Figure 4.7; Higo et al., 2016). Before spermiogenesis at early stages of development, antheridia methylation is already high at repetitive elements compared to the thallus but shows similar patterns across the genome, with high CG and CHG and low CHH methylation over repetitive elements and low global methylation (Figure 4.6b; Figure 4.6c). However, genic CAH islands have substantial levels of methylation, indicating that these are established in early spermiogenesis or earlier during development. During the middle stage of antheridia development, methylation in additional contexts including CAH increases over repetitive elements and the genic CAH regions, while global methylation becomes detectable to a small degree (Figure 4.6b; Figure 4.6c). During the late stages of sperm development when DNA compaction occurs, global DNA methylation also increases substantially (Figure 4.6b; Figure 4.6c). Thus, the two DNA methylation reprogramming events in *M. polymorpha* sperm development are at least partially temporally separate. First, there is methylation reinforcement over repetitive elements and a subset of genes that do not typically have methylation in the sporophyte or thallus, and second there is global methylation occurring primarily at the late stages of development.

Previous attempts to identify the methylases expressed during male sexual development indicate that most of these genes are upregulated (Bowman et al.,

2017; Schmid et al., 2018). However, given our fine-scale developmental methylation patterns and RNA-seq data, we were curious to explore the patterns of methylase expression in correlation with methylation changes. While most of the methylases displayed no discernible patterns across antheridia development (Supplementary Figure 4.8), MpDNMT3B expression increased in early spermiogenesis during hypermethylation of repetitive elements and sperm genic CAH islands (Figure 4.6d). Surprisingly, MpCMTa expression decreased during this phase, and then increased as the sperm matured (Figure 4.6d). Although I was unable to detect a pattern of expression for MpHEN1, I determined that MpPiwia and MpPiwib are expressed only during the late phase of development (Supplementary Figure 4.8). Together, these results provide indications of the underlying players and dynamics involved in the methylation patterns observed during *M. polymorpha* sperm development.



See page 93 for legend.



See following page for legend.

Figure 4.6 Sperm methylation patterns are established during sperm development in two phases that correlates with methylase expression. a, Cross sections of individual antheridia from *M. polymorpha* at different stages of sperm development. Scale bars, 20 µm. **b**, Average methylation of 50 bp windows across the *M. polymorpha* genome for CG, CHG, CCH, and CTH contexts in the thallus, developing antheridia, and sperm for repetitive regions, non-repetitive regions, and genic CAH islands. **c**, Snapshots of DNA methylation in thallus, developing antheridia, and sperm for CG, CHG, CCH, CTH, and CAH contexts. Sperm CAH methylation regions overlapping thallus CG methylation, sporophytic CAH methylation, and repetitive elements are depicted with light blue bars. Genic CAH islands are depicted with red bars. **d**, Bar graphs showing expression of the candidate 5mC methylases that may play a role during antheridium development.

4.8 Discussion

Drastic changes in chromatin architecture are often known to occur during sexual reproduction but examples of DNA methylation reprogramming in germ cells outside of mammals are only beginning to surface. Here, we have shown the remarkably dynamic and multi-faceted changes in DNA methylation that occur during sexual reproduction in the basal land plant, *M. polymorpha*. We have shown that blanket methylation, primarily in the CHG, CCH, and CTH contexts, is established late during chromatin compaction in sperm, and that regions of hypomethylation are associated with genes that are expressed during sperm development in an analogous manner to vertebrate blanket methylation. We have shown that repetitive elements and a subset of genes gain methylation that includes the CAH contexts earlier during sperm development, allowing these regions to be distinguishable from the global methylation patterns in mature sperm, and suggest that the targeted methylation of specific genes likely plays some critical role due to the strong enrichment of epigenetic and developmental functions. Additionally, I have identified *de novo* methylation resembling SLMs seen in *A. thaliana* sex cells and show that this methylation is also associated with genes. Finally, we explore the correlation between methylase expression and methylation patterns during sperm development and identify MpDNMT3B and MpCMTa as candidates for further study on sperm methylation and development. CRISPR knockout lines are currently being generated to further disseminate the underlying mechanisms of the patterns observed.

It was recently reported that, in addition to a loss of CHG methylation, PpCMT mutants gain CHH methylation at the edges of long transposons and at short transposons as a result of PpDNMT3B activity (Yaari et al., 2019). The decrease in MpCMTa expression, increase in MpDNMT3B expression, and increase in methylation at repetitive elements and the subset of genes in the middle stages of antheridia development, particularly in the CHH context, presents a tantalising parallel between these two situations. Although these patterns need to be confirmed at the protein level, questions about the interaction between CMT and DNMT3B would be raised if this interaction proves to exist during antheridia development. It was originally suggested that an increased accessibility of the chromatin in the absence of CHG methylation in PpCMT mutants allowed PpDNMT3B to methylate the CHH context (Yaari et al., 2019). However, antheridia do not lose CHG methylation within the short period that MpCMTa expression is decreased. Such competition is unlikely to be a result of H3K9 methylation binding, as similar competition is not seen with CMT2 and CMT3 mutants in flowering plants (Stroud et al., 2014). A negative interaction between CMT and DNMT3B may also explain the evolutionary loss of DNMT3B and the evolution of CMT2 in flowering plants. Investigating the competition between MpCMTa and MpDNMT3B in somatic and sperm cells by examining knock out mutants, as well as devising experiments to maintain MpCMTa expression during sperm development would be important to determine the interplay between these proteins in *M. polymorpha*. Additional experiments to understand the interacting partners of MpDNMT3B and MpCMTa would also be of interest, as it may be that binding partners are sequestered.

The lack of obvious changes in expression for genes with sperm-specific DNA methylation indicates that the role of this methylation is outside of sperm development, such as in imprinting. This is particularly likely given the absence of such methylation in the archegonia, i.e. the female egg surrounded by a single layer of somatic tissue (data not shown; Schmid et al., 2018). Such a phenomenon would be the first demonstrated outside of mammals and flowering plants (Rodrigues and Zilberman, 2015). How this would manifest in *M. polymorpha* remains to be seen, although there are specialised ‘transfer cells’ in the foot of the sporophyte that connects it with the parent female gametophyte to act as a ‘placenta’ (Shimamura, 2015). Alternatively, this methylation may be critical earlier in male development during the switch in meristem identity from the vegetative phase to the sexual phase, although why the methylation would be absent in female sexual development is unclear. How these genes are targeted is also of strong interest. It is possible that

the introns share sequence homology to regularly methylated repetitive elements and that they are targeted due to the stronger reinforcement of methylation across the genome. How such a system of controlling gene expression originally evolved is a challenge for the future. The close study of genes with targeted methylation is therefore warranted, either by comparing expression during development in methylation mutants, or in single cases such as by deleting the relevant introns. Similarly, the study of the sporophytic-specific methylated regions is currently underway, for example by examining methylation and gene expression in sporophytes lacking a functional MpDRMb.

The presence of blanket methylation in *M. polymorpha* sperm is particularly unexpected, with no other examples outside of vertebrates being previously identified. While *M. polymorpha* methylation covers a much broader range of cytosine contexts, the striking similarities with mammalian sperm methylation suggest a potentially shared role between these two organisms. The correlation of blanket methylation in *M. polymorpha* with protamine expression points to DNA compaction as the most likely function. Protamines interact with the major groove of DNA where methyl groups are presented (Balhorn, 2007), and it would be interesting to determine how methylation influences protamine interaction with DNA. Whether such a mechanism is also true in mammals would be interesting to explore. Alternatively, the global methylation could be utilised to repress gene expression, as mammalian germ cell expression and development is affected in methylation mutants (Zeng and Chen, 2019). Examining the compaction of DNA in *M. polymorpha* global methylation mutants, as well as any effect on global gene repression, is thus a priority for further study and is possible given the ability to propagate mutants asexually.

While MpCMTa is the only typical methylase that shows an increase in expression during late sperm development, it is difficult to justify its involvement in blanket methylation with its known targeting mechanism via the chromodomain (unless there is also global H3K9 methylation reprogramming or a change in the targeting mechanism of MpCMTa). Notably, however, the similar preference of CCH and CTH over CAH between blanket methylation in the sperm and for repetitive elements in the thallus indicates that the same methylase(s) may be utilised in both situations with a change in target specificity. Close examination of sperm methylation in methylase CRISPR knockout mutants will help to illuminate this situation.

Following fertilisation, the paternal genome of vertebrates is rapidly demethylated by

TET enzymes (Wu and Zhang, 2017). While *M. polymorpha* does not have a discernible version of TET, two ROS1 DNA glycosylases are present, one of which is located on the female X chromosome that is thus absent in male *M. polymorpha* and is named ROS1x (Bowman et al., 2017). Although ROS1x may be important for demethylation during female sex development, it is also a prime candidate for paternal genome demethylation following fertilisation. While such drastic DNA base-excision repair on the whole genome seems counter-intuitive due to the possibility of mutation, it should be noted that the paternal genome typically undergoes a multitude of breaks during DNA compaction in sperm to allow topoisomerases to change its structure (Laberge and Boissonneault, 2005). The distinction between CAH methylation at repetitive elements compared to the global CHG, CCH and CTH methylation across the genome may allow the removal of blanket methylation rapidly without the removal of methylation at regions that need to be constantly repressed, for example if the demethylation machinery cannot recognise CAH methylation or is inhibited by it. Alternatively, the methylation may mark these regions for reinforced silencing via histone marks to withstand demethylation following fertilisation and/or the histone-protamine exchange during sperm maturation. Such speculation should be approached with a degree of scepticism but is easily tested. Unlike other organisms, pro-nuclei are slow to fuse in *M. polymorpha* zygotes (Tetsuya Hisanaga, 2018), making this an attractive model to study DNA methylation following fertilisation and during zygote development in both wild-type to explore demethylation and in mutants, for example to examine whether paternal repetitive elements are targeted for demethylation along with the rest of the genome in a DNMT3B mutant.

Together, our work has established a myriad of dynamic methylation patterns during the sexual reproduction of *M. polymorpha* with likely new functions of DNA methylation to be determined. In particular, the mechanisms by which mosaicism could have progressed to global methylation in mammals can be explored as *M. polymorpha* provides a highly suitable model to explore these two mechanisms in parallel. Further studies of DNA methylation in *M. polymorpha*, and the relationship of DNA methylation with histone marks and chromatin during *M. polymorpha* sperm development, will undoubtedly reveal further exciting phenomena.

4.9 Materials and Methods

Plant material and growth conditions

Male and female *M. polymorpha* accessions Takaragaike-1 (Tak-1, male) and Takaragaike-2 (Tak-2, female), kindly provided by L. Dolan (Oxford University, UK), were used as wild-type. Plants were grown on plates containing 1/2× Gamborg's B5 medium with 1% agar (Sigma) under constant light at 21 °C, and thallus tissue was collected from 2-week old plants. For sperm and sporophytes, plants were transferred to jiffy pellets after 2 weeks on plates and placed in a growth chamber at 21 °C with 70% humidity under constant light with far-red irradiation for induction of sexual reproduction, as described previously (Chiyoda et al., 2008).

Antheridia extraction and sectioning

Developing antheridia were manually dissected as described previously (Higo et al., 2016). Individual antheridia were transversely cut in half using fine needles (0.5mm x 25 mm). One half was immediately frozen in liquid nitrogen, while the other half was fixed in fixation buffer (4% paraformaldehyde, PBS, pH 7.5). After a series of ethanol dehydration steps followed by resin embedding using a Technovit 7100 kit, the fixed antheridium halves were cut into 5 micron sections and stained with Toluidine blue (Feng and Dickinson, 2010). DNA and RNA were extracted simultaneously from the frozen antheridium halves using the Dynabeads™ mRNA DIRECT™ Micro Kit (Invitrogen).

Sequencing-library construction and analysis

For the thallus, sporophyte and sperm, single-end bisulphite-sequencing libraries for Illumina sequencing were constructed with Ovation Ultralow Methyl-Seq Library Systems (Nugen, 0336) and EpiTect Fast Bisulphite Conversion (Qiagen, 59802) kits according to manufacturer's instructions, except for the incorporation of two rounds of bisulphite conversion. For the antheridia samples, bisulphite sequencing library preparation and mRNA library preparation was carried out as previously described (Picelli et al., 2014; Smallwood et al., 2014). Sequencing was performed on site at the John Innes Centre using a nextseq 500. DNA methylation analysis was performed by using the miniature-sniffle-mapper developed by Dr Martin Vickers (https://github.com/martinjvickers/miniaturesniffle-mapper), which is based on Bismark (Krueger and Andrews, 2011), but with the additional ability to map ambiguous reads.

Thallus, sporophyte, and pooled antheridia RNA-seq data utilised herein has previously been published (SRR896229, SRR896223 and DRR050349, respectively; Bowman et al., 2017; Higo et al., 2016)

Transposon and gene meta-analysis (end analysis)

Transposon and gene meta-analysis was performed as described previously (Ibarra et al., 2012). Repetitive elements were merged if they occurred within 100 bp of each other, regardless of strandedness.

Determination of methylated and low-methylated regions

Thallus CG methylated regions, sperm and sporophyte CAH islands, and low-methylated regions (LMRs) were determined using methpipe (Dolzhenko and Smith, 2014) with specific cytosine contexts at the 50 bp window level. 241 genic CAH islands in sperm were determined from sperm CAH islands that overlapped genes but not thallus CG islands, sporophyte CAH islands, or repetitive elements. LMRs were determined by finding CHG, CTH and CCH low-methylated regions and keeping regions that overlapped at least two of these three groups of LMRs (excluding overhangs).

Sporophytic non-CG hyper DMRs were determined in a similar manner to Chapter Two (Walker et al, 2018). Fractional methylation in 50 bp windows for C-methyl in the sporophyte was determined. We first selected windows with C-methyl > 0.05. The selected windows were merged to generate larger methylation islands if they occurred within 200 bp, and the fractional methylation difference for each island was determined between the sporophyte and thallus in CG, CHG, and CHH contexts (CG_diff, CHG_diff, and CHH_diff). Islands were retained if they covered at least 100 bp, had significantly different levels of non-CG methylation (Fisher's exact-test P value < 0.001) and met the following criteria: CG_diff >0 and CHG_diff > 0.05 and CHH_diff > 0.1 and (CHG_diff + CHH_diff) > 0.3. This resulted in the identification of 7037 nonCG DMRs. The same criteria, except with thallus C-methyl windows and the relationship between the sporophyte and thallus reversed, gave 2 non-CG DMRs.

The list of 7,037 sporophytic non-CG DMRs was further refined by excluding those that overlap thallus CG methylated regions and/or show greater than 0.1 CG methylation in thallus to obtain 443 sporophytic islands.

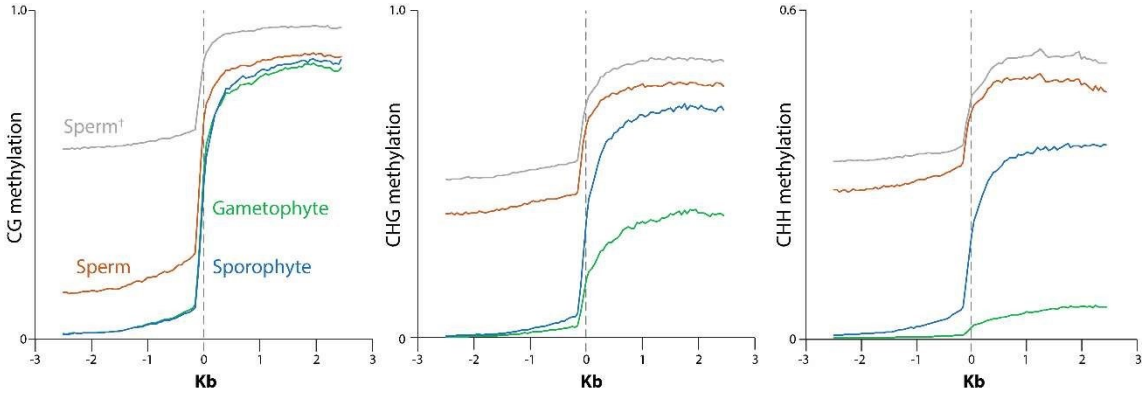
CAH islands, sporophytic islands, and LMRs were considered to overlap repetitive elements only if more than 20% of their length overlapped a repeat annotation after merging, while for genes any degree of overlap was permitted.

GO term enrichment was calculated using Panther at http://geneontology.org/ (Mi et al., 2017).

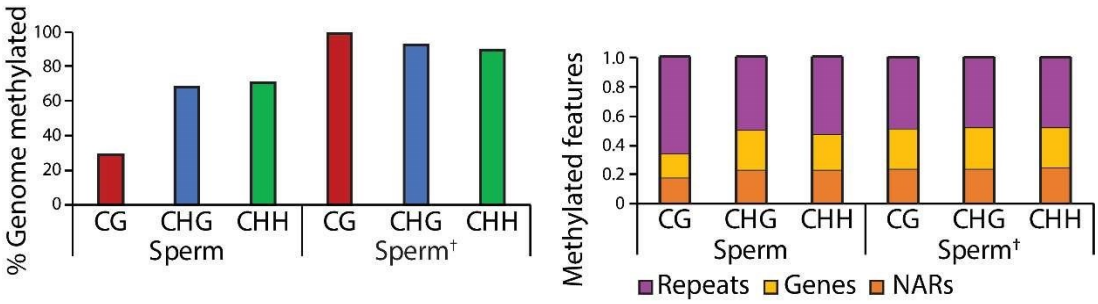
Box plots and bar graphs

All box plots follow this format: each box encloses the middle 50% of the distribution, the horizontal line marks the median, and vertical lines mark the minimum and maximum values falling within 1.5 times the height of the box. All boxes were generated with 50 bp windows with at least 5 informative sequenced cytosines, unless otherwise stated. Bar graphs were similarly generated with 50 bp windows with at least 5 informative sequenced cytosines, unless otherwise stated. For repetitive elements, only 50 bp windows with CG methylation greater than 0.5 were considered. For non-annotated regions, 50 bp windows that were within 500 bp of repetitive elements were excluded.

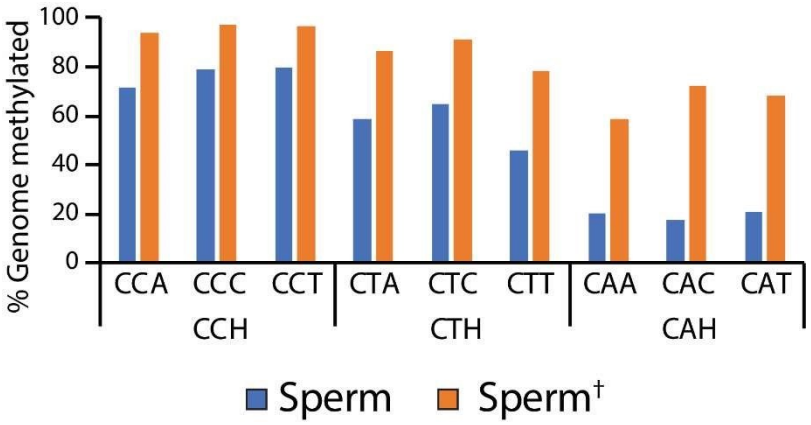
4.10 Supplementary Figures



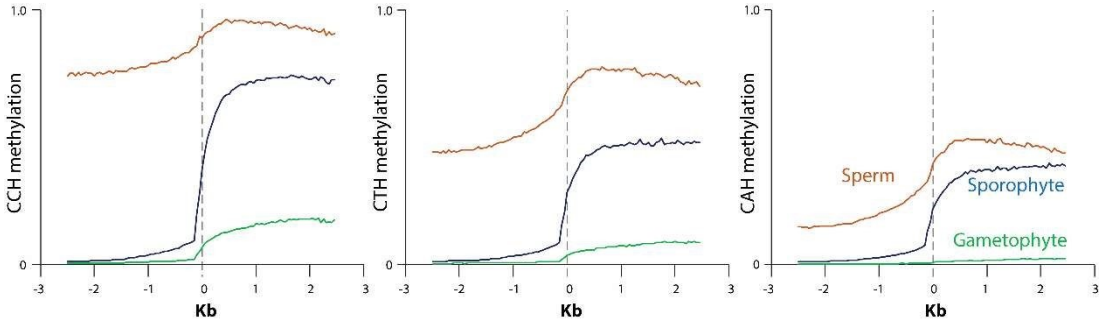
Supplementary Figure 4.1. Ends analysis of repetitive regions in thallus, sporophyte, sperm, and published sperm data (denoted by a dagger symbol; Schmid et al., 2018). Repetitive regions were merged if they were within 100 bp, aligned at one end, and average methylation levels in the CG, CHG, or CHH context for each 100 bp interval are plotted. The dashed line at zero represents the point of alignment.



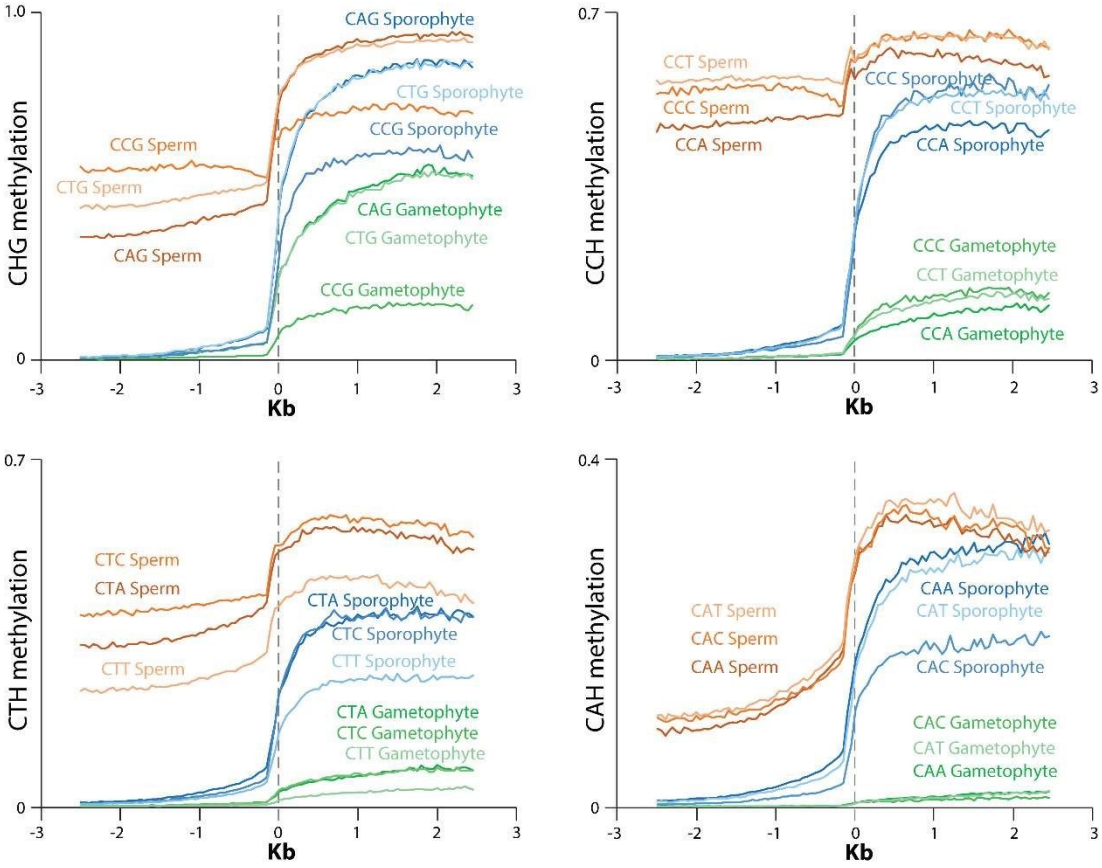
Supplementary Figure 4.2. Bar charts showing the percentage of 50 bp windows across the *M. polymorpha* genome with more than 10% methylation and the association of 50 bp windows that have more than 10% methylation exclusively with either repetitive elements, genes, or non-annotated regions (NARs) in our sperm and published sperm data (denoted by a dagger symbol; Schmid et al., 2018)



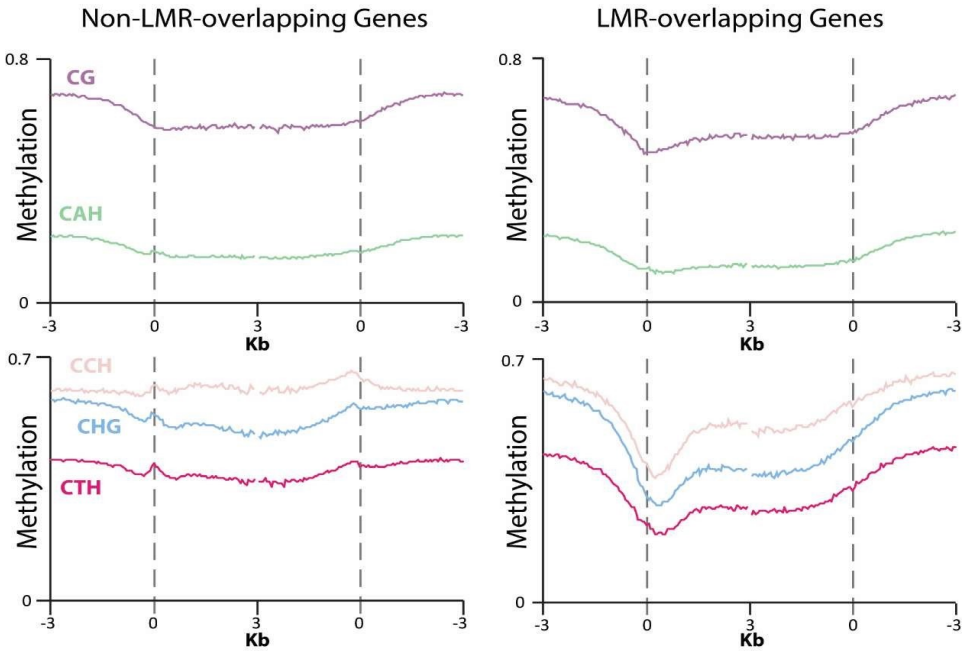
Supplementary Figure 4.3. Bar graph showing the percentage of 50 bp windows above 10% methylation for specific CHH contexts in our sperm data and published sperm data (denoted by a dagger symbol; Schmid et al., 2018).



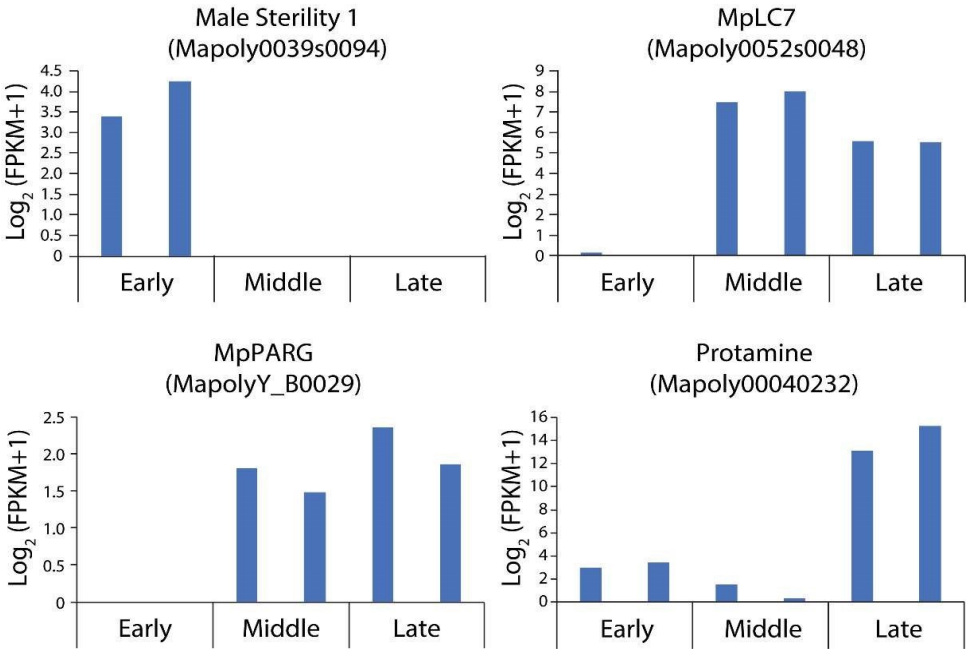
Supplementary Figure 4.4. Ends analysis of repetitive regions in thallus, sporophyte, and sperm in the CCH, CTH, or CAH context.



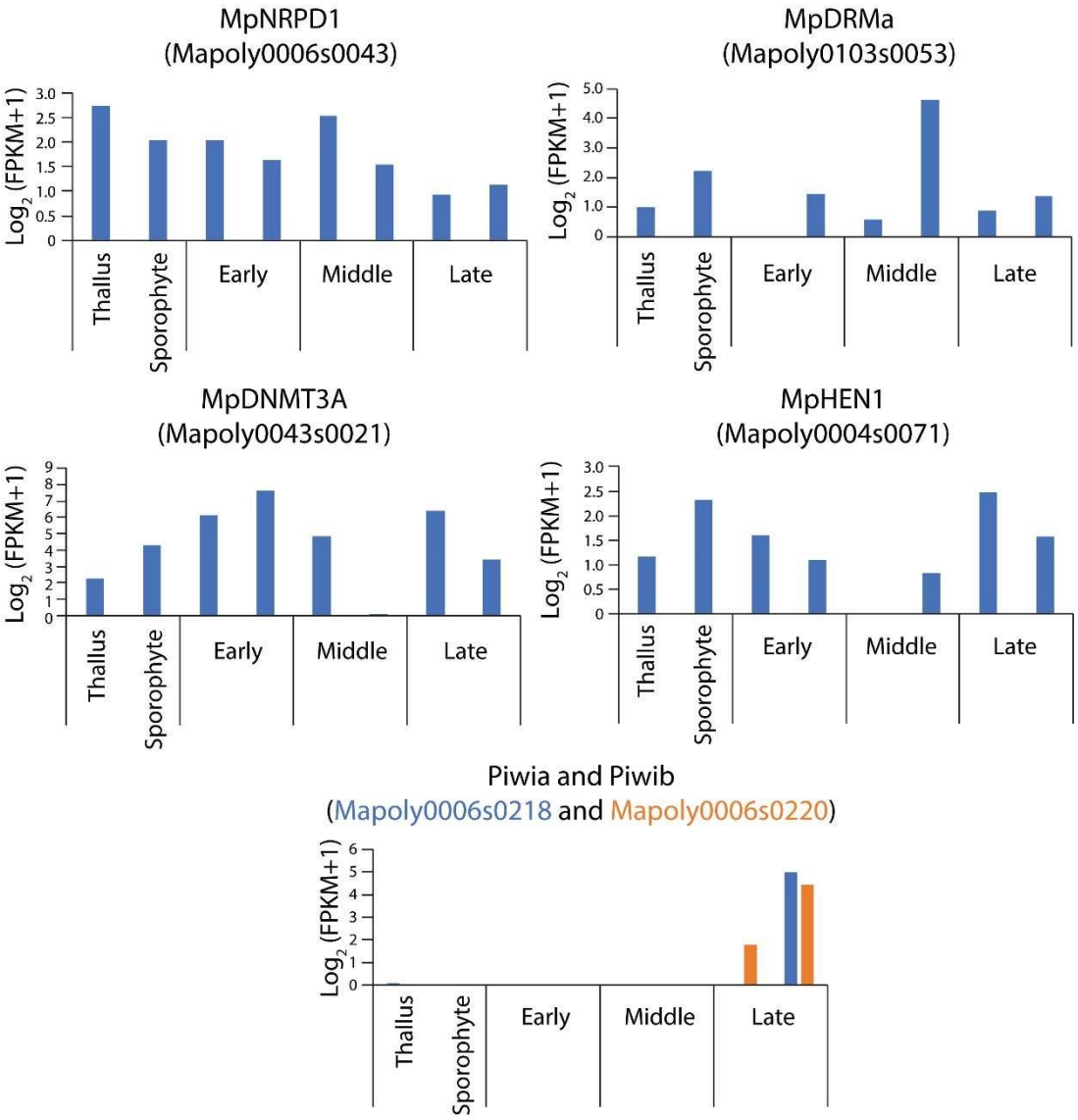
Supplementary Figure 4.5. Ends analysis of repetitive regions in thallus, sporophyte, and sperm in specific CHG, CCH, CTH, and CAH contexts.



Supplementary Figure 4.6. Ends analysis of genes lacking LMRs or with LMRs for CG and CAH methylation or CHG, CTH and CCH methylation in published M. polymorpha sperm methylation data (Schmid et al., 2018).



Supplementary Figure 4.7. RNA expression data during the three stages of antheridia development for four genes with previously tested expression levels during antheridium development via RNA in situ hybridisation (Higo et al., 2016). Male Sterility 1 is only expressed early during antheridium development. MpLC7 and MpPARG are expressed during spermiogenesis into late development, with the former expressed more strongly earlier in development and the latter expressed more strongly during late development. Protamine is expressed late during DNA compaction.



Supplementary Figure 4.8. RNA expression data in the thallus, sporophyte, and three stages of antheridia development for *M. polymorpha* methylases with no discernible patterns of expression during antheridia development, as well as expression patterns of NRPD1 and HEN1, Piwia and Piwib. Only the best fitting transcripts that match the annotation are considered.

Chapter Five
Main Discussion

"I spent all afternoon collecting sperm and managed to get 15 mL but now I'm exhausted and I think I got some on my elbow!" – James Walker, 2018

In this thesis, I describe male sexual-lineage-specific hypermethylation at a subset of RdDM loci (i.e. canonical SLHs) accompanied by RdDM-driven *de novo* methylation associated with genes (i.e. SLMs) in *A. thaliana*. I show that loss of this *de novo* methylation affects gene repression and splicing, with the latter disrupting meiosis. I explore the patterns underlying these specific methylation patterns by examining preliminary sRNA data from *A. thaliana* meiocytes, finding evidence that the *de novo* methylation is a result of sRNA produced at canonical SLHs. I suggest that CLSY3 and possibly CLSY4 are responsible for the RdDM activity at these hypermethylated loci, with a potential role of the tapetum as a source of CLSY3derived sRNAs for the sexual lineage. I further describe the methylation patterns during sexual reproduction in *M. polymorpha*, identifying a blanket methylation phenomenon that is established late during sperm development. This global methylation is accompanied by specific reinforcement of methylation at transposons and targeted *de novo* methylation of genes, although the latter are notably different from SLMs in *A. thaliana*. Additionally, I identify potential methylases involved in these processes. I also characterise gene-associated *de novo* methylation in the sporophytes of *M. polymorpha*, the stage of development undergoing meiosis. This *de novo* methylation more closely resembles SLMs in *A. thaliana* and is accompanied by transposon hypermethylation and upregulation of MpDRMb, suggesting the involvement of RdDM. Together, I show that DNA methylation during the sexual development of plants is much more dynamic than was previously speculated, with roles to play in transposon and gene regulation.

5.1 Importance of SLMs associated with tRNA genes

In Chapter Two, I showed that a subset of SLMs clearly overlap numerous tRNA genes in the *A. thaliana* genome. Like other SLMs, sRNA mapping to pre-tRNA genes in meiocytes have a significant number of mismatches, and their origins can be traced back to meiotic RdDM loci and particularly canonical SLHs (data not shown). It is known that tRNA fragments can target gypsy elements (and copia elements) for repression (Martinez, 2017). However, as canonical SLHs are depleted for gypsy elements (and copia elements are not particularly enriched), the link between pre-tRNA gene SLMs and transposon silencing is unclear. sRNA produced from canonical SLHs themselves could be beneficial for supressing gypsy and/or copia elements, with pre-tRNA gene methylation being a by-product of this process. It was recently shown that pollen easiRNAs (21 – 22 nt sRNAs that repress transposons; He et al., 2015) are dependent on POLIV activity (Martinez et al., 2018). Close analysis of tRNA-associated sRNAs in meiocytes and pollen as well as transposon repression in various RdDM mutants would be interesting to pursue.

5.2 Importance of *de novo* methylation in meiosis

The presence of gene-targeted methylation resembling RdDM activity in the meiotic phase of development in both *A. thaliana* and *M. polymorpha* suggests an importance of this phenomenon during meiosis. Although we have shown that gene-associated RdDM activity in male sex cells is important for sex cell development in *A. thaliana* (Chapter Two; Walker et al., 2018) and may be important for such gene regulation in other plant species, it is likely that the prime purpose of mismatched sRNA targeting from RdDM activity in these cells is to identify related transposons and target them for repression. Thus, while it may be tempting to only examine the effects on gene function when considering sex-cell mismatched sRNA targeting in plants in the future, it will be important to also consider the detection and repression of transposons (and other duplications or chromosomal rearrangements) if we are to identify the underlying evolution of this system.

An interesting parallel to the *de novo* methylation described in this thesis is the meiotic defence systems determined in certain fungi. A homology-based genome defence system called Repeat-Induced Point mutation (RIP) causes permanent silencing of repeated elements in the fungus *Neurospora crassa* (a species without many repeat elements) by causing G:C to A:T mutations during the premeiotic stage of the sexual cycle (Irelan and Selker, 1996). A related process to RIP known as MIP (Methylation Induced Pre-meiotically) methylates but does not mutate repeats in other fungal species (interestingly MIP is then retained in the vegetative phase of

development; Amselem et al., 2015; Irelan and Selker, 1996). An indirect consequence of these phenomena is the prevention of gene duplication as such elements will be targeted for repression and elimination in the same way that transposons are. The heightened detection of repeats during meiosis and the mechanism of shared sequence homology detection is strikingly familiar with the methylation in the male sexual lineage of *A. thaliana*.

5.3 A mechanistic understanding of sexual-lineage-specific methylation

This thesis has provided evidence for the mechanisms underlying specific RdDM activity and SLM establishment in *A. thaliana*. However, significant aspects of this system remain to be determined. Prime among these is the method by which relaxed targeting of RdDM is allowed in meiocytes. Pursuing the answers to these questions, for example by examining the RdDM components that are upregulated in meiocytes, will undoubtedly aid understanding of the dynamics of RdDM. Preliminary work has established that *ago4*, *ago5*, *ago6*, and *ago9* single mutant plants do not completely lose SLMs, suggesting a redundancy of these genes. Similarly, preliminary work from *drm1* and *drm2* single mutants show a presence of SLMs in the sperm (data not shown), which is perhaps expected given the strong expression of DRM1 in meiocytes (Chapter Three). Interestingly, *drm1drm2* double mutant *A. thaliana* plants don't lose sRNA at clsy3-dependent loci despite a loss of methylation (Zhou et al., 2018), uncoupling the known reinforcement of RdDM. Thus, while we have shown the loss of methylation in *drm1drm2* double mutants and an associated mis-regulation of expression for some genes, it is possible that the sRNAs are still present in these mutants and are able to repress a subset of other SLM-associated genes. Examining gene and transposon expression in *rdr2* mutants would be of interest to compare the influence of sRNA and DNA methylation.

5.4 Evolution of SLMs

While *M. polymorpha* does not have an immediately identifiable structure that resembles the tapetum of flowering plants (Wallace et al., 2011), the meiocytes do have sister cells (elater mother cells) that could provide support in a similar way (Shimamura, 2015). These sister cells also undergo a form of programmed cell death (like the tapetum of flowering plants) to form elaters (dried springy cells) that aid spore dispersal (Shimamura, 2015). Thus, one could speculate that these cells could act as a source of sRNA. Interestingly, the presence of only one CLASSY orthologue in *M. polymorpha* correlates with the universal transposon hypermethylation in the sporophyte rather than the hypermethylation of specific

transposons in the male sex cells of *A. thaliana* utilising specialised CLASSY proteins. This highlights the question of why such specificity exists in *A. thaliana* and other organisms. *M. polymorpha* is known for having few instances of gene duplications and polyploidisation events during its evolution (Berrie, 1960; Villarreal A et al., 2016). Gene-capture events and other instances of duplicated sequences may also be less tolerated in this species, removing the need for selective RdDM activity. However, further work is needed to establish that the *de novo* methylated regions in the *M. polymorpha* sporophyte are indeed caused by RdDM, as they are in *A. thaliana*, and whether they are targeted in trans.

Preliminary results suggest the presence of *de novo* methylation over genes in the meiocytes and sperm of *Zea mays*, and that this methylation correlates with mismatched phasiRNAs which are produced in the maize tapetum (data not shown; Zhai et al., 2015). If true, this would indicate a convergence of sRNA targeted methylation of genes in the male sexual lineage of plants. *A. thaliana* doesn't harbour phasiRNAs (Xia et al., 2019), while the CLSY3 orthologue in maize (RMR1) is thought to play a more central role in RdDM (Hale et al., 2007). The existence of gene-targeted methylation in maize is perhaps slightly surprising given the high occurrence of gene-capture events by helitrons in the maize genome (Barbaglia et al., 2012). However, typical RdDM activity again seems to be reduced in these sex cells and few instances of transposon hypermethylation resembling RdDM activity could be detected (data not shown), suggesting a similar phenomenon of reduced sRNA availability as we have proposed in the male meiocytes of *A. thaliana*. While further work is needed to confirm these findings, it would nonetheless be interesting to explore the existence and mechanisms underlying gene-targeted methylation in the sex cells of eudicots and other plants where phasiRNAs have been detected (Xia et al., 2019) while CLSY3 is likely to act in a similar way to CLSY3 in *A. thaliana*.

5.5 Female sexual-lineage-specific methylation

While there is evidence for soma-germline sRNA transfer in the male sexual lineage, it is important to consider the potential existence of canonical SLHs and SLMs in the female sexual lineage as it could have implications for fertilisation and embryo development. While none of the genes associated with SLMs have been associated with imprinting, a gene important for seed dormancy which is paternally repressed (ALLANTOINASE) is controlled by RdDM in its promoter region (Iwasaki et al., 2019) and this area of the genome is also enriched for canonical SLHs (data not shown). Preliminary results examining central cells heterozygous for *dme* (and thus

with reduced DNA demethylation) and wild-type gynoecium data (kindly provided by Dr. Jean-Philippe Vielle Caldaza) shows a weak but detectable presence of a small number of SLMs and canonical SLHs (data not shown). While this low level and occurrence of detectable methylation in the female sexual lineage might reflect DME activity (in the central cell) or somatic tissue methylation states (in the gynoecium), it could also reflect a difference in RdDM dynamics. CLSY4 expression can be detected in *A. thaliana* egg cells but CLY3 (and CLSY1/2) expression is absent from eggs, synergid, and central cells (Wuest et al., 2010), and so CLSY4 may play a role in producing a portion of SLMs in the female sexual lineage while clsy3-dependent methylation may be absent. Examining the correlation between clsy3-dependent and clsy4-dependent RdDM loci and the female sexual lineage methylation profile would be of interest. It also remains to be seen whether there are any female-specific methylated regions not present in the male sexual lineage. Examining CLSY protein expression patterns with reporter lines and methylation in various RdDM mutants will allow a clearer comparison to be made between male and female sexual methylation dynamics.

5.6 Concluding remarks

It is clear from this thesis and the evidence emerging from others that DNA methylation is remarkably dynamic during sexual development across land plants. The association of this reinforced and *de novo* methylation with transposons and importantly with genes indicate that this epigenetic mark is particularly relied upon for transcriptional regulation during this phase of development. The involvement of sRNAs in directing methylation and the diversity of sRNA populations during sexual reproduction across eukaryotes highlights the potential diversity of genome targeting and an exploration into transcriptional regulation by sRNAs and DNA methylation in other species is an exciting prospect. The dynamics of chromatin modification during sexual development have also been highlighted in this thesis by the dual targeted and blanket methylation of *M. polymorpha* sperm, phenomena that I was serendipitous enough to find. Together, I hope this analysis has set a foundation for future discoveries of the roles of DNA methylation in transcriptional regulation and other processes.

List of abbreviations

4mC 4-methylcytosine

5mC 5-methylcytosine

6mA 6-methyladenosine

AGO4 ARGONAUTE 4 (AT2G27040)

AGO6 ARGONATURE 6 (AT2G32940) AGO9

ARGONAUTE 9 (AT5G21150)

bp nucleotide (base pairs)

CLSY1 (Chr38) CLASSY1 (AT3G42670)

CLSY2 (Chr42) CLASSY2 (AT5G20420)

CLSY3 (Chr31) CLASSY 3 (AT1G05490)

CLSY4 (Chr40) CLASSY 4 (AT3G24340)

CMT2 CHROMOMETHYLASE 2 (AT4G19020)

CMT3 CHROMOMETHYLASE 3 (AT1G69770)

DCL3 DICLER-LIKE 3 (AT3G43920)

DCL3 DICLER-LIKE 4 (AT5G20320)

DME DEMETER (AT5G04560)

DML2 DEMETER-LIKE 2 (AT3G10010)

DML3 DEMETER-LIKE 3 (AT4G34060)

DMR Differentially Methylated Region

DNMT DNA (cytosine-5)-methyltransferase

DRD1 DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (AT2G16390)

DRM1 DOMAINS REARRANGED METHYLTRANSFERASE 1 (AT5G15380)

DRM2 DOMAINS REARRANGED METHYLTRANSFERASE 1(AT5G14620)

FPKM fragments per kilobase per million reads

HEN1 HUA ENHANCER 1 (AT4G20910)

kb kilobases

LMR Low-methylated region

LTR Long Terminal Repeat

MET1 METHYLTRANSFERASE 1 (AT5G49160)

MIP methylation induced pre-meiotically

MpCMTa CHROMOMETHYLASE a (Mapoly0060s0056)

MpCMTb CHROMOMETHYLASE b (Mapoly0091s0051)

MpDNMT3A *M. polymorpha* DNA (cytosine-5)-methyltransferase 3A (Mapoly0043s0021)

MpDNMT3B *M. polymorpha* DNA (cytosine-5)-methyltransferase 3B (Mapoly0095s0030)

MpDRMa *M. polymorpha* DOMAINS REARRANGED METHYLTRANSFERASE a (Mapoly0103s0053)

MpDRMb *M. polymorpha* DOMAINS REARRANGED METHYLTRANSFERASE b (Mapoly0109s0015)

MpHEN1 *M. polymorpha* HUA ENHANCER 1 (Mapoly0004s0071)

MpLC7 *M. polymorpha* LIGHT CHAIN 7(Mapoly0052s0048)

MpMET1 METHYLTRANSFERASE 1 (Mapoly0038s0027)

MpPACRG *M. polymorpha* PARKIN CO-REGULATED GENE (MapolyY_B0029)

MpPiwia *M. polymorpha* Piwi a (Mapoly0006s0218)

MpPiwib *M. polymorpha* Piwi b (Mapoly0006s0220)

MPS1/PRD2 MULTIPOLAR SPINDLE 1/PUTATIVE RECOMBINATION INITIATION DEFECTS 2 (AT5G57880)

nt nucleotides

POLIV RNA POLYMERASE IV

POLV RNA POLYMERASE V

PpCMT *P. patens* CHROMOMETHYLASE

PpDNMT3B *P. patens* DNA (cytosine-5)-methyltransferase

RdDM RNA-directed DNA Methylation

RDR2 RNA-DEPENDENT RNA POLYMERASE 2 (AT4G11130) RDR6

RNA-DEPENDENT RNA POLYMERASE 6 (AT3G49500)

RIP Repeat-induced point mutation

RMR1 REQUIRED TO MAINTAIN REPRESSION (Zm00001d038113)

ROS1/DML1 REPRESSOR OF SILENCING1/DEMETER-LIKE 1 (AT2G36490)

RPKM reads per kilobase per million reads

SDSA Synthesis Dependent Strand Annealing

SHH1 SAWADEE HOMEODOMAIN HOMOLOG 1 (AT1G15215)

SLH Sexual-lineage-specific hypermethylation

SLM Sexual-lineage-specific methylation

SUVH SUPPRESSOR OF VARIEGATION 3-9 HOMOLOG

Tak Takaragaike

TES Transcription end site

TET Ten-eleven translocation methylcytosine dioxygenase

tRFs tRNA fragments

TSS Transcription start site

UHRF1 Ubiquitin-like, containing PHD and RING finger domains, 1

VIM1 VARIANT IN METHYLATION 1 (AT1G57820)

VIM2 VARIANT IN METHYLATION 2 (AT1G66050)

VIM3 VARIANT IN METHYLATION 3 (AT5G39550)

VIM4 VARIANT IN METHYLATION 4 (AT1G66040) VIM5

VARIANT IN METHYLATION 5 (AT1G57800)

References

1. Ambrosi, C., Manzo, M., and Baubec, T. (2017). Dynamics and Context-Dependent Roles of DNA Methylation. Journal of Molecular Biology 429, 1459-1475.

2. Amselem, J., Lebrun, M.-H., and Quesneville, H. (2015). Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics 16, 141.

3. Aravind, L., Burroughs, A.M., Zhang, D., and Iyer, L.M. (2014). Protein and DNA modifications: evolutionary imprints of bacterial biochemical diversification and geochemistry on the provenance of eukaryotic epigenetics. Cold Spring Harb Perspect Biol 6, a016063.

4. Balhorn, R. (2007). The protamine family of sperm nuclear proteins. Genome Biol 8, 227.

5. Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909-912.

6. Barbaglia, A.M., Klusman, K.M., Higgins, J., Shaw, J.R., Hannah, L.C., and Lal, S.K. (2012). Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics 190, 965-975.

7. Bargsten, J.W., Folta, A., Mlynárová, L., and Nap, J.-P. (2013). Snf2 Family Gene Distribution in Higher Plant Genomes Reveals DRD1 Expansion and Diversification in the Tomato Genome. PLOS ONE 8, e81147.

8. Baubec, T., Finke, A., Scheid, O.M., and Pecinka, A. (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in *Arabidopsis*. EMBO reports 15, 446-452.

9. Baucom, R.S., Estill, J.C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J.-M., Westerman, R.P., Sanmiguel, P.J., and Bennetzen, J.L. (2009). Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS genetics 5, e1000732-e1000732.

10. 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.

11. Berrie, G.K. (1960). The Chromosome Numbers of Liverworts (Hepadcae and Anthocerotae). Transactions of the British Bryological Society 3, 688-705.

12. Bestor, T.H. (1999). Sex brings transposons and genomes into conflict. Genetica 107, 289-295.

13. Bewick, A.J., Ji, L., Niederhuth, C.E., Willing, E.M., Hofmeister, B.T., Shi, X., Wang, L., Lu, Z., Rohr, N.A., Hartwig, B., et al. (2016). On the origin and evolutionary consequences of gene body DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 113, 9111-9116.

14. Bilichak, A., Ilnystkyy, Y., Hollunder, J., and Kovalchuk, I. (2012). The progeny of *Arabidopsis thaliana* plants exposed to salt exhibit changes in DNA methylation, histone modifications and gene expression. PloS one 7, e30515.

15. Borg, M., and Berger, F. (2015). Chromatin remodelling during male gametophyte development. The Plant Journal 83, 177-188.

16. Borges, F., Parent, J.S., van Ex, F., Wolff, P., Martinez, G., Kohler, C., and Martienssen, R.A. (2018). Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. Nat Genet 50, 186-192.

17. Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., and Colot, V. (2017). DNA methylation dynamics during early plant life. Genome Biology 18, 179.

18. Bowman, J.L., Araki, T., and Kohchi, T. (2016). *Marchantia*: Past, Present and Future. Plant Cell Physiol 57, 205-209.

19. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S.,
Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. Cell 171, 287304.e215.

20. Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., OnishiSeebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B., et al. (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell 55, 277-290.

21. Burghgraeve, N., Simon, S., Barral, S., Fobis-Loisy, I., Holl, A.-C., Poniztki, C., Schmitt, E., Vekemans, X., and Castric, V. (2018). Base-pairing requirements for small RNAmediated gene silencing of recessive self-incompatibility alleles in *Arabidopsis* halleri. bioRxiv, 370239.

22. Calarco, Joseph P., Borges, F., Donoghue, Mark T.A., Van Ex, F., Jullien, Pauline E., Lopes, T., Gardner, R., Berger, F., Feijó, José A., Becker, Jörg D., et al. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. Cell 151, 194-205.

23. Casas, E., and Vavouri, T. (2014). Sperm epigenomics: challenges and opportunities. Front Genet 5, 330-330.

24. Catoni, M., Griffiths, J., Becker, C., Zabet, N.R., Bayon, C., Dapp, M., LiebermanLazarovich, M., Weigel, D., and Paszkowski, J. (2017). DNA sequence properties that predict susceptibility to epiallelic switching. The EMBO Journal 36, 617-628.

25. Chelysheva, L., Gendrot, G., Vezon, D., Doutriaux, M.-P., Mercier, R., and Grelon, M. (2007). Zip4/Spo22 is required for class I CO formation but not for synapsis completion in *Arabidopsis thaliana*. PLoS genetics 3, e83-e83.

26. Chen, C., Farmer, A.D., Langley, R.J., Mudge, J., Crow, J.A., May, G.D., Huntley, J.,
Smith, A.G., and Retzel, E.F. (2010). Meiosis-specific gene discovery in plants: RNASeq applied to isolated *Arabidopsis* male meiocytes. BMC plant biology 10, 280.

27. Chiyoda, S., Ishizaki, K., Kataoka, H., Yamato, K.T., and Kohchi, T. (2008). Direct transformation of the liverwort *Marchantia polymorpha* L. by particle bombardment using immature thalli developing from spores. Plant Cell Reports 27, 1467.

28. Coleman-Derr, D., and Zilberman, D. (2012). Deposition of Histone Variant H2A.Z within Gene Bodies Regulates Responsive Genes. PLOS Genetics 8, e1002988.

29. Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., and Lyko, F. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Developmental cell 29, 102-111.

30. De Muyt, A., Pereira, L., Vezon, D., Chelysheva, L., Gendrot, G., Chambon, A., LainéChoinard, S., Pelletier, G., Mercier, R., Nogué, F., et al. (2009). A

High Throughput Genetic Screen Identifies New Early Meiotic Recombination Functions in *Arabidopsis thaliana*. PLOS Genetics 5, e1000654.

31. De Veylder, L., Larkin, J.C., and Schnittger, A. (2011). Molecular control and function of endoreplication in development and physiology. Trends Plant Sci 16, 624-634.

32. Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. Nature Reviews Genetics 20, 417-431.

33. Ding, Y., Wang, X., Su, L., Zhai, J., Cao, S., Zhang, D., Liu, C., Bi, Y., Qian, Q., Cheng, Z., et al. (2007). SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice. The Plant cell 19, 9-22.

34. Dolzhenko, E., and Smith, A.D. (2014). Using beta-binomial regression for highprecision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments. BMC Bioinformatics 15, 215.

35. Dowen, R.H., Pelizzola, M., Schmitz, R.J., Lister, R., Dowen, J.M., Nery, J.R., Dixon, J.E., and Ecker, J.R. (2012). Widespread dynamic DNA methylation in response to biotic stress. Proceedings of the National Academy of Sciences 109, E2183-E2191.

36. Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell 18, 1166-1176.

37. Feng, S., and Jacobsen, S.E. (2011). Epigenetic modifications in plants: an evolutionary perspective. Curr Opin Plant Biol 14, 179-186.

38. Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic Reprogramming in Plant and Animal Development. Science 330, 622-627.

39. Feng, X., and Dickinson, H.G. (2010). Tapetal cell fate, lineage and proliferation in the *Arabidopsis* anther. Development 137, 2409-2416.

40. Feng, X., Zilberman, D., and Dickinson, H. (2013). A Conversation across Generations: Soma-Germ Cell Crosstalk in Plants. Developmental Cell 24, 215-225.

41. Gage, M.J.G. (2012). Complex sperm evolution. Proceedings of the National Academy of Sciences 109, 4341-4342.

42. Godwin, J.L., Vasudeva, R., Michalczyk, Ł., Martin, O.Y., Lumley, A.J., Chapman, T., and Gage, M.J.G. (2017). Experimental evolution reveals that sperm competition intensity selects for longer, more costly sperm. Evol Lett 1, 102-113.

43. Goodrich, R.J., Anton, E., and Krawetz, S.A. (2013). Isolating mRNA and small noncoding RNAs from human sperm. Methods Mol Biol 927, 385-396.

44. Gouil, Q., and Baulcombe, D.C. (2016). DNA Methylation Signatures of the Plant Chromomethyltransferases. PLOS Genetics 12, e1006526.

45. Grunewald, S., Paasch, U., Glander, H.J., and Anderegg, U. (2005). Mature human spermatozoa do not transcribe novel RNA. Andrologia 37, 69-71.

46. Hale, C.J., Stonaker, J.L., Gross, S.M., and Hollick, J.B. (2007). A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize. PLoS biology 5, e275.

47. Hardcastle, T.J., Müller, S.Y., and Baulcombe, D.C. (2018). Towards annotating the plant epigenome: the *Arabidopsis thaliana* small RNA locus map. Scientific Reports 8, 6338.

48. He, H., Yang, T., Wu, W., and Zheng, B. (2015). Small RNAs in pollen. Sci China Life Sci 58, 246-252.

49. He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. Cell Res 21, 442-465.

50. He, Y., and Ecker, J.R. (2015). Non-CG Methylation in the Human Genome. Annu Rev Genomics Hum Genet 16, 55-77.

51. Henikoff, S., and Smith, M.M. (2015). Histone variants and epigenetics. Cold Spring Harb Perspect Biol 7, a019364.

52. Higgins, J.D., Armstrong, S.J., Franklin, F.C., and Jones, G.H. (2004). The *Arabidopsis* MutS homolog AtMSH4 functions at an early step in recombination: evidence for two classes of recombination in *Arabidopsis*. Genes Dev 18, 2557-2570.

53. Higo, A., Niwa, M., Yamato, K.T., Yamada, L., Sawada, H., Sakamoto, T., Kurata, T., Shirakawa, M., Endo, M., Shigenobu, S., et al. (2016). Transcriptional Framework of Male Gametogenesis in the Liverwort *Marchantia polymorpha* L. Plant Cell Physiol 57, 325-338.

54. Hollister, J.D., and Gaut, B.S. (2007). Population and evolutionary dynamics of Helitron transposable elements in *Arabidopsis thaliana*. Mol Biol Evol 24, 25152524.

55. Hollister, J.D., and Gaut, B.S. (2009). Epigenetic silencing of transposable elements:
a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome research 19, 1419-1428.

56. Honys, D., and Twell, D. (2003). Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol 132, 640-652.

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

58. Hsieh, P.-H., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R.L., Zilberman, D., and Feng, X. (2016). *Arabidopsis* male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues. Proceedings of the National Academy of Sciences 113, 15132-15137.

59. 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.

60. Huang, J., Wang, C., Wang, H., Lu, P., Zheng, B., Ma, H., Copenhaver, G.P., and Wang, Y. (2019). Meiocyte-Specific and AtSPO11-1–Dependent Small RNAs and Their Association with Meiotic Gene Expression and Recombination. The Plant Cell 31, 444-464.

61. Huff, J.T., and Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell 156, 1286-1297.

62. 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.

63. Ikeda, Y., Nishihama, R., Yamaoka, S., Arteaga-Vazquez, M.A., Aguilar-Cruz, A., Grimanelli, D., Pogorelcnik, R., Martienssen, R.A., Yamato, K.T., Kohchi, T., et al. (2018). Loss of CG Methylation in *Marchantia polymorpha* Causes Disorganization of Cell Division and Reveals Unique DNA Methylation Regulatory Mechanisms of Non-CG Methylation. Plant and Cell Physiology 59, 2421-2431.

64. Irelan, J.T., and Selker, E.U. (1996). Gene silencing in filamentous fungi: RIP, MIP and quelling. Journal of Genetics 75, 313-324.

65. Ishizaki, K., Nishihama, R., Yamato, K.T., and Kohchi, T. (2016). Molecular Genetic Tools and Techniques for *Marchantia polymorpha* Research. Plant and Cell Physiology 57, 262-270.

66. Iwasaki, M., Hyvärinen, L., Piskurewicz, U., and Lopez-Molina, L. (2019). Noncanonical RNA-directed DNA methylation participates in maternal and environmental control of seed dormancy. eLife 8, e37434.

67. 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.

68. Jeltsch, A. (2001). The Cytosine N4-Methyltransferase M.PvuII Also Modifies Adenine Residues. In Biological Chemistry, pp. 707.

69. Jiang, H., Wang, F.-F., Wu, Y.-T., Zhou, X., Huang, X.-Y., Zhu, J., Gao, J.-F., Dong, R.B., Cao, K.-M., and Yang, Z.-N. (2009). MULTIPOLAR SPINDLE 1 (MPS1), a novel coiled-coil protein of *Arabidopsis thaliana*, is required for meiotic spindle organization. The Plant Journal 59, 1001-1010.

70. Jiang, J., Zhang, Z., and Cao, J. (2013). Pollen wall development: the associated enzymes and metabolic pathways. Plant Biol (Stuttg) 15, 249-263.

71. Johnson, L.M., Du, J., Hale, C.J., Bischof, S., Feng, S., Chodavarapu, R.K., Zhong, X., Marson, G., Pellegrini, M., Segal, D.J., et al. (2014). SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature 507, 124128.

72. Johnson, N.R., Yeoh, J.M., Coruh, C., and Axtell, M.J. (2016). Improved Placement of Multi-mapping Small RNAs. G3 (Bethesda) 6, 2103-2111.

73. Joly-Lopez, Z., and Bureau, T.E. (2014). Diversity and evolution of transposable elements in *Arabidopsis*. Chromosome Research 22, 203-216.

74. 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 biology 6, e194-e194.

75. Kapitonov, V.V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. TRENDS in Genetics 23, 521-529.

76. Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-Wide Profiling and Analysis of *Arabidopsis* siRNAs. PLOS Biology 5, e57.

77. Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. Curr Biol 13, 421-426.

78. Kawakatsu, T., Nery, J.R., Castanon, R., and Ecker, J.R. (2017). Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biology 18, 171.

79. Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R.J., Nery, J.R., Urich, M.A., Han, X., Lister, R., Benfey, P.N., et al. (2016). Unique cell-type-specific patterns of DNA methylation in the root meristem. Nature Plants 2, 16058.

80. Kawashima, T., and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. Nature Reviews Genetics 15, 613.

81. Kim, M.Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. Trends in plant science 19, 320-326.

82. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and
Janulaitis, A. (1989). Sequence motifs characteristic of DNA[cytosineN4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic Acids Res 17, 9823-9832.

83. Krieg, A.M. (2006). Therapeutic potential of Toll-like receptor 9 activation. Nat Rev Drug Discov 5, 471-484.

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

85. Kubo, N., Toh, H., Shirane, K., Shirakawa, T., Kobayashi, H., Sato, T., Sone, H., Sato, Y., Tomizawa, S.-i., and Tsurusaki, Y. (2015). DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. Bmc Genomics 16, 624.

86. Laberge, R.M., and Boissonneault, G. (2005). On the nature and origin of DNA strand breaks in elongating spermatids. Biol Reprod 73, 289-296.

87. Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics Chapter 11, Unit 11.17.

88. Larkins, B.A., Dilkes, B.P., Dante, R.A., Coelho, C.M., Woo, Y.m., and Liu, Y. (2001). Investigating the hows and whys of DNA endoreduplication. Journal of Experimental Botany 52, 183-192.

89. Law, J.A., Ausin, I., Johnson, L.M., Vashisht, A.A., Zhu, J.K., Wohlschlegel, J.A., and Jacobsen, S.E. (2010). A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in *Arabidopsis*. Curr Biol 20, 951-956.

90. Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M., Strahl, B.D., Patel, D.J., and Jacobsen, S.E. (2013). Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498, 385-389.

91. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11, 204.

92. Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet 32, 42-56.

93. Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13, 1192-1200.

94. Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. Trends in Genetics 31, 274-280.

95. Liu, Z.W., Shao, C.R., Zhang, C.J., Zhou, J.X., Zhang, S.W., Li, L., Chen, S., Huang, H.W., Cai, T., and He, X.J. (2014). The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet 10, e1003948.

96. Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429, 318-322.

97. Mak, J., and Kleiman, L. (1997). Primer tRNAs for reverse transcription. J Virol 71, 8087-8095.

98. Malik, G., Dangwal, M., Kapoor, S., and Kapoor, M. (2012). Role of DNA methylation in growth and differentiation in Physcomitrella patens and characterization of cytosine DNA methyltransferases. FEBS Journal 279, 4081-4094.

99. Malone, T., Blumenthal, R.M., and Cheng, X. (1995). Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol 253, 618-632.

100. Martinez, G. (2017). tRNA-derived small RNAs target transposable element transcripts. 45, 5142-5152.

101. Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L.L., DeFraia, C., Slotkin, R.K., and Köhler, C. (2018). Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. Nature Genetics 50, 193-198.

102. Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15, 394-408.

103. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D.
(2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic acids research 45, D183-D189.

104. Miller, D., Brinkworth, M., and Iles, D. (2010). Paternal DNA packaging in spermatozoa: more than the sum of its parts? DNA, histones, protamines and epigenetics. Reproduction 139, 287-301.

105. Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W.R., Hannon, G.J., and Smith, A.D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell 146, 1029-1041.

106. Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. (2008). PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proceedings of the National Academy of Sciences of the United States of America 105, 3145-3150.

107. Noy-Malka, C., Yaari, R., Itzhaki, R., Mosquna, A., Auerbach Gershovitz, N., Katz, A., and Ohad, N. (2014). A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of Physcomitrella patens. Plant Mol Biol 84, 719-735.

108. Nuthikattu, S., McCue, A.D., Panda, K., Fultz, D., DeFraia, C., Thomas, E.N., and Slotkin, R.K. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. Plant Physiol 162, 116-131.

109. O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne) 9, 402-402.

110. Oliveira, P.H., Touchon, M., and Rocha, E.P. (2014). The interplay of restrictionmodification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res 42, 10618-10631.

111. Oliver, C., Santos, J.L., and Pradillo, M. (2016). Accurate Chromosome Segregation at First Meiotic Division Requires AGO4, a Protein Involved in RNA-Dependent DNA Methylation in *Arabidopsis thaliana*. Genetics 204, 543-553.

112. Osman, K., Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., and Franklin, F.C. (2011). Pathways to meiotic recombination in *Arabidopsis thaliana*. New Phytol 190, 523-544.

113. Parish, R.W., Phan, H.A., Iacuone, S., and Li, S.F. (2012). Tapetal development and abiotic stress: a centre of vulnerability. Functional Plant Biology 39, 553-559.

114. Park, K., Kim, M.Y., Vickers, M., Park, J.-S., Hyun, Y., Okamoto, T., Zilberman, D., Fischer, R.L., Feng, X., Choi, Y., et al. (2016). DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. Proceedings of the National Academy of Sciences 113, 15138-15143.

115. Patel, P., Mathioni, S., Kakrana, A., Shatkay, H., and Meyers, B.C. (2018). Reproductive phasiRNAs in grasses are compositionally distinct from other classes of small RNAs. New Phytol 220, 851-864.

116. Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nature protocols 9, 171.

117. Pikaard, C.S., and Mittelsten Scheid, O. (2014). Epigenetic Regulation in Plants. Cold Spring Harbor Perspectives in Biology 6, a019315.

118. Quilichini, T.D., Grienenberger, E., and Douglas, C.J. (2015). The biosynthesis, composition and assembly of the outer pollen wall: A tough case to crack. Phytochemistry 113, 170-182.

119. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

120. Regulski, M., Lu, Z., Kendall, J., Donoghue, M.T.A., Reinders, J., Llaca, V., Deschamps, S., Smith, A., Levy, D., McCombie, W.R., et al. (2013). The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Research 23, 1651-1662.

121. Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64-69.

122. Rodrigues, J.A., and Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. Genes & Development 29, 2517-2531.

123. Schmid, M.W., Giraldo-Fonseca, A., Rövekamp, M., Smetanin, D., Bowman, J.L., and Grossniklaus, U. (2018). Extensive epigenetic reprogramming during the life cycle of *Marchantia polymorpha*. Genome Biology 19, 9.

124. Schmitz, R.J., Lewis, Z.A., and Goll, M.G. (2019). DNA Methylation: Shared and Divergent Features across Eukaryotes. Trends in Genetics.

125. Schorn, A.J., Gutbrod, M.J., LeBlanc, C., and Martienssen, R. (2017). LTR-Retrotransposon Control by tRNA-Derived Small RNAs. Cell 170, 61-71.e11.

126. Secco, D., Wang, C., Shou, H., Schultz, M.D., Chiarenza, S., Nussaume, L., Ecker, J.R., Whelan, J., and Lister, R. (2015). Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. eLife 4, e09343.

127. Shimamura, M. (2015). *Marchantia polymorpha* : Taxonomy, Phylogeny and Morphology of a Model System. Plant and Cell Physiology 57, 230-256.

128. Shu, Z., Row, S., and Deng, W.M. (2018). Endoreplication: The Good, the Bad, and the Ugly. Trends Cell Biol 28, 465-474.

129. Slieker, R.C., Roost, M.S., van Iperen, L., Suchiman, H.E.D., Tobi, E.W., Carlotti, F., de Koning, E.J., Slagboom, P.E., Heijmans, B.T., and de Sousa Lopes, S.M.C. (2015). DNA methylation landscapes of human fetal development. PLoS genetics 11, e1005583.

130. Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11, 817820.

131. Steer, M.W. (1977). Differentiation of the tapetum in Avena. I. The cell surface. J Cell Sci 25, 125-138.

132. Stewart, K.R., Veselovska, L., and Kelsey, G. (2016). Establishment and functions of DNA methylation in the germline. Epigenomics 8, 1399-1413.

133. Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nat Struct Mol Biol 21, 64-72.

134. Stroud, H., Greenberg, Maxim V.C., Feng, S., Bernatavichute, Yana V., and Jacobsen, Steven E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the *Arabidopsis* Methylome. Cell 152, 352-364.

135. Sunil K. Kenchanmane Raju, Eleanore Jeanne Ritter, Chad E. Niederhuth; Establishment, maintenance, and biological roles of non-CG methylation in plants. Essays Biochem 20 December 2019; 63 (6): 743–755.

136. Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet 9, 465-476.

137. Takuno, S., and Gaut, B.S. (2011). Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. Molecular Biology and Evolution 29, 219-227.

138. Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. Nature Plants 2, 15222.

139. Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11, 264-275.

140. Talbert, P.B., Meers, M.P., and Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nature Reviews Genetics 20, 283-297.

141. Tetsuya Hisanaga, S.F., Keiji Nakajima (2018). Maternally expressed MpKNOX1 triggers embryo development in *Marchantia polymorpha* (The 25th International Congress on Sexual Plant Reproduction: Nara Institute of Science and Technology).

142. Tran, R.K., Henikoff, J.G., Zilberman, D., Ditt, R.F., Jacobsen, S.E., and Henikoff, S. (2005). DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. Current Biology 15, 154-159.

143. Underwood, C.J., Choi, K., Lambing, C., Zhao, X., Serra, H., Borges, F., Simorowski, J., Ernst, E., Jacob, Y., Henderson, I.R., et al. (2018). Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome research 28, 519-531.

144. Vanrobays, E., Thomas, M., and Tatout, C. (2018). Heterochromatin positioning and nuclear architecture. Annual Plant Reviews online, 157-190.

145. Villarreal A, J.C., Crandall-Stotler, B.J., Hart, M.L., Long, D.G., and Forrest, L.L. (2016). Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytologist 209, 1734-1746.

146. Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J.D., and Feng, X.
(2018). Sexual-lineage-specific DNA methylation regulates meiosis in *Arabidopsis*. Nature Genetics 50, 130-137.

147. Wallace, S., Fleming, A., Wellman, C.H., and Beerling, D.J. (2011). Evolutionary development of the plant and spore wall. AoB Plants 2011, plr027-plr027.

148. Wang, X., Hu, L., Wang, X., Li, N., Xu, C., Gong, L., and Liu, B. (2016). DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice. Molecular Plant 9, 305-307.

149. Weick, E.-M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. Development 141, 3458.

150. Weiss-Schneeweiss, H., and Maluszynska, J. (2001). Molecular Cytogenetic Analysis of Polyploidization in the Anther Tapetum of Diploid and Autotetraploid *Arabidopsis thaliana* Plants. Annals of Botany 87, 729-735.

151. Williams, B.P., Pignatta, D., Henikoff, S., and Gehring, M. (2015). Methylationsensitive expression of a DNA demethylase gene serves as an epigenetic rheostat. PLoS Genet 11, e1005142.

152. Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu Rev Genet 25, 585-627.

153. Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. PLoS Genet 4, e1000156.

154. Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell 156, 45-68.

155. Wu, L., Mao, L., and Qi, Y. (2012). Roles of DICER-LIKE and ARGONAUTE Proteins in TAS-Derived Small Interfering RNA-Triggered DNA Methylation. Plant Physiology 160, 990-999.

156. Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nature Reviews Genetics 18, 517.

157. Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., von Mering, C., and Grossniklaus, U. (2010). *Arabidopsis* Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes. Current Biology 20, 506-512.

158. Xia, R., Chen, C., Pokhrel, S., Ma, W., Huang, K., Patel, P., Wang, F., Xu, J., Liu, Z., Li, J., et al. (2019). 24-nt reproductive phasiRNAs are broadly present in angiosperms.
Nature Communications 10, 627.

159. Xue, M., Wang, J., Jiang, L., Wang, M., Wolfe, S., Pawlowski, W.P., Wang, Y., and He, Y. (2018). The Number of Meiotic Double-Strand Breaks Influences Crossover Distribution in *Arabidopsis*. The Plant cell 30, 2628-2638.

160. Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A., and Ohad, N. (2019). RdDMindependent *de novo* and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. Nat Commun 10, 1613.

161. Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. The Plant Journal 65, 503-516.

162. Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., and Henderson, I.R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in *Arabidopsis*. Genes & development 29, 2183-2202.

163. Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K.,
Harmer, Stacey L., and Zilberman, D. (2013). The *Arabidopsis* Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell 153, 193-205.

164. Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr Biol 20, R780-785.

165. Zeng, Y., and Chen, T. (2019). DNA Methylation Reprogramming during Mammalian Development. Genes (Basel) 10.

166. Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V., and Meyers, B.C. (2015). Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. Proceedings of the National Academy of Sciences of the United States of America 112, 3146-3151.

167. Zhang, X., Henderson, I.R., Lu, C., Green, P.J., and Jacobsen, S.E. (2007). Role of RNA polymerase IV in plant small RNA metabolism. Proceedings of the National Academy of Sciences of the United States of America 104, 4536-4541.

168. Zheng, S., Li, J., Ma, L., Wang, H., Zhou, H., Ni, E., Jiang, D., Liu, Z., and Zhuang, C. (2019). OsAGO2 controls ROS production and the initiation of tapetal PCD by epigenetically regulating OsHXK1 expression in rice anthers. Proceedings of the National Academy of Sciences 116, 7549.

169. Zhou, J., Liu, L., Li, Q., Xu, W., Li, K., Wang, Z.-W., and Sun, Q. (2019). Intronic heterochromatin prevents cryptic transcription initiation in *Arabidopsis*. bioRxiv, 610832.

170. Zhou, M., Palanca, A.M.S., and Law, J.A. (2018). Locus-specific control of the *de novo* DNA methylation pathway in *Arabidopsis* by the CLASSY family. Nature Genetics, 50, 865–873.

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

172. Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., et al. (2011). Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet 7, e1002389.