Genetic and epigenetic control of germline specification in Arabidopsis pollen

Filipe de Sousa Borges

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

Genetic and Epigenetic Control of Germline Specification in Arabidopsis Pollen

Filipe de Sousa Borges

DOUTORAMENTO EM BIOLOGIA
(Biologia Celular)

Tese co-orientada pelo Doutor Jörg D. Becker
e pelo Professor Doutor José A. Feijó

2012

ACKNOWLEDGMENTS

Esta tese é dedicada à família, Mariana e malta da Boca, os grandes responsáveis pelo que sou hoje. Um agradecimento especial aos meus Pais por terem apoiado incondicionalmente a minha formação.

Jörg, thanks for unfurling my research skills and critical thinking when others failed, for your guidance and friendship. Your continuous help and support throughout the years reflect all our scientific achievements. I never imagined that applying to that fellowship could become one of the best decisions in my life.

Prof. Feijó, thanks for your honesty and inspiring words, which I will always remember as a daily stimulus to keep on doing more and better.

Thanks to the IGC direction and staff for the excellent research conditions, and to all my colleagues that made the IGC such a fun place to work. Special thanks to Rui Gardner and Telma Lopes for their dedication and help with FACS, to all former and current members of the Plant Dev and Plant Genomics labs, PMB and PSS labs for the daily discussions, materials, methods and reagents.

Thanks to Rob Martienssen, Joe Calarco and all the researchers in the Martienssen lab for the technical support and fruitful discussions, especially to Keith Slotkin for mentoring my first months in the lab.

Thanks to Ignacio Rubio for the help with the mimic constructs.

Thanks to Fundação para a Ciência e a Tecnologia for my PhD fellowship (SFRH/BD/48761/2008), and to Fundação Calouste Gulbenkian for additional funding support.

RESUMO

Em plantas superiores, os gâmetas masculinos encontram-se no interior da célula vegetativa do pólen, e são transportados pelo tubo polínico até ao saco embrionário onde ocorre a fertilização dupla. O estudo dos mecanismos de diferenciação e maturação dos dois tipos de células que formam o grão de pólen é essencial para compreender a reprodução em plantas, reprogramação epigenética e *imprinting* genómico, mas até à data, este tipo de abordagens na planta modelo *Arabidopsis thaliana* não eram possíveis pela falta de métodos disponíveis para o isolamento das células espermáticas. Nesta tese descrevemos pela primeira vez um método que utiliza a citometria de fluxo para separar as células espermáticas e o núcleo vegetativo do pólen de Arabidopsis, o que permitiu a análise do seu transcritoma por *microarrays*. A análise comparativa com o transcritoma de tecidos esporofíticos permitiu concluir que a linha generativa tem um perfil de transcrição distinto e diverso, funcionalmente direcionado para a reparação do DNA, proteólise por ubiquitinação e progressão no ciclo celular. Além disso, a análise de genes associados a vias de metilação do DNA e processamento de RNAs pequenos sugere que a regulação do estado epigenético do genoma paterno pode depender de novos mecanismos até agora desconhecidos.

Os transposões são silenciados por mecanismos epigenéticos através de RNAs pequenos de interferência (*small interfering RNAs*, siRNAs), o que é especialmente importante para os gâmetas, pois contribuem com material genético para a geração seguinte. No pólen de Arabidopsis a expressão de transposões é reativada, mas apenas no núcleo vegetativo que não contribui com o seu DNA para o zigoto. A expressão de transposões na célula vegetativa coincide com ausência de siRNAs e outros elementos essenciais para a formação da

heterocromatina. No entanto, siRNAs de 21nt derivados de retrotransposões *Athila* são processados e acumulam-se no pólen e nas células espermáticas, sugerindo assim que siRNAs derivados da reativação de transposões no núcleo vegetativo podem vir a contribuir posteriormente para o seu silenciamento nos gâmetas. Para compreender os mecanismos de reprogramação epigenética durante o desenvolvimento da linha generativa no pólen de Arabidopsis, analisamos o perfil de metilação do DNA para os três tipos de núcleos que representam o desenvolvimento do gametófito masculino, desde a célula precursora (micrósporo) até ao grão de pólen maduro. Com estes resultados foi possível verificar que a célula vegetativa perde metilação CG e CHG em comparação com as células espermáticas e micrósporos, mas apenas em zonas específicas, nomeadamente transposões adjacentes a genes *imprinted*. Isto sugere que a metilação em CG e CHG é corretamente propagada após a meiose e durante a replicação de DNA que dá origem à linha generativa, mas posteriormente apagada apenas no núcleo vegetativo. No entanto, os níveis de metilação assimétrica CHH, que está relacionada com a metilação do DNA direcionada por RNAs (*RNA-directed DNA Methylation*, RdDM), são mais baixos nos núcleos das células espermáticas do que no vegetativo, o que reflete a acumulação de siRNAs correspondentes no pólen. Em transposões do tipo RC/Helitron e DNA/MuDR, existe um grupo específico que parece ser regulado por siRNAs e metilação em CHH, mantendo assim os níveis de metilação CG e CHG, e outro grupo que não apresenta siRNAs e metilação CHH, perdendo também a metilação CG e CHG. Estes resultados sugerem que a perda de metilação em zonas que não são alvos de RdDM no núcleo vegetativo, pode significar um mecanismo envolvido na reprogramação de *imprinting* genómico.

Tanto em plantas como animais, os vários estudos sobre a importância de microRNAs (miRNAs) no desenvolvimento da linha generativa são contraditórios. A maioria das

proteínas envolvidas nas vias de produção e atividade de miRNAs é conhecida, no entanto, este conhecimento é ainda escasso no que diz respeito à possível existência de mecanismos de regulação pós-transcricional distintos entre diferentes tipos de células. Por esse motivo, procedemos à análise por sequenciação de miRNAs em células espermáticas e no pólen, utilizando diferentes ferramentas moleculares que permitiram a redução da sua atividade durante o desenvolvimento da linha generativa. Conseguimos ainda identificar 25 miRNAs potencialmente novos, assim como variações no tamanho de miRNAs anteriormente conhecidos, o que pode indicar a sua sub-funcionalização por associação com um complexo de silenciamento de genes específico da linha generativa. A proteína Argonaute 5 (AGO5) pode fazer parte desse complexo, uma vez que tem homologia com AGO1 e está localizada no citoplasma das células espermáticas. No entanto, a expressão de transcritos-alvo de miRNAs não se encontra desregulada no pólen do mutante *ago5-4*, indicando assim que AGO5 não está diretamente envolvido no silenciamento pós-transcricional de genes durante o desenvolvimento do pólen. Ainda assim, os nossos resultados sugerem um mecanismo anteriormente descrito para AGO10, que ao sequestrar certos miRNAs previne a sua incorporação em outras proteínas Argonaute envolvidas no silenciamento de genes ao nível pós-transcricional. Para além disso, a atividade de AGO5 durante o desenvolvimento do polén poderá estar ainda relacionada com o silenciamento de genes ao nível da transcrição, uma vez que muitos dos genes sobre-expressos no mutante são flanqueados por transposões normalmente silenciados por RdDM.

Palavras-Chave: Arabidopsis, pólen, linha generativa, transcritómica, metilação do DNA, microRNAs

ABSTRACT

In flowering plants the male gametes are embedded within the cytoplasm of the growing pollen tube, which transports them to the embryo sac where double fertilization occurs. Understanding the mechanisms and conditions by which sperm cells mature and take part in fertilization are crucial goals in the study of plant reproduction, epigenetic reprogramming and genomic imprinting. Analyses in the model plant *Arabidopsis thaliana* were hindered because no method to isolate sperm cells was available. We used Fluorescence-activated cell sorting (FACS) to isolate sperm cells and vegetative nucleus from Arabidopsis pollen, allowing gene expression analysis at a genome-wide level by microarrays. Comparative analysis of the sperm cell transcriptome with those of representative sporophytic tissues and of entire pollen grains showed that sperm cells have a distinct and diverse transcriptional profile, functionally skewed towards DNA repair, ubiquitin-mediated proteolysis and cell cycle progression. Moreover, analysis of the small RNA and DNA methylation pathways suggested that distinct mechanisms might be involved in regulating the epigenetic state of the paternal genome.

Transposable elements (TEs) are suppressed by epigenetic silencing and small interfering RNAs (siRNAs), especially in the gametes that contribute with genetic material to the next generation. In wild-type Arabidopsis pollen, TEs are reactivated and transpose, but only in the vegetative nucleus, which does not provide DNA to the fertilized zygote. TE expression coincides with down-regulation of key elements involved in heterochromatin formation, and many TE-derived siRNAs. However, 21nt siRNAs from *Athila* retro-transposons are generated and accumulate in pollen and sperm, suggesting that siRNA from TEs activated in the vegetative nucleus may target silencing in gametes. We profiled DNA

methylation of the three different types of nuclei observed through pollen maturation, from the post meiotic precursor cell to the mature pollen grain in an attempt to understand epigenetic reprogramming during plant germline development. CG and CHG methylation is reduced in the vegetative cell comparing with that of sperm and microspore, but only at specific loci, which tend to mark TEs surrounding some noted imprinted genes. This indicates that symmetric methylation in the microspore is maintained during DNA replication in the first pollen division, and actively lost at imprinted loci. Intriguingly, CHH methylation, which is tied to RNA-directed DNA methylation (RdDM), is much lower in sperm nuclei than in the vegetative nucleus, overlapping with the accumulation of correspondent siRNAs in pollen. There is a particular subset of RC/Helitron and DNA/MuDR elements that are rich in siRNAs and CHH methylation while maintaining CG and CHG methylation levels, and a second subset of elements that are devoid of siRNAs and CHH methylation, and actively lose CG and CHG. These results suggest that in the VN, DNA demethylation occurs only at TE loci that are not actively targeted by RdDM, in a mechanism that might be involved in re-establishing imprinting marks.

From plants to animals, there is contradictory evidence on the importance of microRNAs (miRNAs) during germ cell development. While most of the core proteins involved in the miRNA pathway in plants have been identified in the Arabidopsis sporophyte, there is very limited understanding about potentially distinct mechanisms of post-transcriptional regulation between different cell lineages. We provide a robust comparative analysis of miRNAs identified in sperm cells and pollen by deep sequencing, using molecular tools to deplete their function along germ cell specification and beyond fertilization. In addition, we present the identification of 25 potentially novel miRNAs processed in the male gametophyte and sperm cells, as well as enriched variations in the sequence length of known

miRNAs that might indicate sub-functionalization by association with a putative germline-specific Argonaute complex. ARGONAUTE 5 (AGO5), by close homology to AGO1 and localizing preferentially to the sperm cell cytoplasm in mature pollen, may be part of such a complex. MicroRNA targets are not misregulated in *ago5-4* pollen, indicating that AGO5 is not part of the canonical miRNA pathway during pollen development. In alternative and similarly to what had been previously reported for AGO10, our data suggests that AGO5 could be involved in miRNA sequestering, thus preventing post-transcriptional gene silencing in the sperm cells. Moreover, AGO5 may be additionally involved in transcriptional gene silencing by RdDM, as some genes up-regulated in the mutant are flanked by TEs and repeat sequences targeted by RdDM.

Keywords: Arabidopsis, pollen, germline, transcriptomics, DNA methylation, microRNAs

ABBREVIATIONS

AGO	ARGONAUTE
BCP	Bicellular pollen
bp	base pair
cDNA	Complementary deoxyribonucleic acid
CEL	cell intensity file
ChIP	Chromatin immunoprecipitation
CMT3	CHROMOMETHYLASE 3
Col	Columbia
cRNA	Complementary ribonucleic acid
DAPI	4',6-diamidino-2-phenylindole
DCL	DICER-LIKE
DDM1	DECREASE IN DNA METHYLATION 1
DIC	Differential Interference Contrast
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside-triphosphate
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
dsRNA	double-stranded RNA
eGFP	enhanced Green Fluorescent Protein
FACS	Fluorescence-Activated Cell Sorting
FDR	False Discovery Rate
FSC	Forward scatter
GC	Generative Cell
GFP	Green Fluorescent Protein
IVT	in vitro transcription
kb	kilo base
Ler	Landsberg
MET1	DNA METHYLTRANSFERASE 1
MGU	Male Germ Unit
mM	milli Molar
mRNA	messenger RNA

miRNA	microRNA
MS	Murashige-Skoog
MP	Mature pollen
nt	nucleotide
PCR	Polymerase chain reaction
PMI	Pollen mitosis I
PMII	Pollen mitosis II
RdDM	RNA-directed DNA methylation
RDR	RNA-dependent RNA polymerase
RFP	Red Fluorescent Protein
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RT-PCR	Reverse transcription - PCR
SC	Sperm cell
SSC	side scatter
siRNA	small interfering RNA
ssRNA	single-stranded RNA
T-DNA	Transfer DNA
TAIR	The Arabidopsis Information Resource
TCP	Tricellular pollen
TE	Transposable element
UNM	Uninucleate microspore
UTR	Untranslated region
VC	Vegetative cell
VN	Vegetative nucleus

TABLE OF CONTENTS

CHAPTER 1

General Introduction

1. INTRODUCTION

Double fertilization: the hallmark of angiosperm biology

Flowering plants evolved a unique sexual reproduction system that ends in a double fertilization, involving two sperm cells and two female gametes (egg cell and central cell). One sperm cell fertilizes the egg cell giving rise to the diploid embryo, while the other sperm cell fuses with the two polar nuclei of the central cell to originate a triploid endosperm. The endosperm functions as a supportive tissue to protect and nourish embryo development inside the seed. Since the discovery of double fertilization in the nineteenth century (GUIGNARD, 1899; NAWASCHIN, 1898), the cellular and molecular mechanisms involved in this process have remained by and large unknown, as it occurs deep in the maternal tissues thus compromising standard cytological analysis. A significant advance came from studies in Torenia, which has a protruding female gametophyte without covering integuments (HIGASHIYAMA *et al*., 1997, 2000). This work demonstrated that specialized cells in the female gametophyte, called synergids, play a major role in short-range attraction of the pollen tube (HIGASHIYAMA *et al*., 2001), but the mechanism controlling the double fusion event remained enigmatic. Recent studies triggered a renewed interest in double fertilization, with the discovery of a number of cell and nuclear marker proteins expressed in the gametes and companion cells in the gametophytes (BERGER *et al*., 2008). These findings allow live imaging of fertilization and karyogamy, and will certainly contribute to a faster understanding of its features.

Microgametogenesis in Arabidopsis

The male gametophyte of higher plants (pollen grain) is haploid, derived from two post-meiotic cellular divisions. Meiosis occurs in the anthers yielding tetrads that release uninucleate microspores. Subsequently, pollen mitosis I (PM I) gives rise to a larger vegetative cell and a smaller generative cell (GC), that undergoes pollen mitosis II (PM II) to produce two sperm cells (SC) (Figure 1.1A). In Arabidopsis, PM II occurs before anthesis, so that tricellular pollen grains are later released from the anthers (BOAVIDA *et al.*, 2005a; DUMAS *et al.*, 1985; YU *et al.*, 1989). When mature pollen lands on the papillae of a receptive stigma, it triggers complex series of cell-cell signalling events that will orchestrate pollen tube growth towards the female gametophyte (embryo sac) (BOAVIDA *et al.*, 2011; FEIJÓ 2010). Upon arrival the pollen tube tip bursts, discharging the two sperm cells inside the embryo sac wherein double fertilization occurs (BOAVIDA *et al.*, 2005b, HAMAMURA *et al.*, 2011), but the mechanisms underlying double fusion remain unknown.

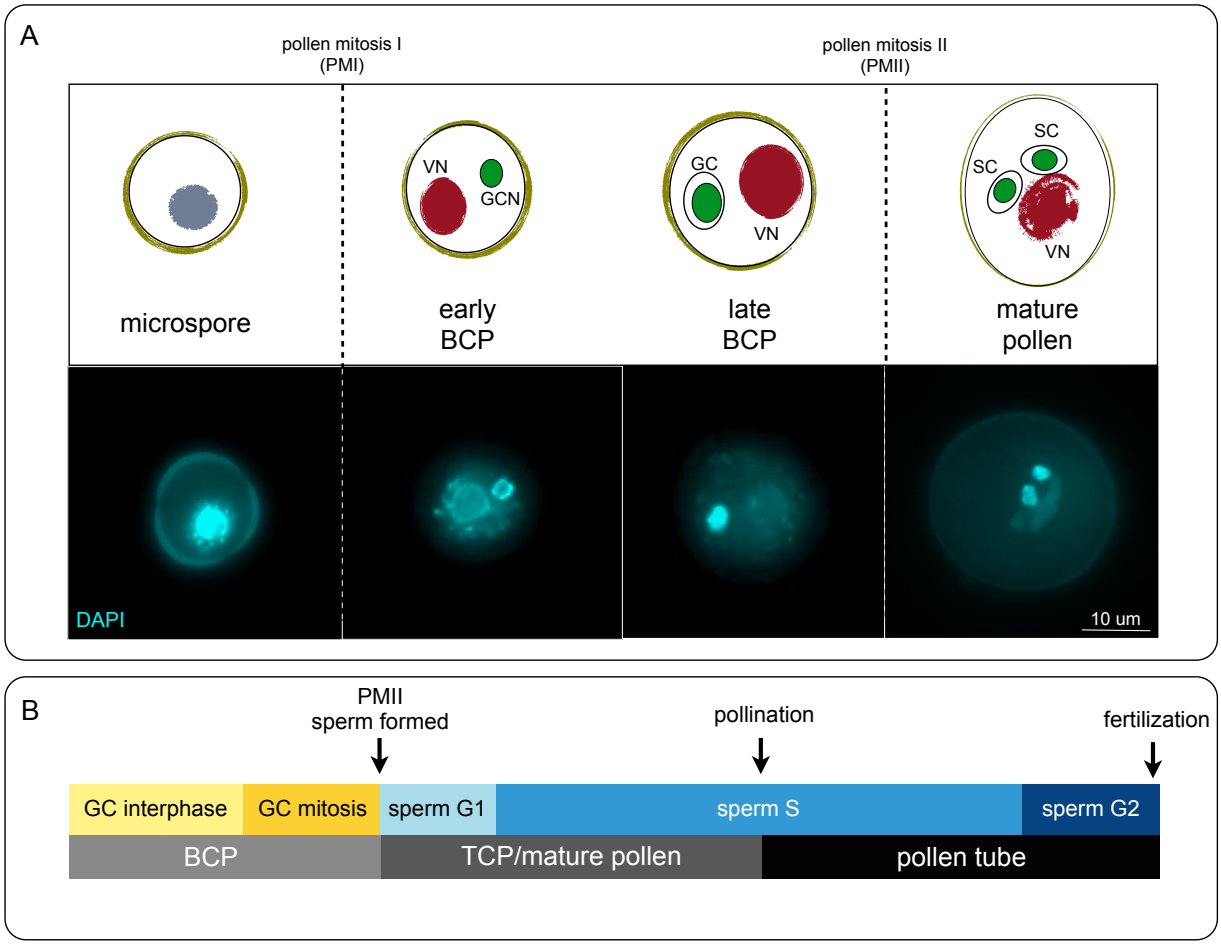


Figure 1.1 - Male gametophyte development. Schematic diagram representing microgametogenesis in Arabidopsis pollen with correspondent DAPI-stained cells **(A),** and cell cycle transitions in the germline **(B)**. The microspore undergoes a first asymmetric cell division originating a larger vegetative cell and smaller generative cell (GC). The generative cell divides further to generate two identical sperm cells (SC). At pollination the sperm cells are stalled in S phase, which is prolonged during pollen tube growth, reaching G2 just prior to fertilization (Adapted from FRIEDMAN, 1999). VN – v§egetative cell nucleus, BCP - bicellular pollen, TCP - tricellular pollen.

Polarity and first asymmetric division

The first asymmetric division that leads to germline formation is under the control of multiple genetic pathways. In Arabidopsis, the microspore nucleus migrates to the future germ cell pole, generating a polarized microspore (TWELL *et al.*, 1998), possibly under a

microtubule-based fashion as suggested by the identification of TUBG1 and TUBG2, two functionally redundant c-tubulin genes required for spindle and phragmoplast organization in Arabidopsis microspores (PASTUGLIA *et al.*, 2006). More recently, microtubule arrays were observed in polarized tobacco microspores, including a cortical microtubule array associated with the nucleus at the germ cell pole (OH *et al.*, 2010), originating a nuclear cap of microtubules connecting the nuclear envelope and cell cortex. This work described TMBP200, the tobacco ortholog of Arabidopsis MOR1/GEM1, for which a mutation showed microspore division defects (PARK *et al.*, 1998). The authors suggested that this structure could be important in generating asymmetry as well as maintaining nuclear position prior to division. Other mutations have been described showing proper division but failure in cytokinesis. This is the case of the *two-in-one* (*tio*) mutant microspores, resulting in binucleate pollen. *TIO* encodes a Ser/Thr protein kinase and has an essential role in cell plate expansion (OH *et al.*, 2005).

It is clear that microtubule-based mechanisms are associated with maintaining the cell polarity required for the first asymmetric pollen division, but is not clear yet how is it established intracellularly, as the determinants of polarity are still unknown. As reported in the Arabidopsis female gametophyte, these factors might consist of regulatory RNAs and proteins possibly controlled by auxin gradients (PAGNUSSAT *et al.*, 2009), though this hypothesis is still controversial.

Genetic determinants of germline specification

Germ cell differentiation occurs in sexually reproducing organisms, but along their evolution, several species developed distinct spatial and temporal control over germline proliferation. In plants, germ cells are specified from somatic cells at the end of ontogeny by

completion of flower organ development. By contrast in animals, germ cells are set aside of the somatic cell lineage early during embryonic development. Their germline is highly protected and strictly regulated during migration through the somatic tissues. In the gonads, germ cells acquire sex-specific differentiation and the ability to undergo meiosis to produce gametes. Three major features of germline proliferation are known to be conserved among several animal species; the transcriptional repression of somatic differentiation, cell-to-cell signalling between germ cells and their neighbouring somatic cell lineage, and post-transcriptional regulation of germ cell-specific functions while keeping somatic genes suppressed to prevent transdifferentiation (CINALLI *et al.*, 2008). For example in the worm and fly germline, a repression of RNA Pol II transcriptional elongation (GHOSH and SEYDOUX, 2008; HANYU-NAKAMURA *et al.*, 2008; MARTINHO *et al.*, 2004) was shown, and expectedly, mutations in core genes within these pathways lead to expression of genes characteristic for somatic cells (MARTINHO *et al.*, 2004). Like in animals, the male germline in plants is also essentially regulated by transcriptional control. Haerizadeh and colleagues have shown that a germline-restrictive silencing factor (GRSF) is ubiquitously expressed in non-germ cells, and its down-regulation results in expression of male germ cell-specific gene products (HAERIZADEH *et al.*, 2006). However, plant germ cell mechanisms to repress the transcriptional program of their somatic precursors remain unknown.

One of the first mutants characterized by showing impaired germ cell specification, was a null mutation in Arabidopsis *DUO1*, a MYB transcription factor (ROTMAN *et al.*, 2005). Pollen in *duo1* plants fail to go through pollen mitosis II, resulting in a single non-functional generative cell. The fact that this single sperm cell was not able to be specified into a gamete and fertilize, as other mutants showing similar phenotypes, suggested that *DUO1* has a prominent role in this process that is beyond cell cycle progression. Recent studies have

shown that DUO1 is indeed a key transcription factor regulating gene expression during germ cell specification, and being itself germline-specific, it activates expression of several genes preferentially expressed in sperm (BORG *et al.*, 2011; BROWNFIELD *et al.*, 2009a). However, what activates *DUO1* expression after PMI is not known. The fact that the *DUO1* promoter contains a putative binding site for GRSF suggested a mechanism similar to what was described in lily (HAERIZADEH *et al.*, 2006), but this possibility was ruled out by showing that a shorter *DUO1* promoter without the GRSF binding site is still specifically active only in the germline (BROWNFIELD *et al.*, 2009a).

DUO3, a novel nuclear protein, is also required for sperm cell specification, G2/M transition during PM II and expression of a subset of DUO1 activated genes by an unknown mechanism. However, unlike DUO1, the requirement for DUO3 in the germline during G2/M transition is independent of cell cycle regulators such as CYCB1;1 (BROWNFIELD *et al.*, 2009b). Interestingly, DUO3 seems to be a widely conserved protein, as it is related to the *Caenorhabditis elegans* Gonadless-4 (GON-4), a cell lineage regulator of gonadogenesis (FRIEDMAN *et al.*, 2000). In addition, a DUO3 ortholog isolated from *Physcomitrella patens* could restore Arabidopsis *duo3* phenotypes (BROWNFIELD *et al.*, 2009b).

Cell cycle

The precise control over cell cycle transitions is important during Arabidopsis gametogenesis. Unlike the gametes of other eukaryotes that fuse in G1, Arabidopsis replicate DNA before fertilization and fuse in G2 (Figure 1.1B). As such, the differentiated sperm cells inside the pollen grain are actively replicating their DNA within the pollen tube, reaching G2 just before double fertilization occurs (FRIEDMAN, 1999). This suggests that cell cycle

transitions must be well synchronized between gametophytes before fertilization and karyogamy. A mutation in the Arabidopsis *CDC2A/CDKA;1* gene was initially reported as having a paternal effect, whereby mutant pollen produces only one sperm cell that exclusively fertilizes the egg cell (NOWACK *et al.*, 2006). Although this mutant arrested embryo development early, its single fertilization event somehow triggered autonomous proliferation of the endosperm. More recently, it was shown that *cdka;1* pollen is delayed in cell cycle progression and a significant proportion of *cdka;1* pollen delivers two sperm cells and not only one as initially thought. This work clarifies that sperm entry is sufficient to initiate central cell division, but the paternal genome is required to complete endosperm development (AW *et al.*, 2010). In contrast, a mutation in the retinoblastoma-related protein (RBR) causes hyperproliferation of both vegetative and generative cells, which is prevented in *cdka;1* mutant pollen (CHEN *et al.*, 2009). As in animals, retinoblastoma in plants should also be involved in repressing E2F transcription factors, thus preventing G1/S transitions (Figure 1.2). Interestingly, in *rbr* pollen the resulting generative cells seem to be properly specified into gametes, while the vegetative loses cell fate specification (CHEN *et al.*, 2009). The loss of activity of the Chromatin Assembly Factor 1 (CAF1) also prevents PM II, but in this case the resulting single sperm cell could fertilize either female gamete (CHEN *et al.*, 2008). F-Box proteins were also shown to play a key role controlling progression through pollen mitotis, as the Arabidopsis FBL17 was identified as being part of the E3 SCF-type Ubiquitin ligase complex in pollen. Loss-of-function mutants of *FBL17* fail to undergo PM II, resulting in single sperm cell that fertilizes exclusively the egg cell (GUSTI *et al.*, 2009). FBL17 seems to act mainly in the generative cell, depleting the CDK inhibitors KRP6 and KRP7 (KIM *et al.*, 2008).

These observations suggested that before fertilization the two identical sperm cells

would be already committed to fuse with one or the other female gamete. However, recent semi *in vivo* experiments have shown that Arabidopsis sperm cells exhibit no preference for either female gamete (HAMAMURA *et al.*, 2011).

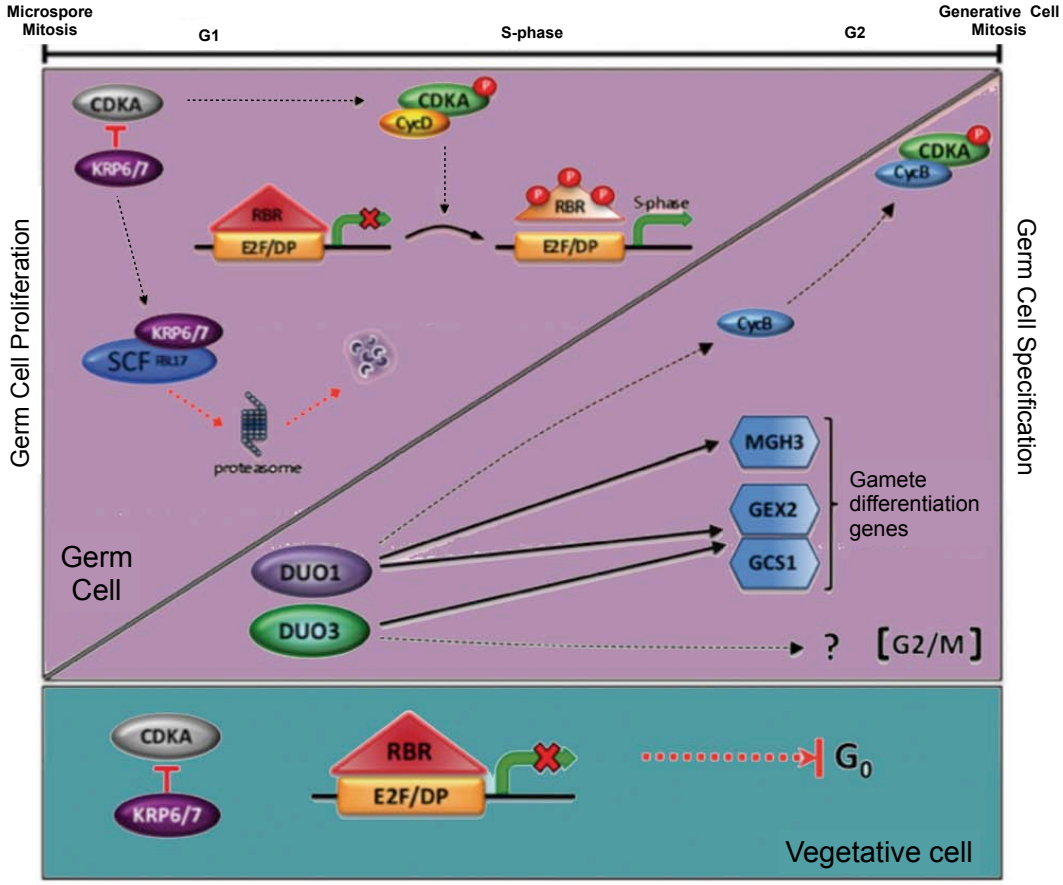


Figure 1.2 - A model of the regulatory events that distinguish germ cell and vegetative cell fate in Arabidopsis pollen. The cell-cycle inhibitors KRP6 and KRP7 are present in both the germ cell and vegetative cell after microspore mitosis, and transient expression of *FBL17* specifically in the germ cell leads to the degradation of these cell cycle inhibitors, allowing CDKA to complex with Cyclin D to phosphorylate RBR. Repression of the E2F/DP pathway is mediated by RBR allowing progression through S-phase. As *FBL17* is not expressed in the vegetative cell, KRP6 and KRP7 levels are maintained, resulting in continued repression of the G1/S transition pathways and preventing entry into the cell cycle. Sperm cell specification begins after microspore mitosis, where the expression of *DUO1* and *DUO3* leads to the activation of germline genes. CYCB1;1 is activated by DUO1 and controls generative cell-cycle progression and entry into mitosis. DUO3 also controls G2/M-phase transition by

unknown mechanisms independent of CYCB1;1. Fully differentiated sperm cells express a full set of germline genes such as GCS1, GEX2 and MGH3 that are required for successful gamete specification (Adapted from BORG and TWELL, 2010).

Pollen transcriptome

Early studies used serial analysis of gene expression (SAGE) to produce 4211 unique tags from Arabidopsis pollen (LEE and LEE, 2003), which was a significant advance when compared with the few genes described before by reverse genetic approaches. However, the most comprehensive analyses of the pollen transcriptome derive from microarray-based studies that have been successfully used to decipher the molecular basis inherent to male gametophyte development. Two different studies used the first available Affymetrix AG Genechip array for Arabidopsis, which represented 26% of the annotated genes in Arabidopsis at the time (BECKER *et al.*, 2003; HONYS and TWELL, 2003). With availability of the second generation of Arabidopsis GeneChips (Affymetrix ATH1), which includes 73% of the currently annotated Arabidopsis genes, the transcriptional profile during male gametophyte development and of mature pollen grains was characterized at a whole-genome scale (HONYS and TWELL, 2004; PINA *et al.*, 2005). Here roughly 7000 genes expressed in pollen were identified, 11% being selectively expressed and 26% enriched compared with those expressed in vegetative tissues (PINA *et al.*, 2005, BECKER AND FEIJÓ 2007). This study highlighted the fact that the pollen transcriptome stores only a limited number of transcripts, functionally skewed towards pollen tube growth and cell-cell signalling processes (Figure 1.3). Moreover, transcriptional profiling of the several independent stages representing microsporogenesis revealed that the number of genes expressed decrease along pollen development, while increasing its complexity and uniqueness (HONYS and TWELL, 2004).

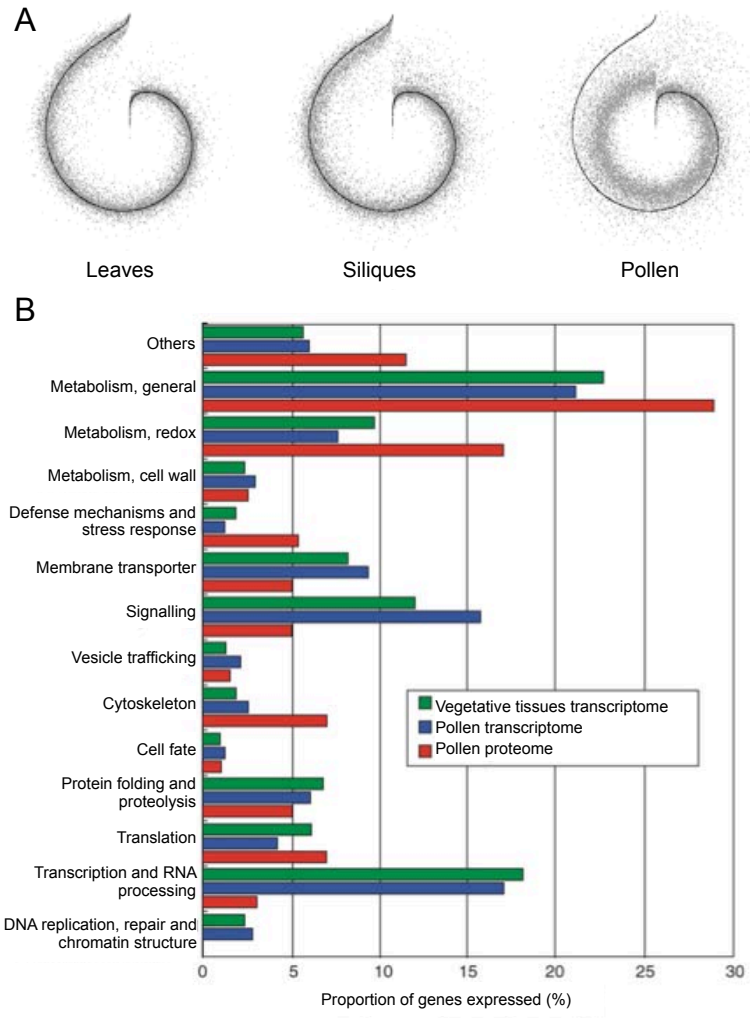


Figure 1.3 - Pollen transcriptomics. (A) Snail view representation of tissue-dependent gene expression patterns. 18,321 genes used for a snail view representation were ranked counterclockwise from top according to decreasing mean expression in seedlings. For each tissue, the mean pattern in seedlings (black line) is co-plotted with the mean pattern of the specific tissue (gray dots). The radius encodes the logarithm of gene expression (values <1 were set to 1 for better visualization). **(B)** Functional categories based on transcriptome and proteome data. A total of 8463 genes represented on the ATH1 GeneChip that were classified into at least one gene ontology category (biological process terms as of September 2003) were regrouped into 14 biological activity classes. The proportion of the genes expressed in each data set assigned to the different activity classes is represented. Transcriptomic data sets are based on PINA *et al*. (2005) with vegetative tissues representing an average of the distribution found in leaves, seedlings and siliques. (Adapted from PINA *et al*., 2005, and BECKER AND FEIJÓ, 2007).

Epigenetic Reprogramming and small RNA silencing

Small RNA pathways and RNA-directed DNA Methylation

In eukaryotes, the expansion of core components of RNA silencing like the RNAse III-like endonuclease Dicer (DCLs), Argonaute proteins (AGOs) and RNA-dependent RNA polymerases (RDRs), allowed the diversification of silencing pathways. A possible explanation for this expansion is that RNAi appeared primarily as a defense mechanism to silence viruses, repeat sequences and transgenes, being then co-opted to regulate expression of endogenous genes. This diversity is illustrated by the distinct functions of the Dicer-Like (DCL) family members. In Arabidopsis, DCL1 produces 21-nt small RNAs called microRNAs (miRNAs) from imperfectly matched stem-loop RNAs, while DCL2 generates 22-nt small interfering RNAs (siRNAs) from double-stranded viral RNAs and transgenes (Figure 1.4). DCL3 is involved in the production of 24-nt siRNAs derived from transposons and transgenes, and DCL4 produces 21-nt *trans-acting* siRNAs (tasiRNAs) from the non-coding transcripts of *TAS* genes, viruses and transgenes (Figure 1.4) (BAULCOMBE, 2004).

The miRNA pathway is an evolutionary conserved and essential pathway that in plants is involved in processes like development, adaptation to biotic and abiotic stresses and hormone responses (MALLORY and BOUCHE, 2008). MicroRNAs act not only through cleavage of highly complementary target mRNAs, but also through translational repression (BRODERSEN *et al.*, 2008; CHEN, 2004), and indirectly through cleavage of *TAS* transcripts giving rise to transacting siRNAs (ta-siRNA) (ALLEN *et al.*, 2005; MONTGOMERY *et al.*, 2008) (Figure 1.4). In Arabidopsis, miRNA biogenesis depends on processing of their stem-loop miRNA precursors by DCL1 (VAZQUEZ, 2006). In the nucleus, miRNA precursors (pri-

miRNAs) are recognized by HYL1, a double-stranded RNA binding protein, and are first converted into an intermediate 'pre-miRNA' and finally into 20–22 nucleotide (nt) miRNA/miRNA* duplexes by DCL1. This duplex is then methylated at the 3' terminus by HEN1 and exported into the cytoplasm by HASTY, an exportin protein. One strand of the duplex (miRNA) is bound by ARGONAUTE1 (AGO1) protein - a component of the RNA-Induced Silencing Complex (RISC) - whereas the miRNA* strand is either degraded or routed into a different AGO complex (CZECH and HANNON, 2011). AGO1 promotes miRNA function by binding the miRNA sequence to the target mRNA and cleavage through base-pair complementarity, or inhibiting translation (BRODERSEN *et al.*, 2008; LANET *et al.*, 2009). While AGO1 is capable of both translational inhibition and transcript cleavage, AGO10 (also named PINHEAD and ZWILLE) acts mainly via translational inhibition (BRODERSEN *et al.*, 2008), or sequestering certain miRNAs preventing their incorporation into a functional AGO1 complex (ZHU *et al*., 2011).

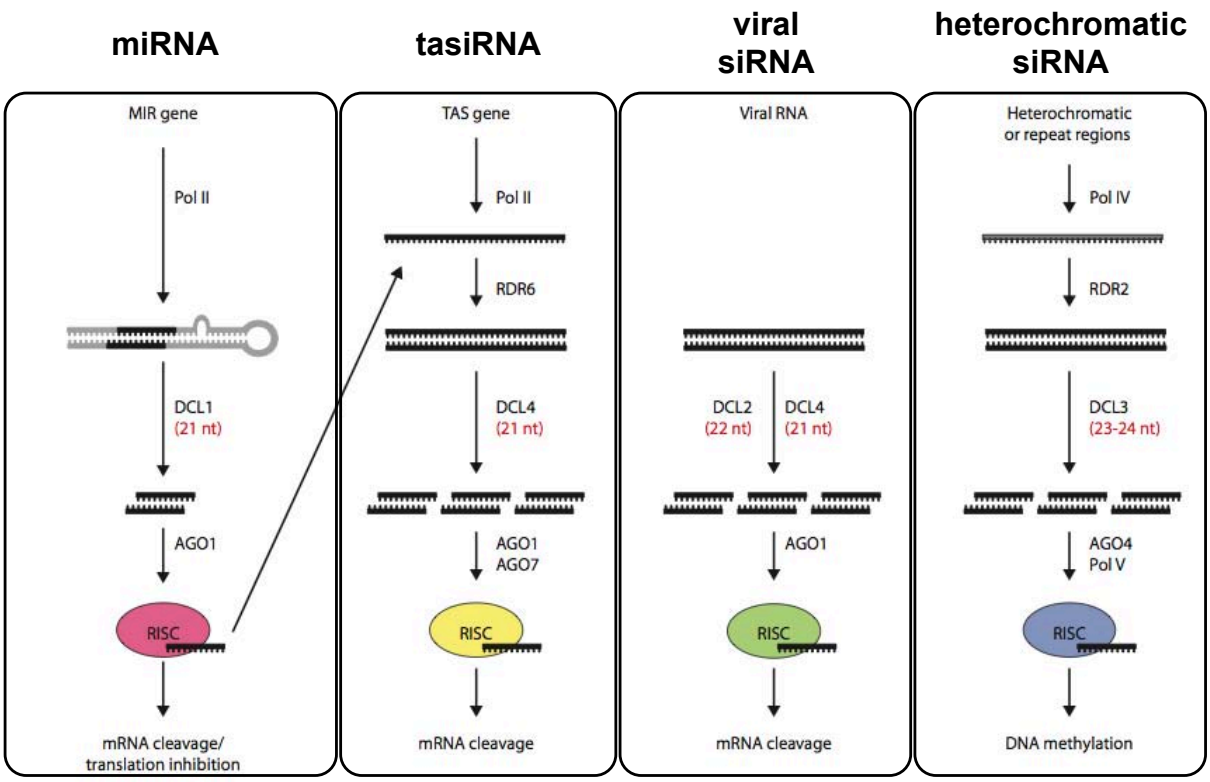


Figure 1.4 - Schematic diagram of small RNA silencing pathways in plants. General overview of the microRNA, trans-acting siRNA, viral siRNA and heterochromatic siRNA pathways. sRNAs are initially processed from perfectly or partially matching double-stranded RNA (dsRNA) by Dicer-like (DCL) proteins. Subsequently, they are incorporated into Argonaute (AGO) proteins to guide mRNA cleavage, translational inhibition or DNA methylation. RNA-dependent RNA polymerases (RDRs) produce dsRNA from transcripts derived from RNA polymerase II (Pol II) and IV (Pol IV) transcription. The members of RDR, DCL and AGO protein families that function in each pathway are indicated. The RNA-induced silencing complex (RISC) is a protein complex that consists of an AGO protein, a small RNA and additional protein factors. The different colors of the RISC complexes indicate a different composition (Adapted from MOLNAR *et al.*, 2011).

Small RNAs are often associated with RNA-directed DNA Methylation (RdDM). In plants, DNA methylation is established at three different sequence contexts (CG, CHG and CHH, where H is A, C or T). While CG and CHG are symmetric and maintained throughout cell divisions by DNA METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE

3 (CMT3), respectively, asymmetric CHH methylation must be established *de novo* in a RNA–directed manner (LAW and JACOBSEN, 2010). As DNA methylation is associated with transcriptional repression and heterochromatin formation, it is preferentially enriched at transposon and repeat sequences (LISTER *et al.*, 2008; ZHANG *et al.*, 2006). Small interfering RNAs (siRNAs) of 24nt size are required for directing DNA methylation at targeted loci. The canonical RdDM pathway is dependent on the plant-specific RNA polymerases Pol IV, Pol V, as well as RDR2, DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), AGO4 and DCL3 (Figure 1.4). Pol IV initiates siRNA biogenesis by producing long ssRNA transcripts, which are recognized by RDR2 to generate dsRNAs. These dsRNAs are then processed into 24-nt siRNAs by DCL3 and loaded into AGO4 (MATZKE *et al.*, 2009). AGO4-bound siRNAs recognize nascent transcripts generated by Pol V, while targeted loci are methylated by DRM2. The recruitment and maintenance of DRM2 to siRNA producing loci is not clear yet, but may result from the interaction between AGO4-bound siRNAs and PolV transcripts (LAW and JACOBSEN, 2010).

Genomic imprinting

Genomic imprinting of paternal and maternal genes is established in the gametes leading to mono-allelic expression after fertilization, and in plants, is especially important for endosperm development. The involvement of small RNA and DNA methylation pathways in this process is an important question. While in mammals imprinting occurs in both embryonic and extra-embryonic lineages, plant imprinted genes known to date show parent-of-origin expression only in the endosperm (KOHLER and WEINHOFER-MOLISCH, 2010). Parent-of-origin expression was also analyzed in the Arabidopsis embryo, showing that activation of the paternal genome is delayed during early embryogenesis, but possibly involving a different

mechanism than those regulating genomic imprinting (AUTRAN *et al.*, 2011). DNA methylation is thought to be the main source of epigenetic marks on imprinting programming during gametogenesis and its maintenance is crucial for stable epigenetic inheritance (MATHIEU *et al.*, 2007; SAZE *et al.*, 2003). This is especially important in the sperm cells that are stalled in S-phase before double fertilization (FRIEDMAN, 1999), the phase of the cell cycle when epigenetic marks such as DNA methylation and histone modifications are established (KLOC *et al.*, 2008). Genetic studies in Arabidopsis suggested that imprinted genes are methylated by default in plant gametes unless selectively demethylated (CHOI *et al.*, 2002; KINOSHITA *et al.*, 2004), but the knowledge remains scarce, as only a few imprinted genes have been fully described. *MEDEA (MEA)*, *FLOWERING WAGENINGEN (FWA)* and *FERTILIZATION INDEPENDENT SEED 2 (FIS2)* are only expressed from the maternal allele (JULLIEN *et al.*, 2006; KINOSHITA *et al.*, 2004; KINOSHITA *et al.*, 1999), and *PHERES1 (PHE1)* is preferentially expressed from the paternal allele (KOHLER *et al.*, 2005; MAKAREVICH *et al.*, 2008). DEMETER (DME), a DNA glycosylase expressed preferentially in the central cell and endosperm, is involved in genome-wide demethylation during endosperm development, including demethylation of *MEA* and *FWA* (GEHRING *et al.*, 2009; GEHRING *et al.*, 2006; HSIEH *et al.*, 2009). The list of imprinted genes increased exponentially in the last years, owing to genome-wide sequencing of allele-specific expression between different Arabidopsis ecotypes (HSIEH *et al.*, 2011; WOLFF *et al.*, 2011).

Epigenetic regulation of transposable elements in pollen

In animals, the stem cell proliferative potential is controlled within a restricted micro-environment of somatic neighbouring cells. This micro-environment, also dubbed niche, is characterized by a diversity of cell-to-cell signalling mechanisms (MORRISON and

SPRADLING, 2008). Recent evidences suggest that epigenetic reprogramming in the plant germline also relies on signals coming from companion non-germ cells. A first insight into small RNA species in Arabidopsis sperm cells resulted from studying transposable element (TE) activity in pollen (SLOTKIN *et al.*, 2009). It was shown that transposons are expressed and transpose in the vegetative nucleus, concordantly with down-regulation of *DECREASE IN DNA METHYLATION 1* (*DDM1)* and loss of TE-derived siRNAs. However, a particular class of 21-nt siRNAs produced from Athila retrotransposons are abundantly accumulated in the sperm cells, where TE loci are highly methylated and transcriptionally silenced (SLOTKIN *et al.*, 2009). These results strongly suggest an active communication between the VN and SCs (Figure 1.5), reviving earlier observations of a cytoplasmic tail that remains after pollen mitosis, connecting the gametes directly to the vegetative nucleus to engender the Male Germ Unit (MCCUE *et al.*, 2011). Future studies should help understanding the functional purpose and extent of an intercellular cross-talk during pollen development and tube growth.

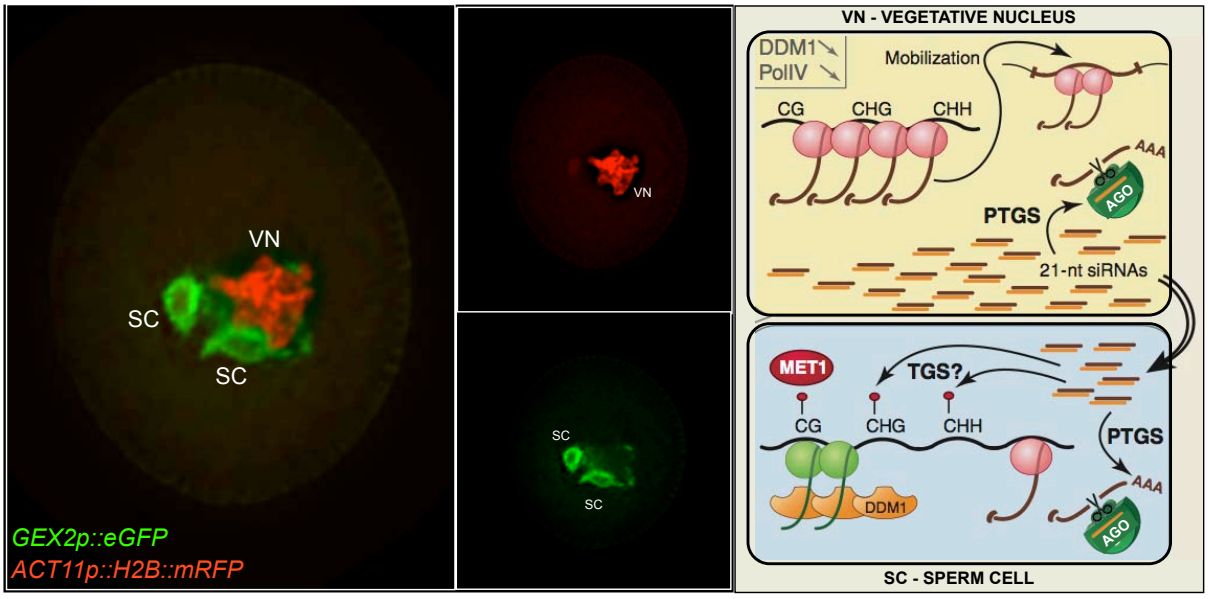


Figure 1.5 – Male Germ Unit and model for siRNA transport. **(A)** *GEX2p::eGFP* and *ACT11p::H2B::mRFP* transgene expression in the sperm cells (SC) and vegetative nucleus

(VN), respectively. Cytoplasmic tail derived from the sperm cells (GFP), connects the germline to the VN (RFP) **(B)** Transposable elements (TEs) are transiently reactivated in the VN, possibly because DDM1 is not expressed leading to DNA hypomethylation and chromatin decondensation. Transcriptionally activated TEs produce siRNAs, which are detected in the SC where TEs are transcriptionally silenced and DDM1 and MET1 are expressed. This observation suggested a mechanism in which reactivation of TEs in the VN lead to the production of epigenetic information that is later transported to the germline. In the VN, 21-nt siRNAs (brown and orange lines) might induce post-transcriptional gene silencing (PTGS) of Pol II-derived transcripts via AGO proteins, so that 21-nt siRNAs delivered to the sperm cells (double-line arrow) can enforce transposon silencing (Adapted from BOURC'HIS and VOINNET, 2010).

A separate study reported that the VN is hypermethylated at non-CG sites at centromeric repeats and a specific class of retrotransposons (SCHOFT *et al.*, 2009). The authors suggested that it could be the result of *DDM1* down-regulation and correspondent demethylation at H3K27me2, which combined with a decondensed chromatin state and loss of centromere identity would facilitate RdDM. This mechanism would explain why the VN is hypermethylated, unlike *ddm1* mutant nuclei. A further study could show that unexpected DME activity in the VN results in active demethylation of TEs surrounding imprinted genes such as *MEA* and *FWA*, as well as the retrotransposon *AtMu1* (SCHOFT *et al.*, 2011), but the biological meaning of DME activity during pollen development and the mechanism controlling reactivation of TE expression remain unknown.

In *Drosophila* and rodents, piRNAs target TE silencing in both male and female germline (BRENNECKE *et al.*, 2007; CARMELL *et al.*, 2007), suggesting that conserved epigenetic mechanisms between plants and animals germline may exist. In addition, "reproductive" Argonaute proteins play equally important roles in plants and animals. In

Drosophila, Argonaute proteins with known germline role include Ago3, Aubergine and Piwi (SETO *et al.*, 2007), while in mice, MIWI, MILI and MIWI2 proteins appear to be functional homologues of Piwi (SETO *et al.*, 2007). The Argonaute gene family in *Arabidopsis* includes 10 identified members. A mutation in *AGO1* has shown a reproductive post-meiotic defect (KIDNER and MARTIENSSEN, 2005), but the molecular function of half the family remains undetermined or not clear. The first Argonaute protein in plants to be shown to have a direct and essential role during reproductive development was MEL1 in rice. *MEL1* expresses solely in reproductive cells, and a loss-of-function mutation was found to disrupt the development of germline precursor anther cells destined to undergo meiosis (NONOMURA *et al.*, 2007). MEL1 functions resonate with that of the Piwi subfamily in animal reproduction, and its closest ortholog in Arabidopsis is *AGO5* that has a function yet to be identified. AGO9 was recently shown to be involved in gamete specification and TE silencing during female gametophyte development (OLMEDO-MONFIL *et al.*, 2010). *AGO9* is also expressed in pollen, but its function during male gametophyte development was not yet reported.

Natural *cis*-antisense siRNAs regulating expression of endogenous genes

Natural *cis*-antisense siRNAs (*cis*-nat-siRNAs) were described in Arabidopsis sperm cells, as playing a crucial role for double fertilization (RON *et al.*, 2010). These siRNAs are produced from the overlapping region of *KPL* and *ARI14* transcripts, and were shown to be involved in *ARI14* down-regulation (RON *et al.*, 2010). The authors proposed an interesting model in which ARI14 lost its E3 ubiquitin ligase activity due to aminoacid mutations along with gene duplication of *ARI 13/14/15*, thus requiring that *ARI14* is efficiently silenced post-transcriptionally to avoid competing with ARI13 protein that is also expressed in sperm. In addition, this study showed that KPL-ARI14 nat-siRNAs are processed by DCL1, and

dependent on a RDR2/SGS3/Pol IV pathway, thus correlating with other previously described nat-siRNAs (BORSANI *et al.*, 2005; KATIYAR-AGARWAL *et al.*, 2006). siRNAs matching to the KPL transcript region were detected by co-expressing *KPL* and *ARI14* ectopically, as well as other 21-nt siRNAs derived from the overlapping region. However, only one 21-nt siRNA that could target *ARI14* was detected in the sperm cell sRNA dataset, originating from the KPL-ARI14 overlap region but outside the predicted area (RON *et al.*, 2010). The fact that other siRNAs were not detected may be explained by their low abundance. Alternatively, this could be a process that occurs earlier during pollen development, thus explaining why *ARI14* transcripts are absent from sperm cells at the mature pollen stage.

Expression and activity of microRNA pathways during male gametophyte development

Arabidopsis pollen at anthesis is tricellular, having originated from a uninucleate microscope after two rounds of mitotic division. Honys and Twell (2004) have shown that throughout male gametogenesis there is a decrease in the number of genes expressed, going along with an enrichment of pollen-specific transcripts. Their data suggested activation of a pollen-specific transcriptional machinery, simultaneously with selective down-regulation of somatic genes as a possible result of miRNA activity. However, this study could not account for the distinct gene expression profiles between the two differentiated cell types in pollen.

Primary evidence indicating that small RNAs could be actively controlling male gametophytic development came from a study in Arabidopsis (KIDNER and MARTIENSSEN, 2005). This work characterized a mutant allele of *AGO1* (*ago1-10*) that is poorly transmitted through the male gametes. MicroRNAs are present and functional in mature pollen, along with expression of the most important genes involved in small RNA silencing such as *DCL1,*

AGO1 and *RDR6* during pollen development (GRANT-DOWNTON *et al.*, 2009a). More recently, two independent studies reported the miRNA repertoire in Arabidopsis pollen using microarrays, qRT-PCR and pyrosequencing, resulting in the identification of approximately 33 known miRNAs (CHAMBERS and SHUAI, 2009; GRANT-DOWNTON *et al.*, 2009b). Grant-Downton et al. (2009b) identified also 7 putative novel miRNAs, including one that targets a F-Box protein specifically expressed in pollen and co-targeted by miR774. Their work showed that ta-siRNAs derived from miR173-directed cleavage of *TAS* transcripts also accumulate in pollen, but intriguingly precursor *TAS* transcripts could not be detected. These results indicate that mature miRNAs and products of miRNA-guided cleavage are present in mature pollen, but the spatial temporal regulation of miRNA activity during pollen development and in plant germ cells remained unexplored.

MicroRNA activity in germ cells has been studied in more detail in animals, leading to the discovery of divergent mechanisms of post-transcriptional regulation between the soma and the germline (KEDDE *et al.*, 2007; MISHIMA *et al.*, 2006). General loss of miRNAs caused by *Dicer* mutations have been shown to impair germline maintenance in *Caenorhabditis elegans* and *Drosophila* (HATFIELD *et al.*, 2005; KNIGHT and BASS, 2001), and some differentiation defects in mouse germ cells (HAYASHI *et al.*, 2008; MAATOUK *et al.*, 2008), but apparently miRNAs are not required for germ cells proliferation in zebrafish (GIRALDEZ *et al.*, 2005). Future studies should certainly elucidate the mechanisms distinguishing miRNA-mediated regulation of multiple target transcripts between different cell lineages, as well as its conservation along the evolution of eukaryotes.

Inheritance of parental microRNAs

A very exciting question is if miRNAs accumulated in sperm cells could be delivered to egg and central cell upon double fertilization, to play a role in the development of early zygote and endosperm, respectively. A similar question is being addressed in the animal field, originating from the discovery of miRNAs and siRNAs in human spermatozoa (OSTERMEIER *et al.*, 2005). In mice, microRNAs delivered to the oocyte can be detected up to the 2-cell stage (AMANAI *et al.*, 2006), but miRNA function in general seems to play no role during oocyte maturation and preimplantation development (SUH *et al.*, 2010).

While the function of miRNAs delivered at fertilization is still a matter of debate (DADOUNE, 2009), there is evidence that paternal mRNAs are transported to oocytes (OSTERMEIER *et al.*, 2004) and that some have a function. For spermatozoon-derived *PSG1* and *HLA-E* mRNA a possible functional role during early embryo development was demonstrated in human (AVENDANO *et al.*, 2009), while in mice the *Kit* transcript delivered by spermatozoa seems to act as an epigenetic modifier of early embryo development (RASSOULZADEGAN *et al.*, 2006). In plants, the strongest indication for a delivery and function of sperm-derived mRNAs upon fertilization to date came from the study on the YDA signaling pathway during embryonic patterning and the characterization of the SHORT SUSPENSOR (SSP) protein. It was found that *SSP* transcripts accumulate in Arabidopsis sperm cells, but are translationally repressed before fertilization, to be translated only during early zygotic development (BAYER *et al.*, 2009). It is possible that in plants paternal miRNAs are also delivered at fertilization, as they could play important functions as signaling molecules, possibly triggering early zygotic patterning and endosperm development. miRNAs would certainly provide an efficient mechanism of reprogramming in the early zygote, before

the gene expression program that defines the maternal to zygotic transition is established. In the light of the parental conflict theory (SPIELMAN *et al.*, 2001, BRANDVAIN AND HAIG 2005) it is also possible that paternally-inherited miRNAs could contribute as an immediate mechanism to regulate the maternally expressed inhibitors of embryo growth. However, small RNA sequencing of egg cells and early zygotes are required to determine if miRNAs accumulated in SCs are indeed inherited by the zygote. The use of artificial miRNA target mimics (TODESCO *et al.*, 2010) to knock down specific miRNA families in sperm cells should clarify if their potential inheritance is of importance.

2. OBJECTIVES AND SYNOPSIS

Evolution of eukaryotes involved adaptive changes in sexual reproduction, which are raising many intriguing questions. Recent studies in the model plant *Arabidopsis thaliana* have paved the way for a genome-wide genetic dissection of the events surrounding pollen development, germ cell differentiation and double fertilization. We aimed to characterize the transcriptome and epigenome of Arabidopsis male gametophyte in combination with well-established reverse genetic approaches, in order to address three main questions: (i) What is the genetic basis underlining male gamete differentiation and fertilization in higher plants, (ii) How do small RNAs and DNA methylation contribute to epigenetic reprogramming and genomic imprinting in the germline, and (iii) What is the role of microRNAs during plant germ cell differentiation and specification, and possibly beyond fertilization?

In order to address these questions we developed tools allowing the isolation of differentiated cell/nuclei in Arabidopsis pollen (vegetative nucleus and sperm cells) by Fluorescence Activated Cell Sorting (FACS). This was possible with the identification of genes preferentially expressed in pollen, whose promoter activity is differentially regulated in the two different cell types that compose the Arabidopsis male gametophyte. These promoters driving expression of fluorescent proteins such as GFP and RFP in sperm cells and vegetative cell, respectively, allowed their separation by FACS (Chapter 2).

Purified fractions were used for whole-genome transcriptome analysis by microarray (Chapter 3), as well as small RNA sequencing and DNA methylation profiling by bisulfite sequencing (Chapter 4). We identified a number of genes and novel pathways with putative

roles during double fertilization and possibly controlling DNA methylation and small RNA in pollen. Small RNA sequencing data were also used to analyze microRNA abundances in pollen and isolated sperm cells (Chapter 5). Comparison of miRNA abundance with transcriptional profiles, allowed the identification of candidate miRNA families possibly involved in germline development and fertilization. To address their function we analyzed mutant lines of genes possibly involved in the small RNA pathway, and used molecular tools to deplete miRNA activity specifically in the germline (Chapter 5).

3. REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Amanai, M., Brahmajosyula, M., and Perry, A.C. (2006). A restricted role for sperm-borne microRNAs in mammalian fertilization. Biol Reprod *75*, 877-884.

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

Avendano, C., Franchi, A., Jones, E., and Oehninger, S. (2009). Pregnancy-specific {beta}-1-glycoprotein 1 and human leukocyte antigen-E mRNA in human sperm: differential expression in fertile and infertile men and evidence of a possible functional role during early development. Human Reprod *24*, 270-277.

Aw, S.J., Hamamura, Y., Chen, Z., Schnittger, A., and Berger, F. (2010). Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in Arabidopsis. Development *137*, 2683-2690.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*, 356-363.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in *Arabidopsis thaliana*. Science *323*, 1485-1488.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol *133*, 713-725.

Becker, J.D., and Feijo, J.A. (2007). How many genes are needed to make a pollen tube? Lessons from transcriptomics. Ann Botany *100*, 1117-1123.

Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization - caught in the act. Trends Plant Sci *13*, 437-443.

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

Boavida, L.C., Borges, F., Becker, J.D., and Feijo, J.A. (2011). Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in Arabidopsis. Plant Physiol *155*, 2066-2080.

Boavida, L.C., Vieira, A.M., Becker, J.D., and Feijo, J.A. (2005b). Gametophyte interaction and sexual reproduction: how plants make a zygote. Int J Dev Biol *49*, 615-632.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. Plant Cell *23*, 534-549.

Borg, M., and Twell, D. (2010). Life after meiosis: patterning the angiosperm male gametophyte. Biochem Soc Trans *38*, 577-582.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell *123*, 1279-1291.

Bourc'his, D., and Voinnet, O. (2010). A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. Science *330*, 617-622.

Brandvain, Y., and Haig, D. (2005). Divergent mating systems and parental conflict as a barrier to hybridization in flowering plants. American Naturalist *166*, 330-338.

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

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science *320*, 1185-1190.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009a). A plant germline-specific integrator of sperm specification and cell cycle progression. PLoS Genet *5*, e1000430.

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P., and Twell, D. (2009b). Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis. Plant Cell *21*, 1940-1956.

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

Chambers, C., and Shuai, B. (2009). Profiling microRNA expression in Arabidopsis pollen using microRNA array and real-time PCR. BMC Plant Biol *9*, 87.

Chen, X.M. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science *303*, 2022-2025.

Chen, Z., Hafidh, S., Poh, S.H., Twell, D., and Berger, F. (2009). Proliferation and cell fate establishment during Arabidopsis male gametogenesis depends on the Retinoblastoma protein. Proc Natl Acad Sci U S A *106*, 7257-7262.

Chen, Z., Tan, J.L., Ingouff, M., Sundaresan, V., and Berger, F. (2008). Chromatin assembly factor 1 regulates the cell cycle but not cell fate during male gametogenesis in *Arabidopsis thaliana*. Development *135*, 65-73.

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

Cinalli, R.M., Rangan, P., and Lehmann, R. (2008). Germ cells are forever. Cell *132*, 559-562.

Czech, B., and Hannon, G.J. (2011). Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet *12*, 19-31.

Dadoune, J.P. (2009). Spermatozoal RNAs: what about their functions? Microsc Res Tech *72*, 536-551.

Dumas, C., Knox, R.B., and Gaude, T. (1985). The Spatial Association of the Sperm Cells and Vegetative Nucleus in the Pollen Grain of Brassica. Protoplasma *124*, 168-174.

Feijo, J.A. (2010). The mathematics of sexual attraction. J Biol *9*, 18.

Friedman, L., Santa Anna-Arriola, S., Hodgkin, J., and Kimble, J. (2000). gon-4, a cell lineage regulator required for gonadogenesis in *Caenorhabditis elegans*. Dev Biol *228*, 350-362.

Friedman, W.E. (1999). Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development *126*, 1065-1075.

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

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

Ghosh, D., and Seydoux, G. (2008). Inhibition of transcription by the *Caenorhabditis elegans* germline protein PIE-1: genetic evidence for distinct mechanisms targeting initiation and elongation. Genetics *178*, 235-243.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. Science *308*, 833-838.

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

Grant-Downton, R., Le Trionnaire, G., Schmid, R., Rodriguez-Enriquez, J., Hafidh, S., Mehdi, S., Twell, D., and Dickinson, H. (2009b). MicroRNA and tasiRNA diversity in mature pollen of *Arabidopsis thaliana*. BMC Genomics *10*, 643.

Guignard, M.L. (1899). Sur les anthérozoïdes et la double copulation sexuelle chez les végétaux angiospermes. Rev Gén de Botanique, 129–135.

Gusti, A., Baumberger, N., Nowack, M., Pusch, S., Eisler, H., Potuschak, T., De Veylder, L., Schnittger, A., and Genschik, P. (2009). The *Arabidopsis thaliana* F-box protein FBL17 is essential for progression through the second mitosis during pollen development. PLoS ONE *4*, e4780.

Haerizadeh, F., Singh, M.B., and Bhalla, P.L. (2006). Transcriptional repression distinguishes somatic from germ cell lineages in a plant. Science *313*, 496-499.

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

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730-733.

Hatfield, S.D., Shcherbata, H.R., Fischer, K.A., Nakahara, K., Carthew, R.W., and Ruohola-Baker, H. (2005). Stem cell division is regulated by the microRNA pathway. Nature *435*, 974-978.

Hayashi, K., Chuva de Sousa Lopes, S.M., Kaneda, M., Tang, F., Hajkova, P., Lao, K., O'Carroll, D., Das, P.P., Tarakhovsky, A., Miska, E.A.*, et al.* (2008). MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. PloS ONE *3*, e1738.

Higashiyama, T., Kuroiwa, H., Kawano, S., and Kuroiwa, T. (1997). Kinetics of double fertilization in Torenia fournieri based on direct observations of the naked embryo sac. Planta *203*, 101-110.

Higashiyama, T., Kuroiwa, H., Kawano, S., and Kuroiwa, T. (2000). Explosive discharge of pollen tube contents in Torenia fournieri. Plant Physiol *122*, 11-13.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S., Kuroiwa, H., and Kuroiwa, T. (2001). Pollen tube attraction by the synergid cell. Science *293*, 1480-1483.

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

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

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

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

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

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J., and Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci U S A *103*, 18002-18007.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Orom, U.A.*, et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

Kidner, C.A., and Martienssen, R.A. (2005). The role of ARGONAUTE1 (AGO1) in meristem formation and identity. Dev Biol *280*, 504-517.

Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D., and Nam, H.G. (2008). Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. Nature *455*, 1134-1137.

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

Kinoshita, T., Yadegari, R., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. Plant Cell *11*, 1945-1952.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol *18*, 490-495.

Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. Science *293*, 2269-2271.

Kohler, C., Page, D.R., Gagliardini, V., and Grossniklaus, U. (2005). The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat Genet *37*, 28-30.

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

Lanet, E., Delannoy, E., Sormani, R., Floris, M., Brodersen, P., Crete, P., Voinnet, O., and Robaglia, C. (2009). Biochemical evidence for translational repression by Arabidopsis microRNAs. Plant Cell *21*, 1762-1768.

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

Lee, J.Y., and Lee, D.H. (2003). Use of serial analysis of gene expression technology to reveal changes in gene expression in Arabidopsis pollen undergoing cold stress. Plant Physiol *132*, 517-529.

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

Maatouk, D.M., Loveland, K.L., McManus, M.T., Moore, K., and Harfe, B.D. (2008). Dicer1 is required for differentiation of the mouse male germline. Biol Reprod *79*, 696-703.

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

Mallory, A.C., and Bouche, N. (2008). MicroRNA-directed regulation: to cleave or not to cleave. Trends Plant Sci *13*, 359-367.

Martinho, R.G., Kunwar, P.S., Casanova, J., and Lehmann, R. (2004). A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. Curr Biol *14*, 159-165.

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

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009). RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol *21*, 367-376.

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

Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. Curr Biol *16*, 2135-2142.

Molnar, A., Melnyk, C., and Baulcombe, D.C. (2011). Silencing signals in plants: a long journey for small RNAs. Genome Biol *12*, 215.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell *133*, 128-141.

Morrison, S.J., and Spradling, A.C. (2008). Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. Cell *132*, 598-611.

Nawaschin, S.G. (1898). Resultate einer revision der befruchtungsvorgaenge bei Lilium martagon und Fritillaria tenella. Bul Acad Imp Sci St Petersburg, 377–382.

Nonomura, K., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurata, N. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell *19*, 2583-2594.

Nowack, M.K., Grini, P.E., Jakoby, M.J., Lafos, M., Koncz, C., and Schnittger, A. (2006). A positive signal from the fertilization of the egg cell sets off endosperm proliferation in angiosperm embryogenesis. Nat Genet *38*, 63-67.

Oh, S.A., Johnson, A., Smertenko, A., Rahman, D., Park, S.K., Hussey, P.J., and Twell, D. (2005). A divergent cellular role for the FUSED kinase family in the plant-specific cytokinetic phragmoplast. Curr Biol *15*, 2107-2111.

Oh, S.A., Pal, M.D., Park, S.K., Johnson, J.A., and Twell, D. (2010). The tobacco MAP215/Dis1-family protein TMBP200 is required for the functional organization of microtubule arrays during male germline establishment. J Exp Bot *61*, 969-981.

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

Ostermeier, G.C., Goodrich, R.J., Moldenhauer, J.S., Diamond, M.P., and Krawetz, S.A. (2005). A suite of novel human spermatozoal RNAs. J Androl *26*, 70-74.

Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., and Krawetz, S.A. (2004). Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature *429*, 154.

Pagnussat, G.C., Alandete-Saez, M., Bowman, J.L., and Sundaresan, V. (2009). Auxin-dependent patterning and gamete specification in the Arabidopsis female gametophyte. Science *324*, 1684-1689.

Park, S.K., Howden, R., and Twell, D. (1998). The *Arabidopsis thaliana* gametophytic mutation gemini pollen1 disrupts microspore polarity, division asymmetry and pollen cell fate. Development *125*, 3789-3799.

Pastuglia, M., Azimzadeh, J., Goussot, M., Camilleri, C., Belcram, K., Evrard, J.L., Schmit, A.C., Guerche, P., and Bouchez, D. (2006). gamma-tubulin is essential for microtubule organization and development in Arabidopsis. Plant Cell *18*, 1412-1425.

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

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469-474.

Ron, M., Saez, M.A., Williams, L.E., Fletcher, J.C., and McCormick, S. (2010). Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev *24*, 1010-1021.

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

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

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

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

Seto, A.G., Kingston, R.E., and Lau, N.C. (2007). The coming of age for Piwi proteins. Mol Cell *26*, 603-609.

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

Spielman, M., Vinkenoog, R., Dickinson, H.G., and Scott, R.J. (2001). The epigenetic basis of gender in flowering plants and mammals. Trends Genet *17*, 705-711.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. Curr Biol *20*, 271-277.

Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D. (2010). A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. PLoS Genet *6*, e1001031.

Twell, D., Park, S.K., and Lalanne, E. (1998). Asymmetric division and cell-fate determination in developing pollen. Trends Plant Sci *3*, 305-310.

Vazquez, F. (2006). Arabidopsis endogenous small RNAs: highways and byways. Trends Plant Sci *11*, 460-468.

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

Yu, H.S., Hu, S.Y., and Zhu, C. (1989). Ultrastructure of Sperm Cells and the Male Germ Unit in Pollen Tubes of Nicotiana-Tabacum. Protoplasma *152*, 29-36.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.H., Liou, L.W., Barefoot, A., Dickman, M., and Zhang, X. (2011). Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell *145*, 242-256.

CHAPTER 2

Purification of Microspores, Differentiated Sperm Cells and Vegetative Nucleus by FACS

1. INTRODUCTION

Plant germ cells differentiate late in ontogeny within gametophytes. In the male gametophyte (pollen grain), this process requires post-meiotic microspores to undergo two subsequent mitotic divisions, giving rise to the male germ unit (MGU) that is composed by the vegetative cell nucleus (VN) and two sperm cells (SC). The vegetative cell arrests cell cycle progression upon pollen mitosis I (PM I), while the two sperm cells originated from pollen mitosis II (PM II) are specified into gametes. The sperm cell lineage is immersed in the pollen vegetative cell, being dependent on this companion cell for transportation towards the embryo sac. However, the gametes are known to have their own unique molecular repertoire (reviewed in TWELL, 2011). Several studies reported different methodologies to isolate generative cells of *Lilium longiflorum* (lily) (OKADA *et al*., 2006; XU *et al*., 1998) and sperm cells of *Oryza sativa* (rice) (GOU *et al*., 2001), *Zea mays* (maize) (ENGEL *et al*., 2003), and more recently from *Nicotiana tabacum* (tobacco) (XIN *et al*., 2011) and the dimorphic sperm cells of *Plumbago zeylanica* (GOU *et al*., 2009). Technical difficulties in obtaining sufficient amounts of pure biological material constituted a major problem for purification of Arabidopsis sperm cells.

Microgametogenesis in Arabidopsis has proven to be an excellent model to identify novel mechanisms controlling cell cycle transitions, cell fate specification and epigenetic reprogramming (BERGER and TWELL, 2011). However, such studies highlighted the importance of analyzing these two cell types independently, as they activate different transcriptional machineries and establish distinct epigenetic states (BORGES *et al.*, 2008;

BROWNFIELD *et al.*, 2009; SLOTKIN *et al.*, 2009). The VN may participate actively in controlling heritable epigenetic modifications in the germline, as it activates expression of transposable elements (TEs) to produce a specific class of small interfering RNAs (siRNAs) that accumulate in the gametes (SLOTKIN *et al.*, 2009). To understand epigenetic reprogramming throughout pollen development we needed a simple and powerful method to co-purify the two differentiated types of nuclei in mature pollen, as well as their precursor microspore.

Here we describe a fast and reliable method to isolate Arabidopsis SC, VN and microspores, based on a combination of previously reported techniques to isolate mature pollen using high-speed cell sorting (BECKER *et al.*, 2003). Our first study described a fluorescent marker line specifically labeling differentiated SC in mature pollen, which allowed their FACS-purification and genome-wide transcriptional profiling (Chapter 3). This method was improved by using stronger fluorescent markers and more efficient methodologies for pollen disruption, resulting in larger amounts of highly pure material at very high rates. As such, it also allowed co-purification of the VN from the same genetic background.

2. RESULTS AND DISCUSSION

Fluorescence-activated sperm cell sorting

Arabidopsis sperm cells expressing eGFP (ENGEL *et al.*, 2005) were isolated using a high-speed cell sorter, from crude fractions resulting from a multi-step purification protocol (see Material and Methods). In brief, inflorescence tissue was collected into a glass beaker and pollen was released in 300mL of sperm extraction (SE) buffer. This suspension was then filtered through Miracloth, before pollen was concentrated on a 10um mesh by vacuum filtration. The mesh is washed with 2mL of SE buffer, and additionally filtered through a 28um mesh into a glass homogenizer to disrupt pollen grains. Released sperm cells in solution were selected based on their size and intracellular refringence (as assessed by the angle of their light scatter properties), their GFP signal, and presence of DNA. Low granulosity (low SSC) and GFP-positive signals were used to identify the sperm cell population (R1 in Figure 2.1A). A low FSC signal (small particles) within the GFP positive population further discriminated between sperm cells and other small entities within the sample (R2 in Figure 2.1B). To assure exclusion of debris within the small-sized population containing GFP-positive sperm cells, we used DRAQ5, a live cell DNA marker (R3 in Figure 2.1C). To guarantee purity, sperm cells were sorted using a logical combination of regions R2 and R3, and we verified that cells sorted from region R1 only yielded the same population (results not shown). Each session yielded around 100,000 cells, from which total RNA was immediately isolated and processed for gene expression analysis by microarray (Chapter 3).

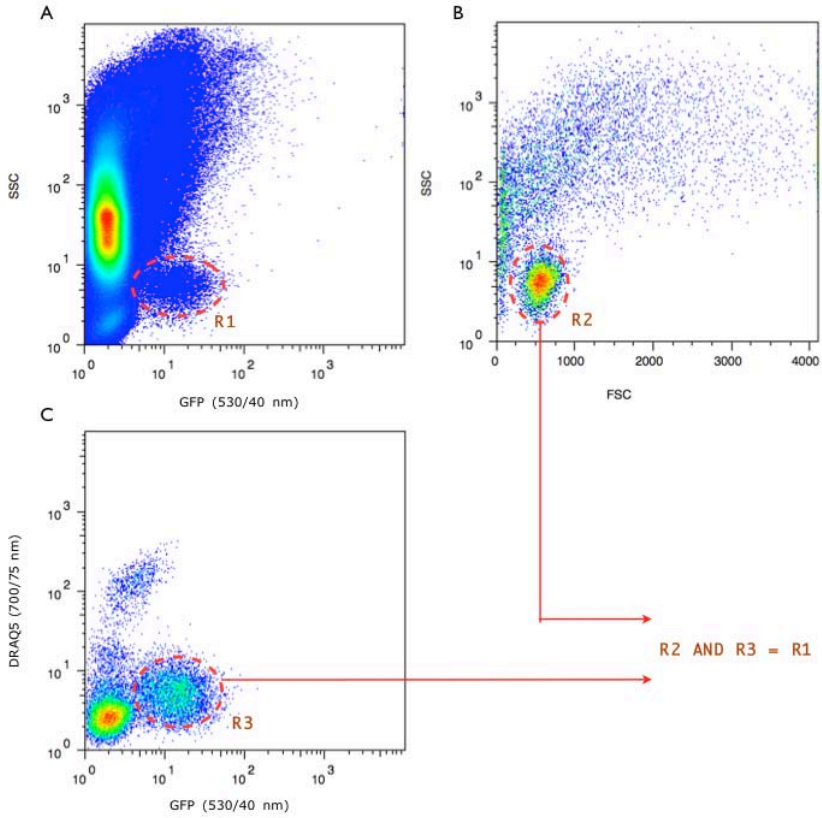


Figure 2.1 - Fluorescence-activated sperm cell sorting. FACS purification based on cell size (FSC), intracellular complexity (SSC), GFP signal, and presence of intracellular DNA, via DRAQ5 staining. **(A)** Low granulosity (low SSC) and GFP positive signals were used to identify the sperm cell population (R1) from the total population. **(B)** To guarantee purity, a low FSC signal (small particles; R2) **(C)** within the GFP/ DRAQ5 double positive population (R3) were used to exclude other small particles.

A high degree of purity of the sorted sperm cell fractions was verified routinely before and during sorting by wide-field fluorescence microscopy (Figure 2.2 A-F), and later by RT-PCR analysis on known vegetative nucleus- and sperm cell-specific genes (Chapter 3, Figure 3.2). The viability of isolated sperm cells was confirmed by fluorescein diacetate staining (Figure 2.2G).

Even though this method allowed obtaining pure and viable sperm cell fractions, it is laborious, time consuming and not very efficient, considering the amount of plants needed as starting material. In addition, the need for DRAQ5 or other DNA dyes may become problematic for certain down-stream applications such as chromatin IPs, as it is known to interfere with chromatin condensation and nucleosome positioning (MARI *et al.*, 2010; WOJCIK and DOBRUCKI, 2008).

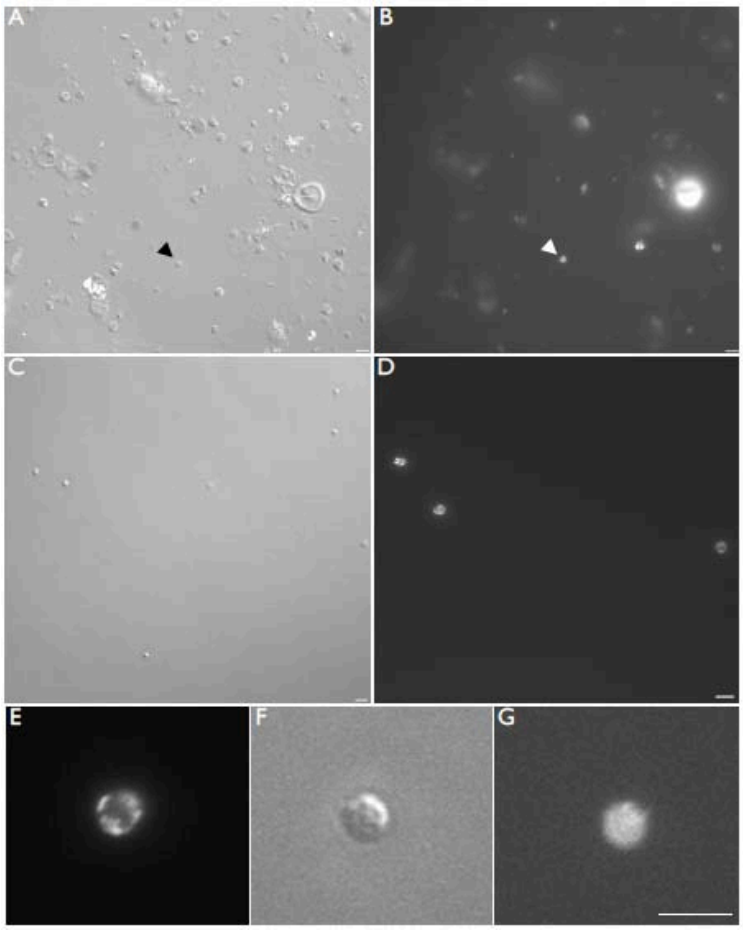


Figure 2.2 - Visualization of FACS-purified sperm cells. Wide-field fluorescence microscopy was used to visualize Arabidopsis sperm cells expressing *AtGEX2::eGFP* (Engel et al., 2005) before **(B)** and after FACS purification **(D)**. Differential interference microscopy (DIC) confirmed that the debris in the filtrate before FACS **(A)** was removed after sorting **(C)**. For DIC imaging of FACS-purified sperm cells **(C)**, we captured and merged several

images along the optical axis. A higher magnification of a sorted sperm cell shows GFP fluorescence **(E)**, cell-shape integrity by DIC microscopy **(F)**, and cell viability using fluorescein diacetate staining **(G)**. The bars represent 5 microns and the arrowheads are pointing to sperm cells.

Co-purification of sperm cell and vegetative nuclei

In order to improve and simplify the SC-sorting method and additionally attempt to co-purify the VN from the same genetic background, we generated a transgenic line expressing distinct fluorescent proteins in both nuclei. The *ACT11* promoter driving histone H2B fused to mRFP was used as a VN marker (ROTMAN *et al.*, 2005). A homozygous plant harboring the *ACT11p::H2B::mRFP* transgene in Col-0 background was crossed with a sperm-specific marker line of the same ecotype, harboring a *MGH3p-MGH3::eGFP* construct, which encodes a male germline-specific histone variant (INGOUFF *et al.*, 2007; OKADA *et al.*, 2005). *MGH3* is expressed as early as bicellular pollen, and is highly abundant in the sperm nuclei of mature pollen (Figure 2.3). Progeny resulting from this cross was allowed to segregate by self-pollination, in order to obtain double homozygous plants. Studying the expression pattern of both transgenes throughout pollen development revealed that *ACT11p::H2B::mRFP* is initially expressed in the generative cell (GC) of bicellular pollen, but preferentially expressed in the VN at the mature pollen stage (Figure 2.3).

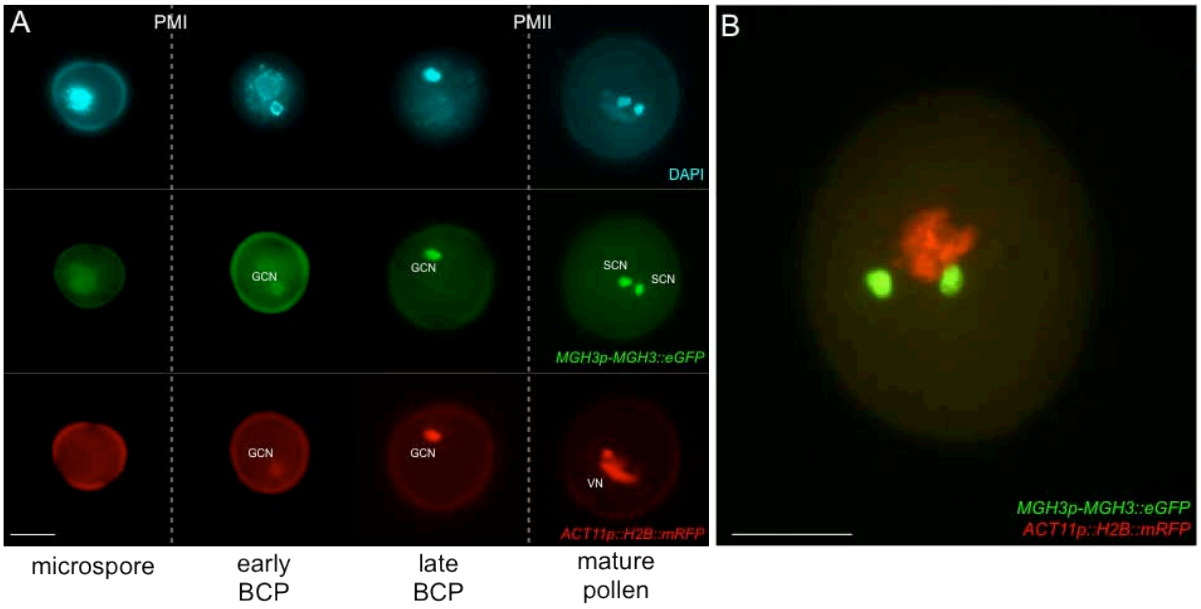


Figure 2.3 - Expression pattern of GFP- and RFP-fusion proteins during pollen development. (A) *MGH3p-MGH3::eGFP* localizes in the generative cell nucleus (GCN) after the first pollen mitosis, and is strongly accumulated in the sperm cell nucleus (SCN) of mature pollen. *ACT11p::H2B::mRFP* is initially expressed in the GCN until late bicellular pollen (BCP). After the second pollen mitosis the expression of this transgene is down-regulated in the germline, and it becomes strongly expressed in the vegetative nucleus (VN). **(B)** Merged magnification of a mature pollen grain expressing both transgenes. Scale bars: 10μm.

Pollen grains were collected from open flowers in a 2mL tube, by vortexing with nuclei extraction buffer (see Materials and Methods). Released pollen was then disrupted by vortexing with glass beads. Enriched filtrates containing SC and VN released from broken pollen were co-purified by FACS based on their distinct fluorescence properties (Figure 2.4A). We inspected purity of sorted samples by microscopy (Figure 2.4B) and RT-PCR on SC and VC-specific transcripts. *MGH3* is only expressed in the SC and was not detected in the sorted VN fraction, while *VEX1*, being a VC-expressed gene was equally not detected in the SC fraction (Figure 2.4D), thus indicating that both fractions were pure. Re-analyzing

sorted populations stained with 4',6-diamidino-2-phenylindole (DAPI) confirmed that the SC and VN populations were consistently more than 99% pure (Figure 2.4C). The ratio SC/VN before and after sorting was consistently 2:1, thus demonstrating a good recovery for both types of nuclei after pollen disruption (data not shown).

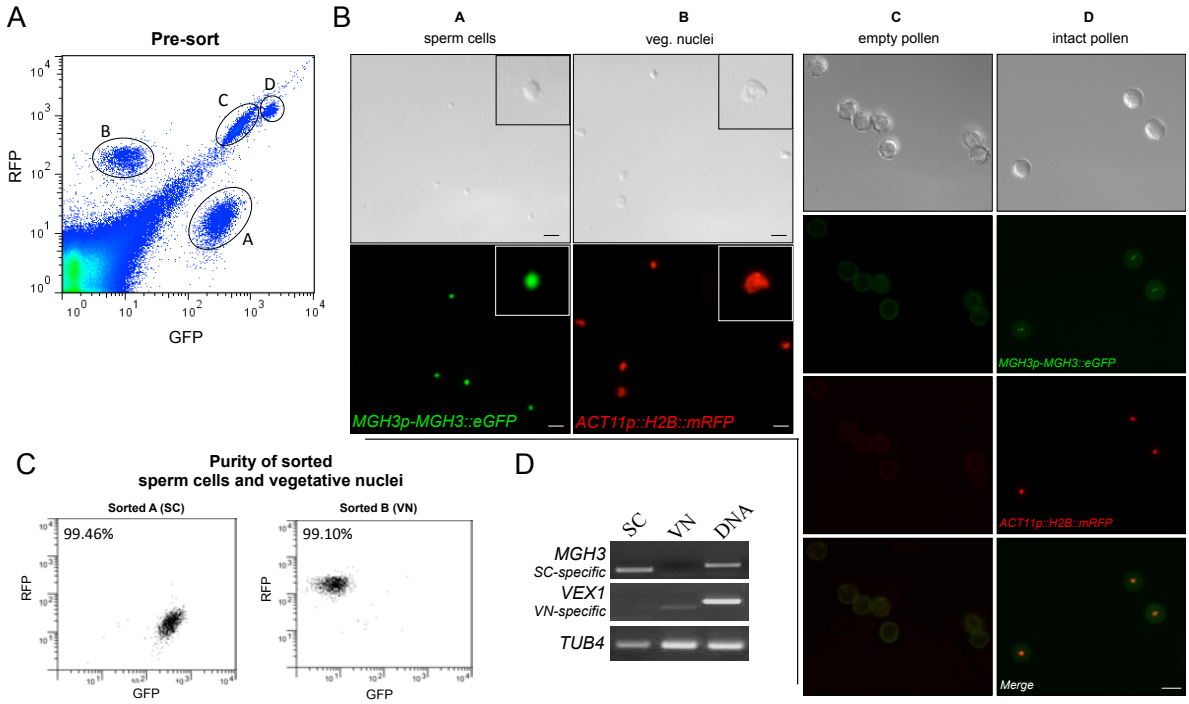


Figure 2.4 - Purification of sperm cells and vegetative nuclei by FACS. (A) Four distinct cell populations are highlighted in the filtrate pre-sorting. Sperm cells (SC) are identified based on their strong GFP signal – Population A, whereas vegetative nuclei (VN) separate towards the opposite axis based on strong RFP signal – Population B. In addition the filtrate contains empty – Population C, and intact pollen grains – Population D. **(B)** Purity of sorted SC and VN fractions was confirmed by DIC and fluorescence microcopy; scale bar: 10um. Populations C and D were confirmed to represent empty and intact pollen, respectively; scale bar: 30um **(C)** Sorted SC and VN samples were run through a flow cytometer to check for purity. Purity is determined based on all DNA-containing particles present in the sorted sample stained with DAPI or Hoechst. **(D)** RT-PCR analyses confirmed that each fraction is enriched for cell-specific transcripts (*MGH3* for SC and *VEX1* for VN), and devoid of contaminating RNAs. *TUB4* was used as control.

These results showed that using fluorescent proteins under the control of strong cell-specific promoters eliminates the need of DNA dyes that could enhance noise and interfere with downstream DNA analysis. SybrGreen staining was previously used to isolate VN and SC nuclei by flow cytometry (SCHOFT *et al.*, 2011; SCHOFT *et al.*, 2009). However, we considered that this method does not provide a clear separation of both fractions, which compromises purity. In addition, we believe that the nuclear markers described in our study are more suitable for live cell imaging, when it comes to identify mutant phenotypes associated with defects in chromatin condensation and MGU dynamics during pollen tube growth and double fertilization.

In order to complement the transcriptional profiling of sperm cells and total pollen (Chapter 3), genomic DNA was extracted from isolated SC and VN to perform whole-genome bisulfite sequencing. DNA methylation profiling throughout pollen development is crucial to understand epigenetic reprogramming along gametogenesis, and how are these distinct features correlated with reactivation of TEs in the VN (SLOTKIN *et al.*, 2009). These findings are presented and discussed in Chapter 4.

Microspore purification by FACS

To understand the genetic and epigenetic mechanisms that control VN and SC differentiation it is essential to analyze their precursor microspore cell. Previous methods to isolate Arabidopsis microspores relied on Percoll density gradients (HONYS and TWELL, 2004), and would not provide a sufficiently pure fraction suitable for genomic analysis at the DNA level, besides the problem of relatively low yield. We used closed flower buds of wild type plants to collect all different stages of pollen development, by grinding in pollen

extraction buffer and filtering through 28um mesh (see details in Materials and Methods). Filtered spores were concentrated by centrifugation, and sorted based on their small size and autofluorescent properties (Figure 2.5A). Sorted microspore population was analyzed by microscopy and confirmed to contain only microspores (Figure 2.5B). We used isolated microspores to analyze DNA methylation at a genome-wide scale by bisulfite sequencing. The results are presented and discussed in Chapter 4.

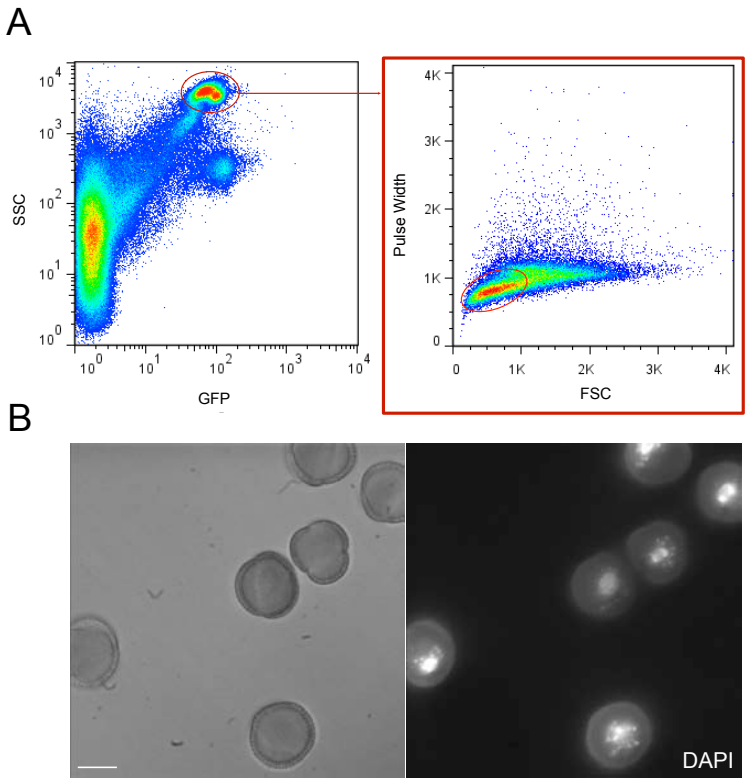


Figure 2.5 - Microspore sorting. (A) Pollen population is characterized by an elevated high angle scatter (SSC) and autofluorescence (observed in the GFP channel). Within this population, microspores (right panel, circular gate) can be differentiated from bicellular and tricellular pollen by their characteristic smaller size, captured by a diminished low angle scatter (FSC) and time-of-flight (Pulse Width), as compared to other stages of pollen development. **(B)** Sorted microspores were inspected by microscopy to confirm purity and integrity, as revealed by DAPI staining. Scale bar: 10μm.

3. CONCLUSIONS

Gametogenesis and fertilization involve a series of genetic and epigenetic reprogramming events, often characterized by transcriptional activation of germline genes and repression of somatic gene expression, as well as dynamic changes of DNA methylation on a genome-wide scale. A comprehensive interpretation of such phenomena relies on analyzing isolated cell populations, and as such it demands methods for efficient and robust purification of each cell type involved in the process. We found that FACS coupled to pollen nuclei-specific fluorescent markers provides fast and reliable results to isolate Arabidopsis sperm cells and vegetative nucleus fractions that are consistently more than 99% pure. In addition, we found that FACS also allows the isolation of pure microspore fractions based on their small size and autofluorescent properties. Combined, these methods provide a complete set of tools that will certainly boost our understanding on the key pathways involved in microgametogenesis. These methods were successfully used to study gene expression by microarray (Chapter 3), DNA methylation profiling by bisulfite sequencing (Chapter 4) and small RNA sequencing (Chapter 5), thus foreseeing its downstream applicability for challenging molecular analysis such as ChIP.

4. MATERIALS AND METHODS

Plant material and growth conditions

Transgenic Arabidopsis (*AtGEX2::eGFP)* plants (ENGEL *et al.*, 2005), were used for both sperm cell and pollen isolation. Plants were sown on soil and grown for 8 weeks in short-day conditions (8 h light at 21ºC-23ºC) and then transferred to long-day conditions (16 h light) to induce flowering. *MGH3p-MGH3::eGFP* was obtained by cloning 1.2kb upstream of the MGH3 transcriptional start site, together with the MGH3 coding sequence without the stop codon, into the pMDC107 vector (CURTIS and GROSSNIKLAUS, 2003) by gateway cloning (in collaboration with R. Keith Slotkin). Primers used for *MGH3p-MGH3::eGFP* cloning are presented in Table S2.1. Double homozygous plants harbouring *MGH3p-MGH3::eGFP* and *ACT11p::H2B::mRFP* (ROTMAN *et al.*, 2005) transgenes, were obtained by crossing individual homozygous lines

Isolation of Arabidopsis sperm cells expressing *AtGEX2::eGFP*

Inflorescences from transgenic plants expressing *AtGEX2*::*eGFP* were collected (filling approximately 300mL of a glass beaker) and placed in a humid chamber at room temperature for 1 hour to hydrate. The sperm extraction (SE) buffer (1.3 mM H_3BO_3, 3.6 mM $CaCl_2$, 0.74 mM KH_2PO_4, 438 mM sucrose, 7 mM MOPS, 0.83 mM $MgSO_4$, in double distilled water), was adjusted to pH 6.0 with NaOH, filter-sterilized (0.22 μm filtration, 150 mL Bottle Top Filter with a 45 mm neck (Corning) and vacuum-degassed. Inflorescences were immersed and mixed in 500 mL of SE buffer, until the pollen grains were released, yielding a yellowish solution. To remove flowers and other plant tissues, the solution was filtered through Miracloth (Calbiochem), then the pollen was concentrated by passing the filtrate through a

10µm pore size mesh filter (SEFAR AG) using a vacuum pump (KNF). The material collected on the filter was subsequently washed with 1-2mL of buffer, to release pollen and impurities, and then filtered through a 28µm mesh filter (SEFAR AG), to dispose of debris larger than pollen, directly into a glass homogenizer (Kontes). Pollen grains were subsequently disrupted by applying 3 circular up and down movements. The resulting solution, containing free sperm cells as well as both burst and intact pollen grains, was filtered through a 15µm pore size mesh filter (SEFAR AG) in order to eliminate most of the debris and the remaining intact pollen grains. To the filtrate enriched in sperm cells, 1µM DRAQ5 (Biostatus) was added in order to stain genomic DNA, and maintained at 4ºC before FACS.

Improved method to purify vegetative and sperm nuclei

Open flowers were collected into a 2 mL tube, filling with approximately 150mg of fresh weight material. Flowers were vigorously vortexed (IKA, maximum speed) in 1.5 mL of Galbraith buffer (45mM $MgCl_2$, 30mM Sodium Citrate, 20mM MOPS, 1% Triton-100, pH to 7.0) for 3 minutes, at room temperature. This crude fraction was then filtered through a 30um mesh (Partec, CellTrics) into a 1.5 mL eppendorf tube containing 10uL of glass beads (435-600um, Sigma), and vortexed (IKA, maximum speed) for additional 3 minutes in order to break mature pollen grains. The cell/nuclei released into solution are ready for FACS at this point, which includes debris resulting from broken pollen. This fraction may be additionally filtered through a 15um mesh (SEFAR) to exclude debris and intact pollen.

Isolation of Microspores from young flower buds

Young flower buds (approximately 20mL) were gently grinded using mortar and pestle in 10mL of pollen extraction buffer (10mM $CaCl_2$, 2mM MES, 1mM KCl, 1% H_3BO_3, 10% Sucrose, pH 7.5) in order to release the spores. This crude fraction was initially filtered through Miracloth (Calbiochem) to remove bigger debris, and concentrated by centrifugation (800g, 5 min) in 1.5mL eppendorf tubes. The resulting yellowish pellet enriched in microspores was resuspended in 1.5mL of PEB, and filtered through a 20um mesh (Partec, CellTrics) before FACS.

Fluorescence Activated Cell Sorting

GEX2p::eGFP sperm cells

Fluorescence-activated sperm cell sorting was performed in a Moflo High-Speed Cell Sorter (Dako-Cytomation), using a 100 μm ceramic nozzle with 30 psi sheath pressure, a 488 nm laser line from a Coherent Sapphire 488-200 CDRH laser for eGFP excitation, and a 632.8 nm laser line from a Spectra-Physics 107B 25mW HeNe laser to excite DRAQ5. GFP and DRAQ5 were detected using a 530/40 nm and a 700/75 nm HQ band pass filter, respectively. SE buffer was used as the sheath solution, both for hydrodynamic stability and subsequent analysis of sorted sperm cells. Sperm cells viability after sorting was evaluated by fluorescein diacetate staining according to a procedure previously described for pollen (HESLOP-HARRISON and HESLOP-HARRISON, 1970). For subsequent RNA extraction, sperm cells were sorted directly into RNeasy extraction buffer RLT (Qiagen) or Tri Reagent (Sigma).

MGH3p-MGH3::eGFP sperm cells and *ACT11p::H2B::mRFP* vegetative nuclei

Fluorescent activated cell sorting was carried out with a MoFlo (Beckman Coulter, Fort Collins, USA) with a 488nm laser (Spectraphysics, Coherent) at 140mW used for scatter measurements (Low angle or Forward Scatter, and High angle or Side Scatter; FSC and SSC, respectively) and for GFP excitation, and a 561nm laser (CrystaLaser) at 38mW for RFP excitation. GFP and RFP were detected using the following bandpass filters, 530/40nm and 630/75nm, respectively. FSC was used for triggering, and threshold had to be low to avoid missing the sperm cell population, whose size is on average 2.5um in diameter. Phosphate Buffer Saline (PBS) was used as sheath, and run at a constant pressure of 60psi. Frequency of drop formation was approximately 96,000Hz. Even though pollen was present in the sample (broken and intact) it did not interfere with drop formation or break-off, as we were able to sort pollen under the same conditions as for SC and VN. Sorting rates were typically just below 1 million SC and 0.5 million VN per hour, i.e. an average rate of 250 and 125 SC and VN per second, respectively. Purity was determined by running aliquots of sorted cells in a CyAn ADP flow cytometer (Beckman Coulter, Fort Collins, USA). A 488nm laser was used to excite both GFP and RFP, detected with 530/30 and 616/21 bandpass filters, respectively. 1uM DAPI (Sigma) was added to the sorted cells and incubated on ice for 5 min, to discriminate between nuclei containing cells and debris. DAPI was excited with a 405nm laser and detected with a 450/40nm bandpass filter.

Microspores

Microspores were purified by flowcytometric cell sorting using a Moflo High-Speed Cellsorter (Dako-Cytomation Inc. Fort Collins, CO, USA). The machine was used in a standard configuration, using a 100 μ ceramic nozzle, the 488 nm laser line from an Innova

C90 Argon Ion Laser at 150 mW (Coherent Inc. Santa Clara, CA, USA) for excitation, 580/20HQ and 670/40HQ nm Bandpass Filters and a 610 DCLP Dichroic Mirror (Chroma Techologies Corp. Brattleboro, VT, USA) for log-amplified detection. Microspores and other stages of pollen development were identified by their elevated high angle scatter (SSC) and autofluorescence properties (observed in the GFP channel). Within this population, microspores were selected by their characteristic smaller size, captured by a diminished low angle scatter (FSC) and time-of-flight (Pulse Width).

RT-PCR

Semi-quantitative RT-PCR was performed with total RNA isolated with Trizol from approximately 100.000 cells/nuclei. First-strand cDNA (Oligo-dT primed) was synthesized in 25ul reactions using the MLV reverse transcriptase - RNaseH minus (Promega) according to manufacturer instructions. 2ul of non-diluted cDNA was used as a template for 30x PCR reactions. Primers used are listed in Table S2.1.

5. REFERENCES

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol *133*, 713-725.

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

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

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009). A plant germline-specific integrator of sperm specification and cell cycle progression. PLoS Genet *5*, e1000430.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol *133*, 462-469.

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

Engel, M.L., Davis, R.H., and McCormick, S. (2005). Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol *138*, 2124-2133.

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

Gou, X.P., Xu, Y., Tang, L., Yan, F., and Chen, F. (2001). Representative cDNA library from isolated rice sperm cells. Acta Bot Sin *43*, 1093-1096.

Heslop-Harrison, J., and Heslop-Harrison, Y. (1970). Evaluation of pollen viability by enzymatically induced fluorescence; intracellular hydrolysis of fluorescein diacetate. Stain Technol *45*, 115-120.

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

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

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

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

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

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

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

Schoft, V.K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA

methylation upon decondensation of constitutive heterochromatin. EMBO Rep *10*, 1015-1021.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell *136*, 461-472.

Twell, D. (2011). Male gametogenesis and germline specification in flowering plants. Sex Plant Reprod *24*, 149-160.

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

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

Xu, H.L., Swoboda, I., Bhalla, P.L., Sijbers, A.M., Zhao, C.X., Ong, E.K., Hoeijmakers, J.H.J., and Singh, M.B. (1998). Plant homologue of human excision repair gene ERCC1 points to conservation of DNA repair mechanisms. Plant J *13*, 823-829.

CHAPTER 3

Comparative Transcriptomics of Arabidopsis Sperm Cells

Chapter section published as original article in peer-review journal:

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

1. INTRODUCTION

Arabidopsis sperm cells have highly condensed chromatin, but the original assumption that they are transcriptionally quiescent was proven wrong by the activity of *AtGEX1* promoter, whose expression is confined to sperm cells after pollen mitosis II, and was not detected in the progenitor generative cell (ENGEL *et al.*, 2005). A few genes had been described as expressed in Arabidopsis sperm cells; some of these appear to be important for pollen tube guidance (VON BESSER *et al.*, 2006) or for regulating their own transcriptional program (OKADA *et al.*, 2005). A cDNA library was constructed from FACS-purified sperm cells of *Zea mays*, which revealed a diverse complement of mRNAs representing at least 2,560 genes (ENGEL *et al.*, 2003). These findings were significantly more than the restricted number of transcripts previously described in plant sperm cells (OKADA *et al.*, 2005; SINGH *et al.*, 2002; XU *et al.*, 1999a, b). In addition they led to the identification of *AtGEX1* and *AtGEX2* (ENGEL *et al.*, 2005), whose promoters drive expression in Arabidopsis sperm cells. Additional sperm-specific genes have been reported (INGOUFF *et al.*, 2007; MORI *et al.*, 2006; VON BESSER *et al.*, 2006), including *DUO1*, a MYB transcription factor in Arabidopsis that is important for the G2/M transition during pollen mitosis II (ROTMAN *et al.*, 2005). Transcriptional profiling of generative cells of *Lilium longiflorum* (OKADA *et al.*, 2006; OKADA *et al.*, 2007) revealed a considerable overlap with the maize sperm ESTs (ENGEL *et al.*, 2003) and the Arabidopsis pollen microarray datasets (BECKER *et al.*, 2003; HONYS and TWELL, 2003). Despite this previous work, the male gamete ESTs from maize, Plumbago, and lily were limited, so that testing the potential importance of sperm-expressed transcripts by means of reverse genetics could not be comprehensive, even when an Arabidopsis homolog could be identified for such ESTs.

We used Fluorescence-Activated Cell Sorting (FACS) to isolate and purify sperm cells from transgenic Arabidopsis plants that were expressing enhanced green fluorescent protein under the control of a sperm-specific promoter *AtGEX2p::eGFP* (Chapter 2). Subsequently, Affymetrix ATH1 Genome Arrays were used to profile the mRNA complement of the male gametes. The direct comparison of their transcriptome with those of pollen and seedlings, as well as with additional ATH1 data sets from a variety of vegetative tissues, showed that the sperm cell transcriptome was distinct, and led to the identification of transcripts that were enriched and/or preferentially expressed in sperm cells. Functional classification of sperm-enriched transcripts showed that DNA repair, ubiquitin-mediated proteolysis and proteins required for progression through the cell cycle were over-represented categories, which correlated with the biological state and cell fate. Our analyses provide a number of hypotheses for testing and point out future challenges towards understanding the role of genes expressed in the male gametes before, during, and possibly after double fertilization.

2. RESULTS AND DISCUSSION

A remarkable diversity of transcripts

To determine the sperm cell transcriptome, we used Arabidopsis GeneChip ATH1 Genome arrays, representing 22,392 different genes. Although microarray data sets for seedlings and pollen exist (PINA *et al.*, 2005), we repeated analyses for these samples using identical starting amounts of total RNA (16 ng), since these amounts were less than those used in previous studies in our lab. Correlation coefficients above 0.97 between the three replicates of each sample type underline the reproducibility of the array data (Figure S3.1). Based on the MicroArray Suite (MAS) 5 detection algorithm, the mean percentages of detected Present calls were 27% for sperm cells, 33% for pollen and 64% for seedlings, corresponding to 5,829, 7,177 and 14,464 genes, respectively. Summarized expression data of all samples are shown in Table S3.1.

A cDNA library from FACS-purified sperm cells of *Zea mays* yielded around 5,000 sequenced ESTs that matched with ~1,385 annotated genes in the Arabidopsis genome (e-values of 1e-8 or smaller) (ENGEL *et al.*, 2003), although 138 of those were not represented on ATH1. A Venn diagram (Figure 3.1) was used to display the number of genes common between our samples and the identifiers of Arabidopsis genes matching the maize sperm cell ESTs (restricted to genes represented on the array). Although differences in the transcriptomes of sperm cells from monocot and dicot plants are expected, it was surprising that only 594 genes were shared between our dataset of Arabidopsis sperm cell-expressed genes and the Arabidopsis genes listed as corresponding to maize sperm ESTs. One explanation of this might be that the maize EST comparisons identified a different family member as the best Arabidopsis match. Another explanation is that cDNA sequencing could

have identified mRNAs that are not detected in the microarray analyses. To test these hypotheses, we analyzed 395 maize sperm ESTs for which their putative Arabidopsis homologs were called Absent by the microarray experiment. This analysis confirmed that most (238) were singletons, while the others were mostly represented by 2, 3 or 4 ESTs. These findings support the idea that rare messages might have been sequenced from the maize sperm library but would not have been detected by the microarray. A few genes were represented by multiple ESTs (e.g. ATPase encoding gene represented by 23 ESTs) and these belong to gene families, so that the gene selected as the best match could have been a different member than the one detected in Arabidopsis sperm cells. The overlap between Arabidopsis sperm cells and seedling encompassed 4,757 genes, representing almost the entire sperm transcriptome, with approximately 2,400 of those genes showing enriched expression in sperm. As expected, the vast majority of genes called Present in Arabidopsis sperm cells were also detected in Arabidopsis pollen (3,813).

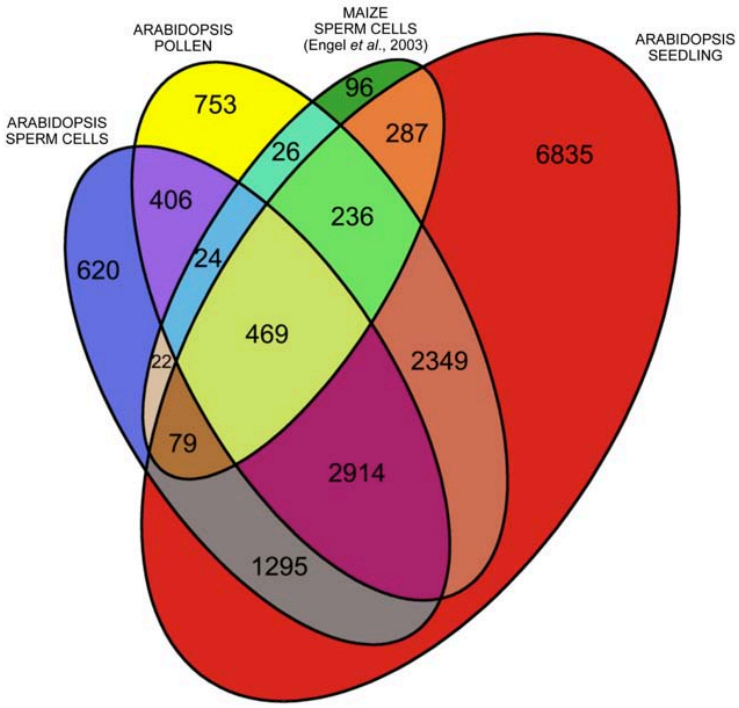


Figure 3.1 - Venn diagram, depicting the overlap between genes whose expression was called Present in sperm cells (5,829), pollen (7,177), seedlings (14,464), and additionally the Arabidopsis genes that are represented on the ATH1 array (1,239) and that match those of maize sperm cells ESTs (Engel et al., 2003). Intersection of common genes between Arabidopsis sperm cells and pollen, maize sperm cells, and Arabidopsis seedlings is 3,813, 594, and 4,757, respectively.

To address the possibility of contamination from RNA derived from the vegetative cell in our FACS-purified sperm cell samples, we screened the dataset for genes reported as pollen-specific (expression confined only to the vegetative cell). A good example can be found in Strompen *et al*. (STROMPEN *et al.*, 2005), where isoforms of the Vacuolar H^+-ATPase subunit E were analyzed by means of comparison of VHA-E1-GFP, E2-GFP and E3-GFP fusions expression in mature pollen. E1, the major isoform, was expressed in sperm cells but not in the vegetative cell, E2 was expressed only in the vegetative cell, and E3 was expressed in both the vegetative cell and sperm cells. The genes encoding all three isoforms are represented on ATH1 (*At4g11150*, *At3g08560* and *At1g64200*, respectively). Both E1 and E3 transcripts were detected in sperm cells, whereas E2 was not. All three transcripts were detected in pollen. Another example is *AtVEX1|At5g62850*, specifically expressed in the vegetative cell cytoplasm and nucleus (ENGEL *et al.*, 2005), and called Absent in the sperm cell sample, but Present in pollen. For further validation of our microarray data, RT-PCR was performed. One chosen target was one of the highest expressed genes in pollen, encoding a carbonic anhydrase protein (*At5g04180*), and whose expression was not detected in sperm cells. In agreement with our microarray data, this transcript could not be detected in sperm cells or in seedling by RT-PCR, but high levels were detected in pollen (Figure 3.2). As a further proof, two additional independent samples of total RNA from sperm cells were used to confirm enrichment for a sperm-specific transcript *AtMGH3|At1g19890* (OKADA *et al.*, 2005),

and absence of transcripts derived from the vegetative cell, *AtVEX1*|*At5g62850* (ENGEL *et al.*, 2005), by RT-PCR (Figure 3.2C).

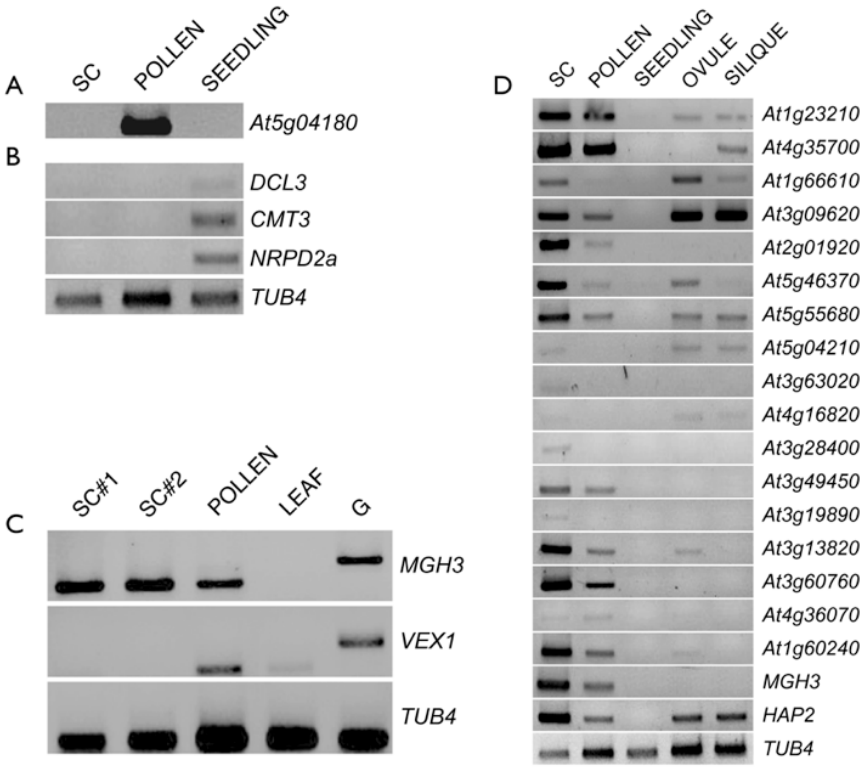


Figure 3.2 - RT-PCR analysis. (A and B) Gel figures presenting confirmatory RT-PCR analysis for a gene highly expressed in pollen but not detected in sperm cells in our microarray analysis, encoding a carbonic anhydrase family protein (*At5g44340*; A); and genes involved in the sRNAs and RdDM pathways, which are not detectable in sperm cells, DCL3 (*At3g43920*), *CMT3* (At1g69770), and NRPD2a (*At3g23780*; B). **(C)** RT-PCR on total RNA from two samples of FACS-isolated sperm cells (SC#1, SC#2), pollen, and leaf, showing enrichment of *MGH3* (*At1g19890*) transcripts and absence of *VEX1* (*At5g62850*) transcripts in both sperm cell replicates. **(D)** Expression of several genes presented in Table II was tested by RT-PCR in sperm cells, pollen, seedling, ovule, and silique cDNA samples. *TUB4* (*At5g04180*) was used as positive control. SC - sperm cells; G - genomic DNA.

Unique transcriptional profile to control DNA repair, ubiquitination and cell cycle

We carried out comparative analyses of the sperm cell transcriptome with transcriptional datasets from representative vegetative tissues of Arabidopsis. Principal component analysis (PCA) is a robust method to project high-dimensional, global expression data onto the three principal components. The closer two points are in the plot, the more similar the samples are in terms of their global gene expression profile. As shown in Figure 3.3B, the first principal component separates sperm cells and pollen from the vegetative tissues, while the second principal component shows a further separation between pollen and sperm cells. The first principal component demonstrates the uniqueness of the expression profile in sperm cells. This was already demonstrated for pollen in previous studies (PINA *et al.*, 2005), but here we can additionally observe that pollen and sperm cells separate along the second principal component, thus emphasizing the differences in their transcriptional profiles. A similar separation of samples can be obtained with hierarchical clustering (Figure 3.3A).

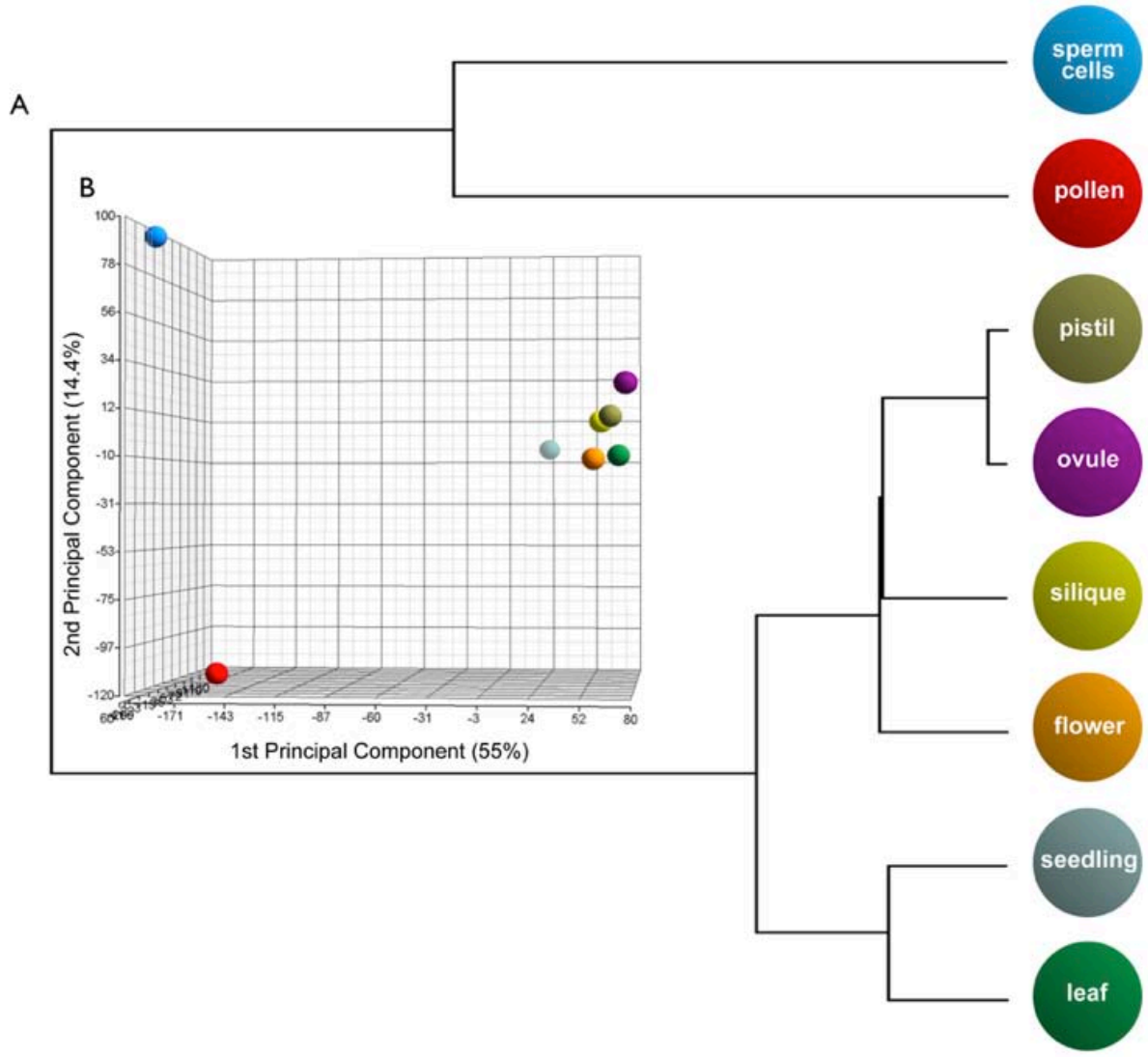


Figure 3.3 - Hierarchical clustering and Principal Compontent Analysis of tissue dependent gene expression patterns. Transcriptome data of seedlings, pollen, and sperm cells were compared with those of leaf, flower, silique, ovule, and unpollinated pistil (BOAVIDA *et al.*, 2011; PINA *et al.*, 2005). **(A)** Hierarchical clustering dendrogram, using Pearson's dissimilarity to calculate row dissimilarity and Ward's method for row clustering. **(B)** Principal Component Analysis (PCA) showing that the transcriptomes of pollen and sperm cells are distinct, noted by a clear separation from sporophytic tissues along the first principal component.

The DAVID Gene Functional Classification Tool (HUANG *et al.*, 2007) was used to condense the list of genes showing enriched expression in sperm cells. This tool agglomerates the genes into functional clusters according to their biological significance and enrichment score, in relation to all genes represented on the array (Table 3.1 and Table S3.2). The distribution through the first clusters highlight over-represented GO categories, i.e. DNA metabolism (DNA replication and repair), ubiquitin-mediated proteolysis (ubiquitin ligase activity) and cell cycle. Although most of the genes expressed in sperm cells were also detected in pollen, it was previously shown that the over-represented GO terms for genes with enriched expression in pollen are signaling, vesicle trafficking and membrane transport (PINA *et al.*, 2005). Previous studies on cell cycle activity of sperm cells (FRIEDMAN, 1999) showed that sperm spend most of their development (from the generative cell undergoing mitosis until entry into the embryo sac) in the S phase of the cell cycle. Plant sperm also have a complex set of transcripts for control of protein fate and degradation through ubiquitination. Interestingly, these processes are also essential during spermatogenesis of mammals, for histone-to-protamine replacement (BAARENDS *et al.*, 1999), as well as for many other key events in gametogenesis and fertilization (reviewed in SAKAI *et al.*, 2004). In concordance with our results, genes involved in DNA repair, activation of cyclins and ubiquitination were also highly represented in ESTs from generative cells of *Lilium longiflorum* (OKADA *et al.*, 2006; SINGH *et al.*, 2002; XU *et al.*, 1998).

Table 3.1 - Functional classification of sperm-enriched genes. Gene ontology charts (biological process, molecular function, and cellular component) for sperm-enriched genes were functionally clustered by DAVID tools according to their enrichment score and against all genes represented on the ATH1 array. The representative annotations terms associated to each cluster were manually selected.

Cluster	Annotation Term	Enrichment Score
1	DNA metabolism/DNA repair	8.34
2	Ubiquitin-mediated proteolysis	6.51
3	Cell cycle	6.42
4	Mitosis	3.94
5	Biopolymer metabolism	3.45
6	Chromosome organization and biogenesis	3.23
7	Heterotrimeric G-protein complex	3.18
8	CDK regulator activity	2.43
9	ATPase activity	2.3
10	Cysteine-type peptidase activity	2.25
11	mRNA metabolism	2.17
12	Cytoskeleton	2.08
13	Intracellular transport	2.06
14	Unfolded protein binding	1.98
15	Damaged DNA binding	1.91

Identification of genes preferentially expressed in the germline

Genes showing preferential or enriched expression in sperm cells are primary candidates for roles in male gamete development and fertilization. To identify sperm-preferentially expressed transcripts we performed a comparative analysis with a number of other microarray datasets: leaves and siliques (PINA *et al.*, 2005), ovules and unpollinated pistils (BOAVIDA *et al.*, 2011) and AtGenExpress datasets (SCHMID *et al.*, 2005) available through the web-based application Genevestigator (ZIMMERMANN *et al.*, 2004), selecting samples that did not contain pollen. Seventy-four genes were identified (Table 3.2) and distributed according to their molecular function (as annotated in TAIR as of May 2007).

Notably, this set includes two genes previously reported as sperm-specific: histone H3, *AtMGH3|At1g19890* (OKADA *et al.*, 2005) and *HAP2(GCS1)|At4g11720* (VON BESSER *et al.*, 2006). We performed RT-PCR analysis for several genes within this list, including *AtMGH3* and *HAP2*, in order to confirm the microarray data reported in this study (sperm cells, pollen and seedling), and additionally with ovule and silique samples (Figure 3.2D). Although we could not amplify any of the transcripts from the seedling sample, we detected expression for several of these genes in both ovule and silique samples. We can exclude contamination of genomic DNA in the cDNA templates as source of these amplification products, because we used primers amplifying intron-spanning products whenever possible. Surprisingly, one gene whose expression was detected in ovules and siliques is *HAP2(GCS1)|At4g11720*. Previous studies in Arabidopsis showed only a male-specific role for HAP2 (JOHNSON *et al.*, 2004; MORI *et al.*, 2006; VON BESSER *et al.*, 2006), but on the other hand, RT-PCR analysis indicated that a *HAP2* homolog in *Chlamydomonas* was expressed in both *plus* and *minus* gametes (MORI *et al.*, 2006). Interestingly, although *HAP2* transcripts were detected in *Chlamydomonas plus* gametes, a recent study has shown that HAP2 protein is essential for fusion in *minus* gametes only (LIU *et al.*, 2008). We do not know if the genes we detected are transcribed in the whole ovule or only in a fraction of the cells comprising the ovule, e.g. the egg cell and/or central cell. It was demonstrated that *GEX2*, previously described as specifically expressed in sperm cells (ENGEL *et al*., 2005), is expressed also within the female gametophyte (ALANDETE-SAEZ *et al.*, 2008).

Table 3.2 - Genes preferentially expressed in sperm cells.

*, Genes analyzed by RT-PCR whose expression has been detected only in sperm cells and in some cases in pollen, but probably being sperm-derived. Shown are 74 genes appearing to be preferentially expressed in sperm cells based on comparisons with microarray data of several vegetative tissues (PINA *et al*., 2005; BOAVIDA *et al*., 2011) and the AtGenExpress database (SCHMID *et al*., 2005), distributed according to their molecular function as annotated in TAIR. AGI, Arabidopsis Genome Initiative.

Probe Set ID	AGI ID	Description	Signal
DNA or RNA Binding			
255815_at	*At1g19890**	ATMGH3/MGH3 (MALE-GAMETE-SPECIFIC HISTONE H3) (OKADA *et al.*, 2005)	11305
256313_s_at	*At1g35850* | *At5g59280*	APUM17 (ARABIDOPSIS PUMILIO 17); APUM16 (ARABIDOPSIS PUMILIO 16)	335
245720_at	*At5g04210*	RNA recognition motif (RRM)-containing protein	129
251572_at	*At3g58390*	Eukaryotic release factor 1 family protein / eRF1 family protein	69
266933_at	*At2g07760*	Zinc knuckle (CCHC-type) family protein	50
253192_at	*At4g35370*	Transducin family protein / WD-40 repeat family protein	647
247016_at	*At5g66970*	GTP binding	54.5
255124_at	*At4g08560*	APUM15 (ARABIDOPSIS PUMILIO 15)	60
Hydrolase Activity			
264896_at	*At1g23210*	Glycosyl hydrolase family 9 protein	10415
261217_at	*At1g32850*	Ubiquitin carboxyl-terminal hydrolase family protein	1865
257378_s_at	*At2g02290* | *At5g23470*	NLI interacting factor (NIF) family protein	389
258725_at	*At3g09620*	DEAD/DEAH box helicase, putative	210
245447_at	*At4g16820*	Lipase class 3 family protein	157
262042_at	*At1g80140*	Glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein	69
261278_at	*At1g05800*	Lipase class 3 family protein	46
258740_at	*At3g05780*	Lon protease, putative	661
Kinase Activity			
263577_at	*At2g17090*	SHORT SUSPENSOR (SSP) (BAYER *et al*., 2009)	3470
255892_at	*At1g17910*	Wall-associated kinase, putative	117
254009_at	*At4g26390*	Pyruvate kinase, putative	1179
253128_at	*At4g36070*	CPK18 (calcium-dependent protein kinase 18)	48
Other Binding			
263304_at	*At2g01920**	Epsin N-terminal homology (ENTH) domain-containing protein / clathrin assembly protein-related	1579
257372_at	*At2g43220*	DC1 domain-containing protein	1300
247737_at	*At5g59200*	Pentatricopeptide (PPR) repeat-containing protein	1212
246388_at	*At1g77405*	Unknown protein	208
267047_at	*At2g34370*	Pentatricopeptide (PPR) repeat-containing protein	122
248898_at	*At5g46370*	KCO2 (CA2+ ACTIVATED OUTWARD RECTIFYING K+ CHANNEL 2)	914
Other Enzyme Activity			
263217_at	*At1g30740*	FAD-binding domain-containing protein	1104
263041_at	*At1g23320*	Alliinase family protein	103
256760_at	*At3g25650*	ASK15 (ARABIDOPSIS SKP1-LIKE 15)	34
Protein Binding			
256408_at	*At1g66610*	Seven in absentia (SINA) protein, putative	76
263301_x_at	*At2g04970* | *At2g06440* | *At2g14140* | *At2g15200*	Heat shock protein (At2g04970, At2g14140); Unknown protein (At2g06440, At2g15200)	56
Transcription Factor Activity			
253153_at	*At4g35700*	Zinc finger (C2H2 type) family protein	16885
264269_at	*At1g60240*	Apical meristem formation protein-related	11364
263417_at	*At2g17180*	Zinc finger (C2H2 type) family protein	567
255514_s_at	*At5g65330*	MADS-box family protein	85
Transferase Activity			
251794_at	*At3g55590*	GDP-mannose pyrophosphorylase, putative	51
Transporter Activity			
260850_at	*At1g21870*	Glucose-6-phosphate/phosphate translocator-related	705
Unknown			
251384_at	*At3g60760**	Unknown protein	4936
249468_at	*At5g39650*	Unknown protein	4326
252258_at	*At3g49450**	F-box family protein	3779

Probe Set ID	AGI ID	Description	Signal
248072_at	*At5g55680*	Glycine-rich protein	1903
249981_at	*At5g18510*	Unknown protein	1830
256047_at	*At1g07060*	Unknown protein	1578
257603_at	*At3g13820*	F-box family protein	1400
247341_at	*At5g63720*	Unknown protein; similar to IMP dehydrogenase/GMP reductase [Medicago truncatula]	962
257443_at	*At2g22050*	Unknown; similar to Cyclin-like F-box; Serine/threonine protein phosphatase, BSU1 [Medicago truncatula]	873
254183_at	*At4g23960*	F-box family protein	730
254883_at	*At4g11720*	GCS1/HAP2 (GENERATIVE CELL-SPECIFIC 1) (VON BESSER *et al.*, 2006)	662
248196_at	*At5g54150*	Unknown protein	658
251201_at	*At3g63020**	Unknown protein	443
258431_at	*At3g16580*	F-box family protein	290
257993_at	*At3g19890**	F-box family protein	255
257980_at	*At3g20760*	Unknown protein	225
267049_at	*At2g34210*	KOW domain-containing transcription factor family protein	1179
248214_at	*At5g53670*	Unknown protein	174
257585_at	*At3g12420*	pseudogene of 3'-5' exonuclease containing protein	163
267399_at	*At2g44195*	Unknown protein	151
254184_at	*At4g23970*	Unknown protein	131
246638_at	*At5g34880*	gypsy-like retrotransposon family	107
245135_at	*At2g45230*	non-LTR retrotransposon family (LINE)	87
258319_at	*At3g22700*	F-box family protein	83
255399_at	*At4g03750*	Unknown protein	82
265606_s_at	*At3g09510* | *At2g31520* | *At2g25550*	Unknown protein	70
249257_at	*At5g41640*	Unknown protein	67
254325_at	*At4g22650*	Unknown protein	61
265838_at	*At2g14550*	pseudogene	61
259465_at	*At1g19030*	Unknown protein	60
247728_at	*At5g59510*	DVL18/RTFL5 (ROTUNDIFOLIA LIKE 5)	51
257998_at	*At3g27510*	Unknown protein	50
264732_at	*At1g62160*	Pseudogene	49
259634_at	*At1g56380*	Mitochondrial transcription termination factor family protein / mTERF family protein	47
262472_at	*At1g50160*	Unknown protein	46
257843_at	*At3g28400**	hAT-like transposase family (hobo/Ac/Tam3)	34
245754_at	*At1g35183*	Unknown protein	26

Even though some of the genes reported in Table 3.2 were detected in ovules and/ or siliques, most of them show enriched expression in sperm cells. A highly expressed gene within this list encodes a protein that belongs to the glycosyl hydrolase family (*At1g23210*). Interestingly, in mammals a glycosyl hydrolase (hyaluronidase) protein, PH-20, located at both the plasma and acrosomal membranes in spermatozoa, is involved in adhesion to the zona pellucida (LATHROP *et al.*, 1990). Another gene of note is *KCO2|At5g46370*, encoding a member of the Arabidopsis K^+ channel family AtTPK(KCO) (VOELKER *et al.*, 2006). Within the same family, *AtTPK4(KCO4)* is predominantly expressed in pollen tubes, and is spontaneously activated and modulated by external Ca^{2+} and cytosolic pH (BECKER *et al.*, 2004). In mammalian sperm, capacitation refers to intracellular alkalinization and changes in

membrane potential, conferring on sperm the capacity to reach and fertilize the egg. Such mechanisms were shown to be driven by outwardly rectifying K^+ currents (NAVARRO *et al.*, 2007), whereby a inward Ca^{2+}-selective current is maximized by $I_{CatSper}$ (KIRICHOK *et al.*, 2006). We suggest that K^+ homeostasis in Arabidopsis sperm cells might be at least partially controlled by AtTPK2(KCO2), and that plant male gametes might also undergo a process similar to spermatozoa capacitation. Other channel proteins might play important roles in sperm cells, such as maintaining turgor; for example, 2 members (MSL2|*At5g10490* and MSL3|*At1g58200*) of the mechanosensitive ion channel family (HASWELL and MEYEROWITZ, 2006) were detected in sperm cells, and both are expressed more highly in sperm than in pollen. The predominant phenotype of mutations in these genes was in chloroplast size and shape (HASWELL and MEYEROWITZ, 2006), but some fertility problems were reported as well.

Two genes (*At3g49450*, *At3g19890*), which according to our RT-PCR analyses are expressed only in sperm cells, encode proteins with an F-Box domain in their N-terminus, and it is notable that the most highly expressed gene in Arabidopsis sperm cells is annotated as an F-box protein (*At3g62230*). We used MAPMAN (THIMM *et al.*, 2004) to display the expression profiles of genes involved in the ubiquitin/26S proteasome pathway. Among the F-Box proteins represented on the array and expressed in both seedlings and sperm cells, most had higher expression values in sperm than in seedlings (Figure 3.4B). Only about 20 of the more than 450 F-Box proteins have had their biological functions elucidated: they are involved in the regulation of diverse cellular processes including cell cycle transitions, transcription, signal transduction, circadian rhythms, floral development (DEL POZO and ESTELLE, 2000; PATTON *et al.*, 1998; SCHWAGER *et al.*, 2007; XIAO and JANG, 2000) and gametophytic self-incompatibility (HUA *et al.*, 2007; SIJACIC *et al.*, 2004). The C-termini of F-Box proteins usually contain domains conferring substrate-binding specificity, being later

targeted for degradation by the ubiquitin-ligase complex SCF (SKP-Cullin-F-Box). Why this diversity and enrichment of F-Box proteins in sperm cells? One possible hypothesis is that proteins needed during earlier stages of microgametogenesis might need to be targeted for degradation in later stages, e.g. cell cycle regulators. Recent studies support this idea, since the activity of UBP3/UBP4 deubiquitinating enzymes (essential for the Ub/26S proteasome system) is crucial for PMII in Arabidopsis (DOELLING *et al.*, 2007), and the F-Box *FBL17* was also shown to be important for degradation of the CDK inhibitors KRP6 and KRP7 in the germline, allowing cell cycle progression and sperm formation (GUSTI *et al.*, 2009; KIM *et al.*, 2008). On the other hand, these and other transcripts or proteins stored in sperm cells might be delivered to both egg and central cells upon fertilization and play a role in early embryonic development.

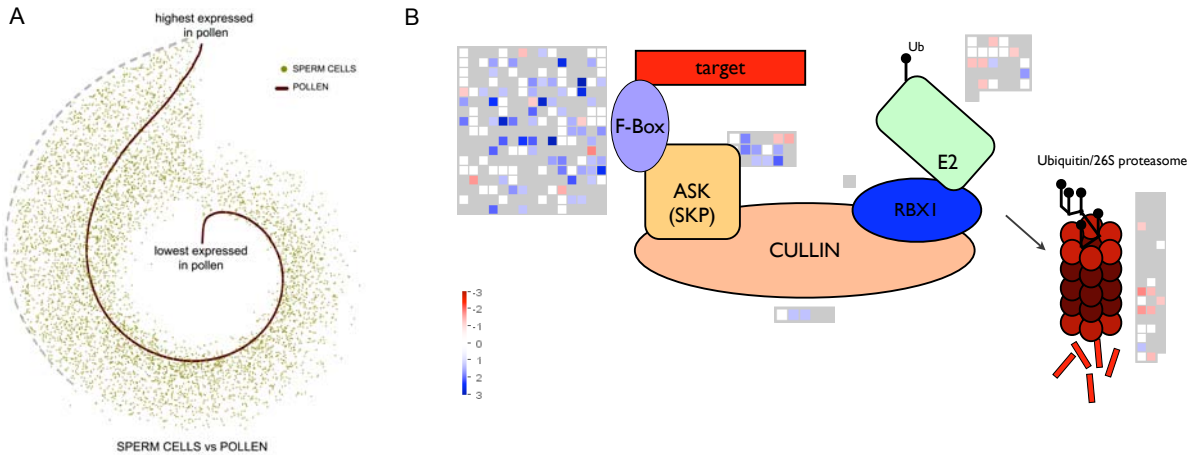


Figure 3.4 – **(A)** Snail View representation of gene expression in sperm cells and pollen. The expression levels of 5,829 expressed genes in Arabidopsis sperm cells (dots) were co-plotted with the expression levels in pollen (line). Gene expression values are represented in angular coordinates, in which the angle defines the gene rank (clockwise from middle) and the radius defines the logarithm of gene expression values. Genes were ranked according to increasing expression in pollen. A clear cut line of dots on the left side of the Snail View (dashed line) indicates that highly-expressed sperm cell transcripts contribute to the corresponding expression level of that given gene in pollen. **(B)** Overview of the SCF E3 complex. Fold-

change (comparing the logarithmic values of gene expression in sperm cells relative to seedlings) is represented using an adapted version of the 26S/ubiquitin-dependent proteasome pathway for the MAPMAN tool (THIMM *et al.*, 2004). Colored squares symbolize genes represented on ATH1 that are associated with each component of the SCF E3 complex. Blue: up-regulated in sperm; Red: down-regulated in sperm; White: no change; Gray: Absent in sperm.

Transcription factors

The *AtGEX1* and *AtGEX2* promoters, as well as other sperm-expressed promoters discussed in (ENGEL *et al.*, 2005) have binding sites for Dof-type transcription factors. Of the 32 Dof-type transcription factors in Arabidopsis, three (*At3g47500*, *At5g39660* and *At5g62430*) were detected in sperm cells; one or more of these probably regulates sperm-specific promoters. There are several MYB-type transcription factors with relatively high expression in sperm cells, while other transcription factor families have restricted representation. For example, there are many homeobox transcription factors in Arabidopsis, but only one of this group (*At3g61150*) has substantial expression in sperm cells. As a group, genes encoding scarecrow transcription factors are mostly not expressed in pollen or in sperm cells, but 3 (*At2g45160*, *At2g29060*, and *At1g63100*) are detected in sperm cells and not in pollen, implying that these 3 might play sperm-specific roles. The bHLH transcription factor family is one of the largest in Arabidopsis (TOLEDO-ORTIZ *et al.*, 2003), but only a few were detected in sperm cells, and one of these (*At2g42300*) is not detected in pollen, again suggesting that it may play a specific role in male gametes. Notably, the zinc finger (C2H2 type) transcription factor *At4g35700* shows the highest expression level of all sperm-preferentially expressed genes identified in this study. This transcription factor might be expressed only in sperm cells although it was also detected in siliques by RT-PCR. Given that

the detection level for this transcript is very high in pollen, and higher in sperm cells than in pollen, it is likely that its weak detection in siliques is pollen-derived.

Signal transduction pathways

Sperm cells might have unique signalling pathways. For example, the small GTPases of the Rop family are important for many aspects of cytoplasmic signalling, including actin cytoskeleton reorganization (NIBAU *et al.*, 2006). The GTP/GDP status of Rops is controlled by the Rop-GEF family, as well as by RopGAPs and ROP-GDIs (KLAHRE *et al.*, 2006; KLAHRE and KOST, 2006; ZHANG and MCCORMICK, 2007). There are 3 Rops expressed in sperm cells (*At2g17800*, *At3g51300* and *At4g35950*), as well as one of the 3 RopGDIs in Arabidopsis (*At3g07880*), and a RopGAP (*At5g61530*). There are 14 genes encoding PRONE-type RopGEFs in Arabidopsis, and 5 of these are pollen-specific or selective (ZHANG and MCCORMICK, 2007). None of the 14 RopGEFs were detected in sperm cells. Another protein, Spike1 (*At4g16340*) also has RopGEF activity (BASU *et al.*, 2005), but its transcript was not detected in pollen or in sperm cells. It is unclear how the GTP/GDP status of Rops in sperm cells can be controlled if RopGEFs are absent. We should not exclude the hypothesis that some known RopGEFs may be in fact expressed in sperm cells, but their transcripts are below the detection threshold.

In sperm cells, some complicated signalling cascades might be simplified. For example, MAP kinase cascades are central to many signaling pathways in plants, and there is often crosstalk between different members in different signaling pathways (MISHRA *et al.*, 2006). Of the 20 MAPKKKs in Arabidopsis, several were detected in sperm cells, and two (*MAPKKK19|At5g67080* and *MAPKKK20|At3g50310*) have extremely high expression; these two were not detected in pollen, implying that sperm-specific signaling pathways might exist. Indeed, these MAPKKKs must act on MAPKK3 (*At5g40440*), the only one of the 10

MAPKKs in Arabidopsis that could be detected in sperm cells. Several MAPKs are also expressed, with *At2g01450* showing the highest expression level. Other examples of gene families with restricted expression in sperm cells include the large lectin receptor kinase family, only a few of which were detected, and the auxin efflux carriers, only one (*At5g01990*) of which is expressed in sperm, while a different one (*At2g17500*) is expressed in pollen.

Small RNA and DNA methylation

From analysis of maize sperm ESTs and their respective Arabidopsis homologs (ENGEL *et al.*, 2003), we had previously hypothesized that small RNA pathways might be partially activated in Arabidopsis sperm cells (PINA *et al.*, 2005), which stood in contrast to an apparent complete lack of expression in mature pollen grains, i.e. Absent calls for all genes in these pathways (PINA *et al.*, 2005). However, our new pollen data set shows Present calls for three of the 15 transcripts that were below the detection threshold in the previous study (*AGO1|At1g48410*, *AGO4|At2g27040* and *AGO9|At5g21150*) (Table S3.1). Given that the biological material was obtained in the same way in the two studies and that the expression values for the three AGO transcripts are relatively low, our best explanation for this discrepancy is that improved chemistry for sample processing, array hybridization and staining resulted in a better signal to noise ratio and thus a higher sensitivity.

The list of genes known or thought to be involved in small RNA pathways has increased significantly compared with previous data sets (PINA *et al.*, 2005). When we include genes implicated in RNA-directed DNA methylation (RdDM), maintenance of DNA-methylation and active demethylation (reviewed in MATZKE *et al.*, 2007; VAZQUEZ, 2006), 18 of the 53 genes are expressed in sperm cells (Table S3.1). Five of these transcripts

(*AGO9*|*At5g21150*, *DDM1*|*At5g66750*, *DRB4*|*At3g62800*, *MET1*|*At5g49160* and *SUVH5*|*At2g35160*) are so highly enriched in sperm cells that their detection in pollen is very likely to be sperm cell-derived. Besides the cytosine methyltransferase (MTase) *MET1*, a number of other transcripts expressed in sperm cells are implicated in maintenance of DNA-methylation: MET1 works in conjunction with HDA6|*At5g63110* to maintain CG methylation during DNA replication. Notably, the SWI2/SNF2 chromatin-remodeling factor DDM1, which is also involved in maintenance of CG methylation, is highly enriched in sperm cells, The histone-H3-lysine-9 MTases SUVH5, SUVH4|*At5g13960* and SUVH6|*At2g22740* are thought to maintain non-CG methylation via control of the DNA MTase CMT3|*At1g69770* (EBBS and BENDER, 2006). Interestingly *CMT3* expression could not be detected in sperm cells as we confirmed by RT-PCR (Figure 3.2B), but the MTase *DRM2* was. This MTase is also involved in the maintenance of non-CG methylation, but more importantly in RdDM by catalyzing de novo methylation of cytosines in all sequence contexts, in conjunction with DRD1 (not represented on the ATH1 array), the Polymerase IVb complex and AGO4 or AGO6. While the largest subunit of Pol IVb (*NRPD1b*|*At2g40030*) is expressed in sperm cells, expression of its second largest subunit *NRPD2a*|*At3g23780* could not be detected (Figure 3.2B). *AGO6* is expressed, but *AGO9*, a member of the AGO4/AGO6 subfamily (ZHENG *et al.*, 2007), shows much higher expression levels, indicating a possible role for AGO9 in the context of de novo methylation in sperm cells. siRNAs serve as triggers for de novo methylation, and their genesis involves RDR2|*At4g11130*, HEN1|*At4g20910* and DCL3|*At3g43920*, with *DCL3* apparently not expressed in sperm cells (Figure 3.2B). Instead, *DCL1*|*At1g01040* expression was detected, as well as very high expression levels for the *AGO1*-homolog *AGO5*|*At2g27880.* Interestingly, the expression of *OsMEL1*, the rice ortholog of *AGO5*, is germ-cell specific and indispensable for premeiotic mitosis and meiosis during sporogenesis (NONOMURA *et al.*, 2007). Taken together these data suggest that sperm cells in

mature pollen grains are actively regulating the epigenetic state of their genome through RdDM and maintenance of DNA methylation. This is in concordance with sperm cells of Arabidopsis being in S-phase at anthesis (DURBARRY *et al.*, 2005; FRIEDMAN, 1999). Perhaps these pathways are necessary to subdue selfish DNA elements after DNA replication. The high expression levels of AGO5 and AGO9 as well as the apparent absence of CMT3 and DCL3 transcripts suggest that novel small RNA pathways might act in sperm cells. The enriched expression of the dsRNA-binding protein *DRB4|At3g62800*, which functions in the trans-acting siRNA pathway (NAKAZAWA *et al.*, 2007) adds a further interesting piece to this puzzle, as does the enriched expression of Morpheus' molecule 1 (*MOM1|At1g08060*), which is involved in a DNA-methylation-independent epigenetic silencing pathway (MITTELSTEN SCHEID *et al.*, 2002; VAILLANT *et al.*, 2006). Notably most of the transcripts mentioned here can be found in clusters one and six of our functional classification (Table 3.1), highlighting the importance of DNA repair and epigenetic processes in sperm cells.

3. CONCLUSIONS

Here we reported the first whole-genome transcriptional analysis of plant sperm. Because so little is known about sperm cell biology, we could raise a number of questions by bio-informatics inference which eventually will open the way to test the respective underlying hypotheses by direct experimental methods. Furthermore, we identified a number of biological processes associated with sperm and identified several targets for reverse genetics analyses as these proteins are likely to play roles during sperm cell maturation and in double fertilization.

The mechanisms controlling eukaryotic germline differentiation rely mostly on transcriptional activation of germline genes while repressing expression of somatic genes. Transcriptional activation in the Arabidopsis germline is known to depend largely on the transcription factor DUO1, but our results provide additional candidate genes encoding transcription factors preferentially expressed in the gametes. The fact that small RNA silencing and particular DNA methylation pathways are enriched in sperm was expected, considering their prominent role in reprogramming the epigenome during gametogenesis. However, as in Arabidopsis most gene families within these pathways suffered large expansion by gene duplication, is important to know exactly which members are expressed in each cell type. This is especially true for the Argonaute protein family, which is composed of 10 members in Arabidopsis, but only AGO5 and AGO9 seem to be enriched in the sperm cells, while all the others are transcriptionally depleted. The supposed absence of AGO1 and AGO4, key proteins involved in the miRNA- and siRNA-directed gene silencing in the sporophyte, respectively, suggest that novel RNAi pathways may be established during pollen development.

4. MATERIALS AND METHODS

Plant material and growth conditions

Twelve 4 week-old seedlings of Arabidopsis ecotype Col-0 were grown on solid medium (B-5 with 0.8% phytoagar; Duchefa, Haarlem, The Netherlands) and pooled for RNA extraction. Pollen isolation and FACS were performed as previously reported (BECKER *et al.*, 2003). Sperm cells were isolated as described in Chapter 2. All sample types were processed in triplicate from three individual RNA extractions.

RNA extraction, biotin-labelling of cRNA and hybridization

Total RNA was isolated using the RNeasy Micro Kit (Qiagen, Germany), from three biological replicates of sperm cells, pollen and seedlings. RNA integrity was assessed using an Agilent 2100 Bioanalyser with a RNA 6000 Nano Assay (Agilent Technologies, Palo Alto, CA), and processed for use on Affymetrix Arabidopsis ATH1 Genome Arrays (Santa Clara, CA, USA) according to the manufacturer's Two-Cycle Target Labeling Assay. Briefly, 16 ng of total RNA containing spiked-in Poly-A RNA controls (GeneChip Expression GeneChip Eukaryotic Poly-A RNA Control Kit; Affymetrix) was used in a reverse transcription reaction (Two-Cycle DNA synthesis kit; Affymetrix) to generate first-strand cDNA. After second-strand synthesis, double-stranded cDNA was used in an *in vitro* transcription (IVT) reaction to generate cRNA (MEGAscript T7 kit; Ambion, Austin, TX). The cRNA obtained was used for a second round of cDNA and cRNA synthesis, resulting in biotinylated cRNA (GeneChip Expression 3'-Amplification Reagents for IVT-Labeling; Affymetrix). The size distribution of the cRNA and fragmented cRNA, respectively, was assessed using an Agilent 2100 Bioanalyzer with a RNA 6000 Nano Assay. Fifteen micrograms of fragmented cRNA was used in a 300-µL hybridization containing added hybridization controls. 200 µL of the

mixture was hybridized on arrays for 16 h at 45°C. Standard post-hybridization washes and double-stain protocols (FS450_0004) were used on an Affymetrix GeneChip Fluidics Station 450, in conjunction with the GeneChip Hybridization Wash and Stain kit (Affymetrix). Arrays were scanned on an Affymetrix GeneChip scanner 3000 7G. All quality parameters for the arrays were confirmed to be in the recommended range. GeneChip datasets for the 9 arrays used in this study are available in a MIAME-compliant format through ArrayExpress (accession no. E-ATMX-35).

Data analysis and gene functional classification

Absent/Present calls were generated from scanned arrays using Affymetrix GCOS 1.4 software and all subsequent analyses were conducted using dChip software as of April 2007 (http://www.dchip.org, Wong Lab, Harvard). For each sample type (sperm cells, pollen and seedling), a microarray experiment was performed with three biological replicates. We used a samplewise normalization to the median probe cell intensity (CEL) of all 9 nine arrays, and for each sample, the median CEL intensity of one replicate was scaled to the median CEL intensity of all arrays and defined as baseline. The remaining replicates of each sample were normalized to the baseline applying an Invariant Set Normalization Method (LI and WONG, 2001). Model-based gene expression was obtained from normalized CEL intensities based on a Perfect Match-only model (LI and HUNG WONG, 2001). Only genes called Present in at least two replicates of each sample type were considered Present. For downstream analysis we used only 16,385 genes, which were filtered according to their consistent expression levels within replicates (called Present within a variation of $0 < \text{median (SD/mean)} < 0.5$) and if called Present at least in one of the 9 arrays. Annotations for all the genes represented on the Arabidopis ATH1 genome array were obtained from the NetAffx database (www.affymetrix.com) as of June 2007 and manually updated according to TAIR7 as of May

2007.

Hierarchical clustering and Principal Component Analysis (PCA) were computed using Partek Genomics Suite (Partek, St. Louis, Missourii, USA). We performed comparative analyses of our samples with datasets of leaves, flowers and siliques (PINA *et al.*, 2005), and with datasets from ovules and unpollinated pistils (BOAVIDA *et al.*, 2011). The graphical representation SnailView (BECKER *et al.*, 2003) was applied to visualize the mean pattern of expression for the 5,829 genes detected in sperm cells, co-plotted with correspondent pollen data (Figure 3.4A).

In order to identify genes for which expression has been only detected in sperm cells, we performed comparisons with microarray data of vegetative tissues from previous studies (PINA *et al.*, 2005) and using the web-based application Genevestigator (ZIMMERMANN *et al.*, 2004). First, we selected genes called Present in sperm cells but Absent in pollen and vegetative tissues. It was also necessary to set a fold-change cut-off for genes enriched in sperm cells (relative to pollen) to infer which genes detected in pollen might be sperm-selective. We found that increasing the fold-change cut-off between the logarithmic expression values of pollen and sperm cells yielded a better correlation and for that reason we used a fold change of 3, without any direct statistical significance. A list of probe IDs representing 254 genes was uploaded on Genevestigator software (ZIMMERMANN *et al.*, 2004), selecting for analysis only high quality arrays of the AtGenExpress Database (SCHMID *et al.*, 2005) from experiments of developmental stages of *Arabidopsis thaliana*. Samples containing mature pollen or pollen undergoing microgametogenesis were excluded. To compare the expression of genes across the selected arrays we used the Digital Northern tool, and genes called Present with a p-value ≤ 0.05 in at least two of the three replicates were excluded.

The DAVID gene functional classification tool (HUANG *et al.*, 2007) was used to condense the list of genes detected in our sample set into functionally related groups. We used the novel agglomeration method to cluster the three main gene ontology charts (Biological Process, Molecular Function and Cellular Component) in a meaningful network context. The following parameters were used: Classification Stringency (Medium), Similarity Term Overlap (3), Similarity Threshold (0.50), Initial Group Membership (3), Final Group Membership (3), Multiple linkage threshold (0.50).

RT-PCR analysis

Non-hybridized cRNA from one replicate of each sample (sperm cells, pollen, seedling, ovule and silique) was used to prepare double-stranded cDNA. For expression analysis, 5 ng of each template cDNA was used in reactions of 35 PCR cycles. First Strand cDNA synthesis by SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA) was performed on two additional samples of total RNA (~50ng each) from sperm cells and used, together with cDNA from pollen and leaf samples, in reactions of 40 PCR cycles. The sequences of all the primers used are available as Table S3.3.

5. REFERENCES

Alandete-Saez, M., Ron, M., and McCormick, S. (2008). GEX3, Expressed in the Male Gametophyte and in the Egg Cell of *Arabidopsis thaliana*, Is Essential for Micropylar Pollen Tube Guidance and Plays a Role during Early Embryogenesis. Mol Plant *4*, 586-598

Baarends, W.M., Hoogerbrugge, J.W., Roest, H.P., Ooms, M., Vreeburg, J., Hoeijmakers, J.H., and Grootegoed, J.A. (1999). Histone ubiquitination and chromatin remodeling in mouse spermatogenesis. Developmental biology *207*, 322-333.

Basu, D., Mallery, E., and Szymanski, D.B. (2005). SPIKE1, a DOCK-family protein, is a guanine nucleotide exchange factor for Rho of plants (ROP) and positively regulates the WAVE/ARP2/3 pathway. In Abstracts of the 16th International Conference on Arabidopsis Research (Madison, WI, USA).

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal control of embryonic patterning in *Arabidopsis thaliana*. Science *323*, 1485-1488.

Becker, D., Geiger, D., Dunkel, M., Roller, A., Bertl, A., Latz, A., Carpaneto, A., Dietrich, P., Roelfsema, M.R.G., Voelker, C.*, et al.* (2004). AtTPK4, an Arabidopsis tandem-pore K+ channel, poised to control the pollen membrane voltage in a pH- and Ca2+-dependent manner. Proc Natl Acad Sci U S A *101*, 15621-15626.

Becker, J.D., Boavida, L.C., Carneiro, J., Haury, M., and Feijo, J.A. (2003). Transcriptional profiling of Arabidopsis tissues reveals the unique characteristics of the pollen transcriptome. Plant Physiol *133*, 713-725.

Boavida, L.C., Borges, F., Becker, J.D., and Feijo, J.A. (2011). Whole genome analysis of gene expression reveals coordinated activation of signaling and metabolic pathways during pollen-pistil interactions in Arabidopsis. Plant Physiol *155*, 2066-2080.

del Pozo, J.C., and Estelle, M. (2000). F-box proteins and protein degradation: An emerging theme in cellular regulation. Plant Mol Biol *44*, 123-128.

Doelling, J.H., Phillips, A.R., Soyler-Ogretim, G., Wise, J., Chandler, J., Callis, J., Otegui, M.S., and Vierstra, R.D. (2007). The Ubiquitin-Specific Protease Subfamily UBP3/UBP4 Is Essential for Pollen Development and Transmission in Arabidopsis. Plant Physiol *145*, 801-813.

Durbarry, A., Vizir, I., and Twell, D. (2005). Male germ line development in Arabidopsis. duo pollen mutants reveal gametophytic regulators of generative cell cycle progression. Plant Physiol *137*, 297-307.

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

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

Engel, M.L., Davis, R.H., and McCormick, S. (2005). Green sperm. Identification of male gamete promoters in Arabidopsis. Plant Physiol *138*, 2124-2133.

Friedman, W.E. (1999). Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development *126*, 1065-1075.

Gusti, A., Baumberger, N., Nowack, M., Pusch, S., Eisler, H., Potuschak, T., De Veylder, L., Schnittger, A., and Genschik, P. (2009). The *Arabidopsis thaliana* F-box protein FBL17 is essential for progression through the second mitosis during pollen development. PLoS ONE *4*, e4780.

Haswell, E.S., and Meyerowitz, E.M. (2006). MscS-like proteins control plastid size and shape in *Arabidopsis thaliana*. Curr Biol *16*, 1-11.

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

Hua, Z., Meng, X., and Kao, T.H. (2007). Comparison of *Petunia inflata* S-Locus F-Box Protein (Pi SLF) with Pi SLF Like Proteins Reveals Its Unique Function in S-RNase Based Self-Incompatibility. Plant Cell *19*, 3593-3609.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2007). The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol *8*, R183.

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

Johnson, M.A., von Besser, K., Zhou, Q., Smith, E., Aux, G., Patton, D., Levin, J.Z., and Preuss, D. (2004). Arabidopsis hapless mutations define essential gametophytic functions. Genetics *168*, 971-982.

Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D., and Nam, H.G. (2008). Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. Nature *455*, 1134-1137.

Kirichok, Y., Navarro, B., and Clapham, D.E. (2006). Whole-cell patch-clamp measurements of spermatozoa reveal an alkaline-activated Ca2+ channel. Nature *439*, 737-740.

Klahre, U., Becker, C., Schmitt, A.C., and Kost, B. (2006). Nt-RhoGDI2 regulates Rac/Rop signaling and polar cell growth in tobacco pollen tubes. Plant J *46*, 1018-1031.

Klahre, U., and Kost, B. (2006). Tobacco RhoGTPase ACTIVATING PROTEIN1 spatially restricts signaling of RAC/Rop to the apex of pollen tubes. Plant Cell *18*, 3033-3046.

Lathrop, W.F., Carmichael, E.P., Myles, D.G., and Primakoff, P. (1990). cDNA cloning reveals the molecular structure of a sperm surface protein, PH-20, involved in sperm-egg adhesion and the wide distribution of its gene among mammals. J Cell Biol *111*, 2939-2949.

Li, C., and Hung Wong, W. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol *2*, RESEARCH0032.

Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A *98*, 31-36.

Liu, Y., Tewari, R., Ning, J., Blagborough, A.M., Garbom, S., Pei, J., Grishin, N.V., Steele, R.E., Sinden, R.E., Snell, W.J.*, et al.* (2008). The conserved plant sterility gene HAP2 functions after attachment of fusogenic membranes in Chlamydomonas and Plasmodium gametes. Genes Dev *22*, 1051-1068.

Matzke, M., Kanno, T., Huettel, B., Daxinger, L., and Matzke, A.J. (2007). Targets of RNA-directed DNA methylation. Curr Opin Plant Biol *10*, 512-519.

Mishra, G., Zhang, W.H., Deng, F., Zhao, J., and Wang, X.M. (2006). A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in Arabidopsis. Science *312*, 264-266.

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

Mori, T., Kuroiwa, H., Higashiyama, T., and Kuroiwa, T. (2006). Generative Cell Specific 1 Is Essential for Angiosperm Fertilization. Nat Cell Biol *8*, 64-71.

Nakazawa, Y., Hiraguri, A., Moriyama, H., and Fukuhara, T. (2007). The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. Plant Mol Biol *63*, 777-785.

Navarro, B., Kirichok, Y., and Clapham, D.E. (2007). KSper, a pH-sensitive K+ current that controls sperm membrane potential. Proc Natl Acad Sci USA *104*, 7688-7692.

Nibau, C., Wu, H.M., and Cheung, A.Y. (2006). RAC/ROP GTPases: 'hubs' for signal integration and diversification in plants. Trends Plant Sci *11*, 309-315.

Nonomura, K., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurata, N. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell *19*, 2583-2594.

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

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

Okada, T., Singh, M.B., and Bhalla, P.L. (2007). Transcriptome profiling of Lilium longiflorum generative cells by cDNA microarray. Plant Cell Rep *26*, 1045-1052.

Patton, E.E., Willems, A.R., and Tyers, M. (1998). Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. Trends Genet *14*, 236-243.

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

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

Sakai, N., Sawada, M.T., and Sawada, H. (2004). Non-traditional roles of ubiquitin-proteasome system in fertilization and gametogenesis. The international journal of biochemistry & cell biology *36*, 776-784.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of *Arabidopsis thaliana* development. Nat Genet *37*, 501-506.

Schwager, K.M., Calderon-Villalobos, L.I., Dohmann, E.M., Willige, B.C., Knierer, S., Nill, C., and Schwechheimer, C. (2007). Characterization of the VIER F-BOX PROTEINE Genes from Arabidopsis Reveals Their Importance for Plant Growth and Development. Plant Cell.

Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S., and Kao, T.H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature *429*, 302-305.

Singh, M.B., Xu, H.L., Bhalla, P.L., Zhang, Z.J., Swoboda, I., and Russell, S.D. (2002). Developmental expression of polyubiquitin genes and distribution of ubiquitinated proteins in generative and sperm cells. Sex Plant Reprod *14*, 325-329.

Strompen, G., Dettmer, J., Stierhof, Y.D., Schumacher, K., Jurgens, G., and Mayer, U. (2005). Arabidopsis vacuolar H+-ATPase subunit E isoform 1 is required for Golgi organization and vacuole function in embryogenesis. Plant J *41*, 125-132.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J *37*, 914-939.

Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell *15*, 1749-1770.

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

Vazquez, F. (2006). Arabidopsis endogenous small RNAs: highways and byways. Trends in plant science *11*, 460-468.

Voelker, C., Schmidt, D., Mueller-Roeber, B., and Czempinski, K. (2006). Members of the Arabidopsis AtTPK/KCO family form homomeric vacuolar channels in planta. Plant J *48*, 296-306.

von Besser, K., Frank, A.C., Johnson, M.A., and Preuss, D. (2006). Arabidopsis HAP2 (GCS1) is a sperm-specific gene required for pollen tube guidance and fertilization. Development *133*, 4761-4769.

Xiao, W., and Jang, J. (2000). F-box proteins in Arabidopsis. Trends in plant science *5*, 454-457.

Xu, H.L., Swoboda, I., Bhalla, P.L., Sijbers, A.M., Zhao, C.X., Ong, E.K., Hoeijmakers, J.H.J., and Singh, M.B. (1998). Plant homologue of human excision repair gene ERCC1 points to conservation of DNA repair mechanisms. Plant J *13*, 823-829.

Xu, H.L., Swoboda, I., Bhalla, P.L., and Singh, M.B. (1999a). Male gametic cell-specific expression of H2A and H3 histone genes. Plant Mol Biol *39*, 607-614.

Xu, H.L., Swoboda, I., Bhalla, P.L., and Singh, M.B. (1999b). Male gametic cell-specific gene expression in flowering plants. P Natl Acad Sci USA *96*, 2554-2558.

Zhang, Y., and McCormick, S. (2007). A distinct mechanism regulating a pollen-specific guanine nucleotide exchange factor for the small GTPase Rop in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. The EMBO journal *26*, 1691-1701.

Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. (2004). GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol *136*, 2621-2632.

CHAPTER 4

DNA Methylation Dynamics During Microgametogenesis

1. INTRODUCTION

Plant germ cells originate from meiocytes at the end of the ontogeny of flower organ development, and they must rapidly reprogramme their epigenome before fertilization. In contrast, animal germ cells are set aside of the somatic cell lineage early during embryonic development. Epigenetic reprogramming along animal germline formation occurs in primordial germ cells (PGCs) in the early embryo to re-establish totipotency, a mechanism that relies on cytidine deaminases such as AID and APOBEC1, which remove methylation marks by cytosine deamination followed by base excision repair (MORGAN *et al.*, 2004). Plants adopted a different system to erase DNA methylation that is dependent on DNA glycosylases such as the Arabidopsis DEMETER (DME) and DME-like (DML) proteins (ORTEGA-GALISTEO *et al.*, 2008; PENTERMAN *et al.*, 2007). *DME* was initially described as being preferentially expressed in the central cell and endosperm (CHOI *et al.*, 2002), where it is required for DNA demethylation of the maternal allele and establishment of genomic imprinting that leads to parent-of-origin expression of MEDEA (MEA) and FERTILIZATION INDEPENDENT SEED 2 (FIS2) Polycomb-group protein genes (GEHRING *et al*., 2006; JULLIEN *et al*., 2006), and the FLOWERING WAGENINGEN (FWA) homeodomain transcription factor gene (KINOSHITA *et al*., 2004). DME is also responsible for genome-wide erasure of DNA methylation in the endosperm (GEHRING *et al.*, 2009; HSIEH *et al.*, 2009), while the function of DML2, 3 and 4 is still mostly unknown. DME was long thought to act only in the endosperm, but a recent study showed that *DME* is also expressed in vegetative cell of the male gametophyte, being responsible for erasure of DNA methylation in the vegetative nucleus of the mature pollen grain (SCHOFT *et al.*, 2011).

Various chromatin modifications reflect the level of DNA packaging, with DNA methylation being one of the best studied. It has been implicated in transcriptional repression and gene silencing, and associated with heterochromatin formation (ROUDIER *et al.*, 2009). Heterochromatin is mostly occupied by transposable elements (TEs), which are heavily methylated in plants (LISTER *et al.*, 2008). Repression of these TEs is important for the integrity of the genome because their mobilization is innately mutagenic (LIPPMAN *et al.*, 2004; TOMPA *et al.*, 2002; TRAN *et al.*, 2005). Mutations in *DDM1* (*DECREASE IN DNA METHYLATION 1*) and *MET1* (*DNA METHYLTRANSFERASE 1*) drastically reduce DNA methylation levels, resulting in massive de-repression of TEs (LIPPMAN *et al.*, 2004). The importance of DNA methylation is further emphasized in double *met1cmt3* mutants. The loss of CMT3 (CHROMOMETHYLASE3), another DNA methyltransferase which methylates at CHG context (where H is A, T or C), further enhances *met1* phenotypes (LISTER *et al.*, 2008). In addition, MET1 activity in the gametophytes was shown to be essential for epigenetic inheritance (JULLIEN *et al.*, 2006; SAZE *et al.*, 2003). Supplementary to CG and CHG methylation, cytosines can be methylated in an asymmetric CHH context (CAO and JACOBSEN, 2002). Because CHH methylation cannot be maintained through replication it has to be re-established with each cell division in a siRNA directed manner.

Many of the components involved in RNA-directed DNA methylation (RdDM) have been identified, including the methyltransferase DRM2, the plant specific polymerases PolIV and PolV that are involved in siRNA production and amplification, as well as additional factors shown to be essential for this process (reviewed in TEIXEIRA and COLOT, 2010). Most of the current understanding of DNA methylation and RdDM in plants results from studies in Arabidopsis sporophytic tissues, while its role in the gametophytes remains fairly unexplored.

Specific retrotransposons are de-repressed and expressed exclusively in the pollen vegetative cell, as genes responsible for defense against TEs such as DDM1 are down-

regulated in the vegetative nucleus (VN) but expressed in the sperm cells (SC) (SLOTKIN *et al.*, 2009). More importantly, transcripts derived from activated TEs seem to contribute with epigenetic information through pollen-specific epigenetically activated 21nt siRNAs (easiRNAs) which accumulate in the sperm cells (SLOTKIN *et al.*, 2009). The exact purpose of these VN-derived 21nt siRNAs is not known yet, but one possible function could be to safeguard the germline before fertilization, preventing aberrant TE expression. Another hypothesis is that these siRNAs play a role after fertilization, during the rapidly occurring chromatin remodelling events throughout early embryogenesis. Regardless of the biological meaning, how this selective de-repression occurs in the VN is an important question. Also in the Arabidopsis female gametophyte, mutation in *AGO9* lead to the formation of multiple germlines, although AGO9 itself does not seem to be expressed in the gametes but in the surrounding sporophytic cells, loading TE-derived 24nt siRNAs that protect the germline against TE expression (OLMEDO-MONFIL *et al.*, 2010).

These findings are illustrative of a prominent role for siRNA-directed epigenetic regulation during gametogenesis, and prompted us to investigate DNA methylation dynamics throughout pollen development. Here we provide a first insight into the process of directed TE de-repression and siRNA production in pollen, by purifying Arabidopsis microspores and differentiated VN and SC, and performing comparative methylome analysis.

2. RESULTS AND DISCUSSION

CG and CHG methylation at TE loci is actively lost in the vegetative nucleus after pollen mitosis I

Sperm cells (SC) and vegetative nuclei (VN) were isolated by FACS, using cell specific promoters driving expression of RFP or GFP, respectively (Chapter 2), and microspores were obtained from young flower buds through a combination of mechanical filtration and purification by FACS, taking advantage of their small size and autofluorescent properties (Chapter 2). Genomic DNA was isolated from each cell/nuclei fraction, treated with sodium bisulfite and sequenced on a HiSeq platform. General comparison of the methylome between the different data sets revealed an average 3-fold change for CG and CHG methylation, and an average of 9-fold change for CHH methylation. Therefore, for further analysis we selected Differentially Methylated Regions (DMR) showing at least a 3-fold change difference. We found CG and CHG hypomethylation at TEs in the VN, in comparison with SC and microspore that share a similar CG and CHG methylation pattern (Table S4.1, Figure 4.1A). Comparing SC and VN, the majority of DMRs for CG methylation are in TE bodies: 469 out of 614 (Table S4.2), including *AtMu1a* (*At4g08680*), of which the LTR region has been previously demonstrated to be demethylated in the VN (SCHOFT *et al.*, 2011; SLOTKIN *et al.*, 2009). Similarly, our analysis revealed 114 CHG DMRs of which 109 are located within TEs (Table S4.2). Interestingly, there was little overlap between CG and CHG DMRs (Figure 4.1B), although they tended to be found in the same class of TE (Figure 4.1C). Identification of TEs corresponding to CG and CHG DMRs does not show an even distribution over the different TE families, as the majority corresponds to DNA/MuDR and RC/Helitron families (Figure 4.1C). As the loss of CG and CHG methylation is not observed genome wide and only present in the VN, this suggests active DNA demethylation, as

microspores and VN are separated by a single mitotic event. This mechanism should rely on the DNA glycosylase DME, which is implicated in DNA demethylation in the endosperm and pollen (GEHRING *et al.*, 2009; HSIEH *et al.*, 2009; SCHOFT *et al.*, 2011).

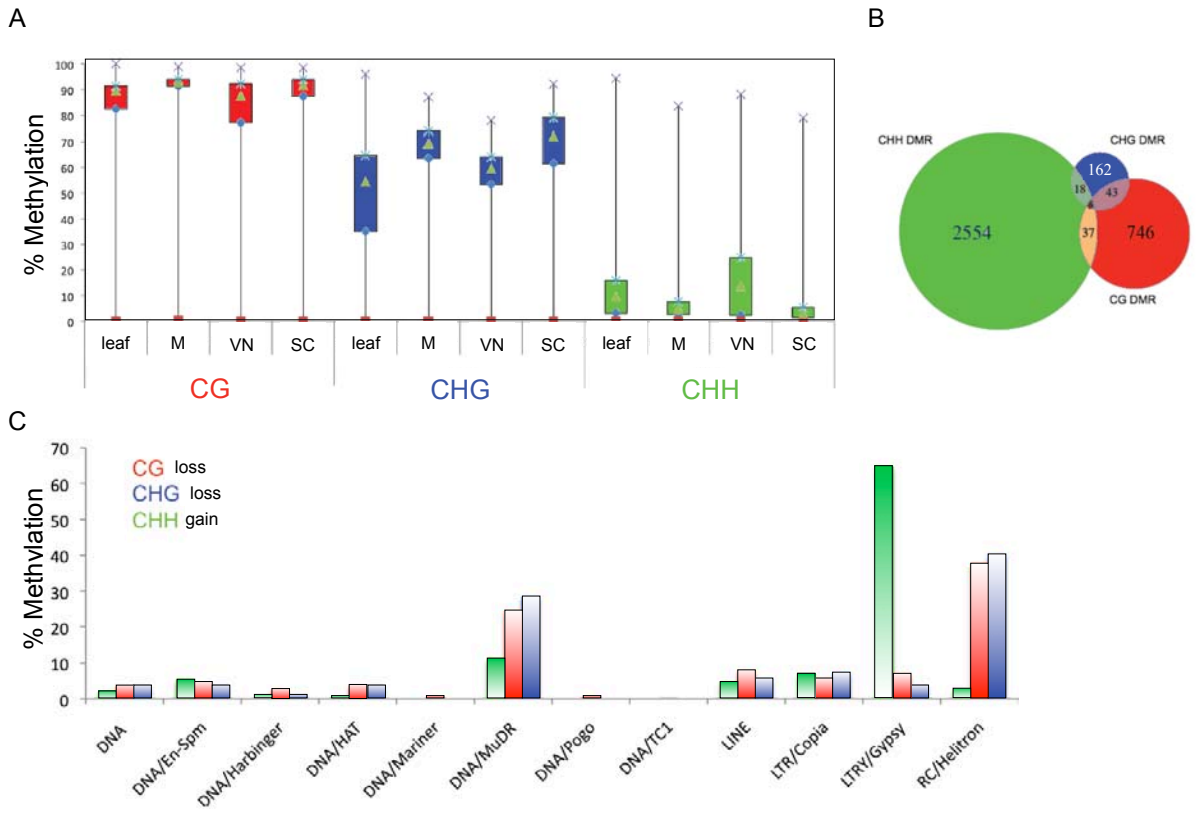


Figure 4.1 - Analysis of differentially methylated regions between microspores, sperm cells and vegetative nucleus. (A) Box plot representing DNA methylation percentages (minimum, lower quartile, median, upper quartile and maximum) at all TEs in the *Arabidopsis thaliana* genome. Overall, CG and CHG methylation is slightly lower in the VN compared to SC, microspore (M) and leaf, while CHH methylation is much higher in the VN compared to the SN. **(B)** Venn diagram showing that TEs with a statistically significant 3-fold change DMR between SN and VN do not overlap. TEs that lose CG or CHG methylation tend to have the same levels of CHH methylation. **(C)** Histogram showing the relative percentage of TE families that either lose CG/CHG or gain CHH in the VN. RC/Helitron and DNA/MuDR elements are the most represented TEs showing CG or CHG hypomethylation, while Gypsy, Copia and some MuDR and Helitrons show CHH hypermethylation.

Asymmetric CHH methylation at particular retrotransposons is erased after meiosis, and re-established specifically in the vegetative nucleus

In stark contrast to CG and CHG methylation, analysis of CHH context revealed the presence of hypomethylated regions in the microspore and SC, in comparison with that of VN and leaf tissue (Figure 4.1A). In fact, this type of methylation is almost completely lost in the microspore when comparing to the VN and SC, with 6635 DMR regions appearing hypomethylated. Roughly half of these DMRs map to unique TEs, the majority (1092) being LTR/Gypsy retrotransposons (Figure 4.1C). DMR analysis of CHH methylation between microspore and VN reveals 5114 TEs hypomethylated in the microspore. Again, the majority (2627) of these are LTR/Gypsy elements. Comparing the leaf dataset with that of microspore, SC and VN datasets reveals that RdDM targets LTR/Gypsy retrotransposon during sporophytic development, but CHH methylation is lost at the microspore stage after meiosis to be re-established later in VN. This is indicative of a resetting of CHH methylation in a subset of TEs throughout pollen development, but it is intriguing that in mature pollen RdDM re-establishes DNA methylation exclusively in the VN and not in the germline.

Distinct DNA/MuDR and RC/Helitron elements are targeted for demethylation at CG and CHG, and hypermethylation at CHH

An interesting observation comes from the analysis of DNA/MuDR and RC/Helitron elements, as different subsets of TEs appear to be differentially targeted by DNA methylation. There are 673 DNA/MuDR elements and 352 RC/Helitron that lose sporophytic levels of CHH methylation in the microspore, being later re-established in the VN, identically to what happens with LTR/Gypsy elements. In contrast, a different subset of RC/Helitron (467) and DNA/MuDR (89) elements retain sporophytic CHH levels in the microspore and differentiated VN and SC. A third group of RC/Helitron and DNA/MuDR elements have

generally low levels of CHH methylation in all the datasets, suggesting that they are not actively targeted by RdDM. Taken together, these findings suggest the existence of two distinct pathways during pollen development, with TEs that maintain or re-establish CHH methylation in the VN, and a second group that either lose CHH methylation or is never targeted by RdDM.

RdDM, as many epigenetic modifications, is linked to siRNAs. Therefore, we investigated the link between siRNA production and TE methylation in pollen. Our analysis of total pollen small RNAs revealed a two fold increase in 24nt siRNAs mapping to LTR/Gypsy, RC/Helitron and DNA/MuDR elements that re-establish or maintain CHH methylation in the VN, in comparison to other TEs of the same families with low levels or without CHH methylation. This observation is in agreement with the predominant model that especially 24nt siRNAs are essential for the *de novo* establishment of CHH methylation (ONODERA *et al.*, 2005). In contrast to CHH DMRs, a similar analysis for TEs corresponding to CG and CHG DMRs showed no fold increase of siRNAs mapping to them.

As unique RC/Helitron and DNA/MuDR elements showed different DMRs (Figure 4.2), we decided to study these more closely. To simplify the analysis we compared only methylation levels between VN and SC. First, we analyzed siRNAs and the status of symmetric methylation across the RC/Helitrons and DNA/MuDR elements that overlapped with CHH DMRs (i.e. hypermethylathed in the VN). Unsurprisingly, these regions showed a significant level of siRNAs mapping, consistent with the presence of asymmetric methylation (Figure 4.2B). In the VN, these regions have CG and CHG methylation levels that are either equal or only slightly lower in the VN than in the SC (Figure 4.2B). When examining the subset of RC/Helitron and DNA/MuDR elements that overlapped with CG or CHG DMR

between VN and SC (i.e. hypomethylated in VN) we found a significantly lower amount of siRNAs mapping to these TEs, thus correlating with the lower level of CHH methylation in both nuclei (Figure 4.2A). This indicates that there is a subset of RC/Helitron and DNA/MuDR elements that are rich in siRNAs and CHH methylation in the VN while maintaining CG and CHG methylation levels, and a second subset of elements that are devoid of siRNAs and CHH methylation, that actively lose CG and CHG in the VN (Figure 4.2). These results suggest that in the VN, DNA demethylation occurs only at TE loci that are not actively targeted by RdDM.

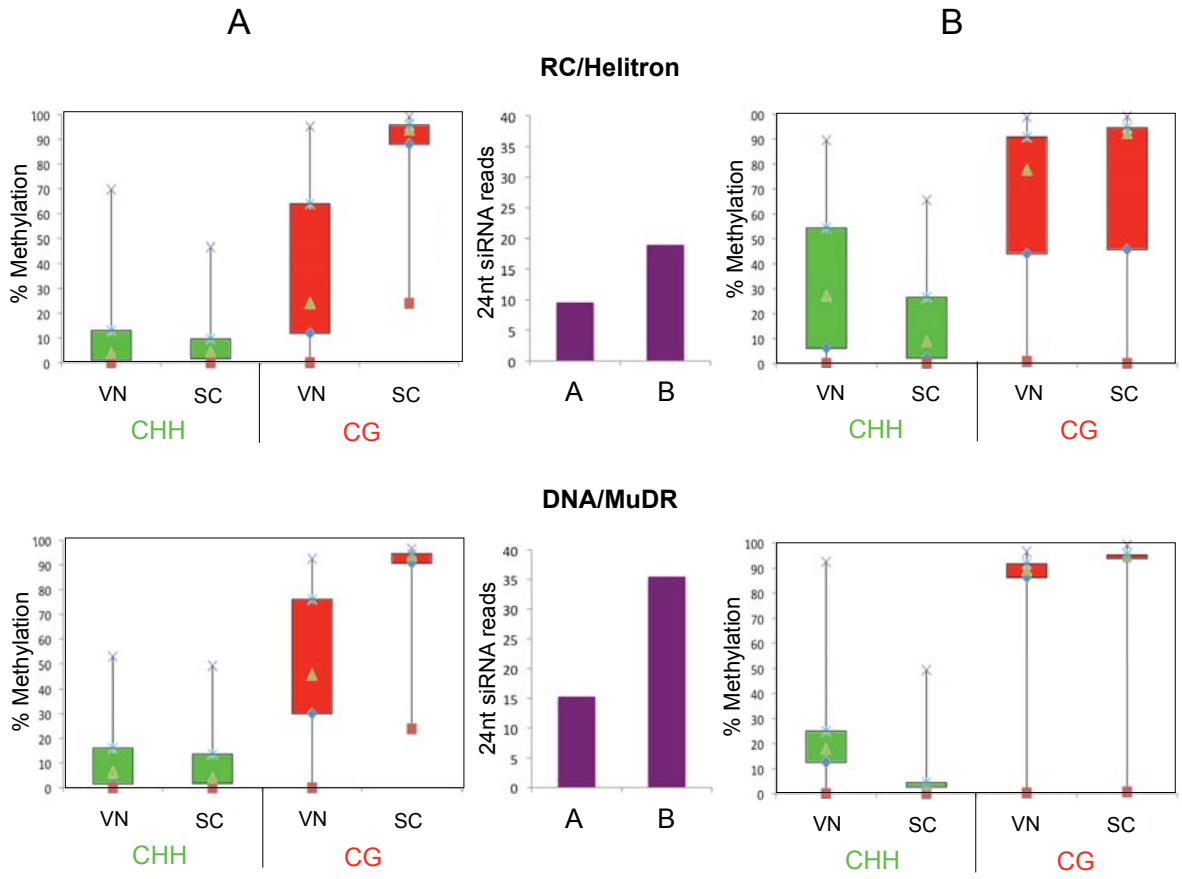


Figure 4.2 - Comparison between the percentage of DNA Methylation and siRNA profiles for RC/Helitron or DNA/MuDR elements. Box plot representation of DNA methylation percentages (minimum, lower quartile, median, upper quartile and maximum) at RC/Helitron and DNA/MuDR elements showing that TEs that are CHH hypermethylated in the VN have high levels of CG methylation **(B)**, while elements that have lower levels of

CHH methylation in the VN also lose CG methylation **(A)**. Regions with high amounts of CHH methylation have substantially more siRNAs mapping to them.

DNA methylation changes in the VN correlate with imprinted loci

There is a global resetting of DNA methylation twice during mammalian development, first during germ line maturation and then early in embryogenesis. It is important to erase methylation genome wide in the mammalian germline to ensure proper establishment of imprinted marks in the parental genomes. The mechanisms responsible for resetting imprinting marks in plants are poorly understood. Our results show that the CG and CHG methylation profiles of Arabidopsis microspores and sperm cells are similar to that of sporophytic tissue (Figure S4.1). Furthermore, our analyses indicate that the absence of CG and CHG methylation in the VN occurs only at particular loci as a result of active demethylation after the first pollen mitosis. The fact that different subsets of RC/Helitron and DNA/MuDR elements show opposite trends of DNA methylation is interesting, as a large proportion of known imprinted genes is known to be neighbored by natural transposable elements (HSIEH *et al.*, 2011; WOLFF *et al.*, 2011).

We analyzed the recently identified paternally imprinted genes having adjacent RC/Helitron or DNA/MuDR elements (HSIEH *et al.*, 2011; WOLFF *et al.*, 2011), and found that CG methylation is similarly high in the leaf, microspore and SC, but not in the VN (Figure 4.3A), which is consistent with the observations above. The well characterized paternally imprinted gene *PHERES1* (*PHE1*) shows a similar profile, being CG hypomethylated in the VN at the 3' downstream tandem repeats (MAKAREVICH *et al.*, 2008) (Figure 4.3B). These regions are devoid of CHH methylation in the microspore, VN and SC, and also tend to have low levels of CHG methylation in all tissues. We confirmed in our data

that imprinted genes normally expressed in the Arabidopsis endosperm upon DME activity, are in fact demethylated in the VN. This is the case of *MEDEA* (*MEA*) and *FWA*, which loose methylation at regulatory loci and are expressed in the vegetative cell of Arabidopsis pollen (SCHOFT *et al.*, 2011). In contrast, *PHE1* is expressed if a tandem repeat downstream of the coding region is methylated, while the maternal allele is demethylated and consequently repressed by the FERTILIZATION INDEPENDENT SEED 2 (FIS2) Polycomb group complex. Our data confirmed that this tandem repeat is hypermethylated in SC and demethylated in the VN (Figure 4.3B). As occurs in the endosperm, demethylation of the 3' region downstream of *PHE1* in the VN is possibly the result of DME activity. Considering both *MEA* and *PHE1* imprinting in the endosperm, it seems that CG methylation is either involved in repressing or activating paternal alleles, or the paternal PHE1 allele is also targeted by DME after fertilization. Interestingly, a demethylated paternal allele of *PHE1* transmitted through *met1* pollen is considerably repressed after fertilization, resulting in predominant expression of the maternal allele in developing seeds (MAKAREVICH *et al.*, 2008). The mechanism controlling *PHE1* imprinting is not yet understood, but by analyzing the results derived from crosses with *met1* mutants, it seems to require that the tandem repeats between the parental alleles remain differentially methylated after fertilization. CG methylation in the gene-body is also involved in the activation of *DME* and *ROS1* (another DNA glycosylase). Expression of these genes is repressed in *met1* mutants because their gene-body CG methylation is lost (MATHIEU *et al.*, 2007). *PHE1*-homolog *PHE2* is also regulated by FIS PcG complex and is specifically expressed in the endosperm (VILLAR *et al.*, 2009). Although the FIS PcG complex targets *PHE2*, it lacks downstream tandem repeats like *PHE1* and for this reason is not imprinted. However, we found CG methylation in the gene body of *PHE2* in both SC and VN, but not in microspores and leaf (Figure 4.3B), suggesting that this methylation mark is established *de novo* during pollen development and is not

removed by DME. It would be interesting to analyze if this methylation mark is important for *PHE2* expression in the endosperm.

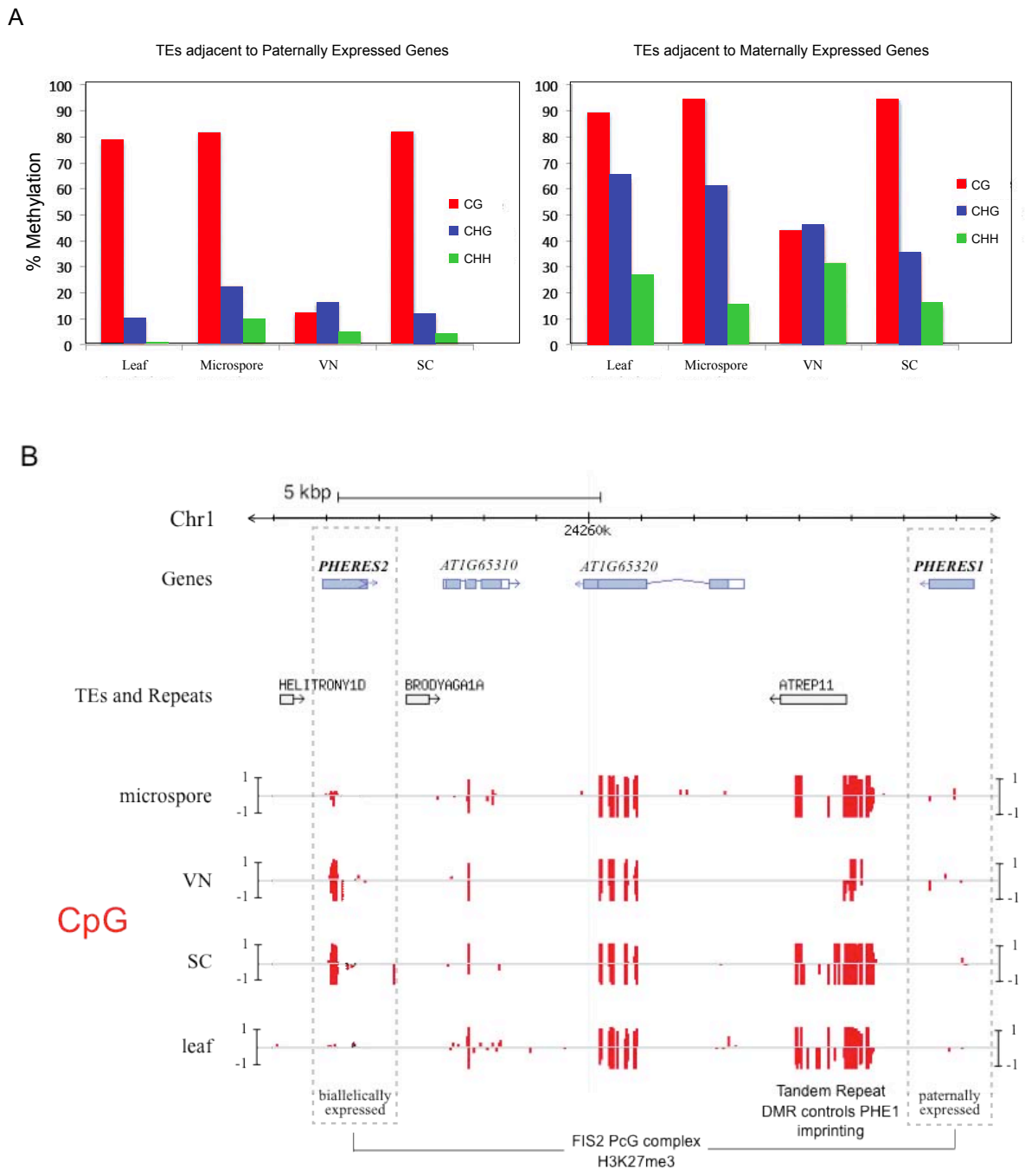


Figure 4.3 - DNA methylation at TEs neighboring paternally and maternally imprinted genes and CG methylation of PHERES genes. **(A)** Relative DNA methylation at TEs neighboring imprinted genes. CHG and CHH methylation in pollen and sporophytic tissues is much lower in paternally expressed genes (PEGs) than that of maternally expressed genes (MEGs). CG is only partially lost in the VN at TEs surrounding MEGs, suggesting that

RdDM might counteract active demethylation by DME at these loci. In contrast, PEGs have low levels of CHG and CHH methylation, and CG is massively lost in the VN. **(B)** Genome browser snapshot showing differentially methylated regions (DMRs) in the CG context, between sperm cell (SC), vegetative nucleus (VN), microspore and leaf DNA samples. PHERES1 (PHE1) is an imprinted gene because the paternal allele is highly methylated on a tandem repeat region at the 3' region downstream of the gene. PHE2 is not imprinted but is also controlled by the FIS2 Polycomb Group (PcG) complex. Demethylation observed in the PHE1 tandem repeats is the consequence of DME activity in pollen, thus reflecting what happens to the maternal allele in the endosperm. PHE2 shows a newly identified CG mark in the gene body, which seems to be established de novo during pollen development in both SC and VN. The correlation of this methylation mark with PHE2 expression in the endosperm is unknown.

In contrast to paternally expressed genes, when examining solely maternally imprinted loci (HSIEH *et al.*, 2011; WOLFF *et al.*, 2011) the situation is more complex, as only a few of them have surrounding RC/Helitron or DNA/MuDR elements. Looking specifically at these elements we found that a subset have high amounts of CG methylation, with levels consistently high in microspore and SC, and only slightly reduced in the VN (Figure 4.3A). With regard to CHH methylation, these elements show a reduction of methylation in the microspore, but maintain high levels of methylation in the VN, consistent with levels seen in leaf tissue (Figure 4.3A). We analyzed a specific group of maternally expressed genes (MEGs) reported previously (WOLFF *et al.*, 2011), and showing CG methylation marks at TEs adjacent to the coding region. In order to understand if DNA methylation is the driving force repressing the paternal allele, WOLFF *et al*. (2011) pollinated wild-type plants with *met1* pollen that lose maintenance of CG methylation. This could show that the paternal alleles of most MEGs remain silenced even without MET1 activity in the sperm cells, but one of these genes became biallelically expressed (*At3g23060*). Interestingly, analyzing DNA methylation around these genes in sperm cells showed that regulatory elements surrounding *At3g23060*

have only CG methylation, while the genes for which the paternal allele remains repressed without CG methylation show additional CHG and CHH methylation, indicating active silencing by RdDM (Figure 4.4). This observation indicates that MEGs *At5g03020* and *At1g60970* are genes with exclusive expression from the maternal allele, because the paternal allele is targeted by RdDM in the sperm cells, and it seems that this repression remains after fertilization. Interestingly, expression of both genes is not deregulated in *fis2* mutants, suggesting that RdDM is the main silencing mechanism. If this is indeed the mechanism responsible for silencing these genes in the endosperm after fertilization, it suggests that RdDM preferentially targets the paternal genome after fertilization, as it seems to occur in the early embryo (AUTRAN *et al.*, 2011).

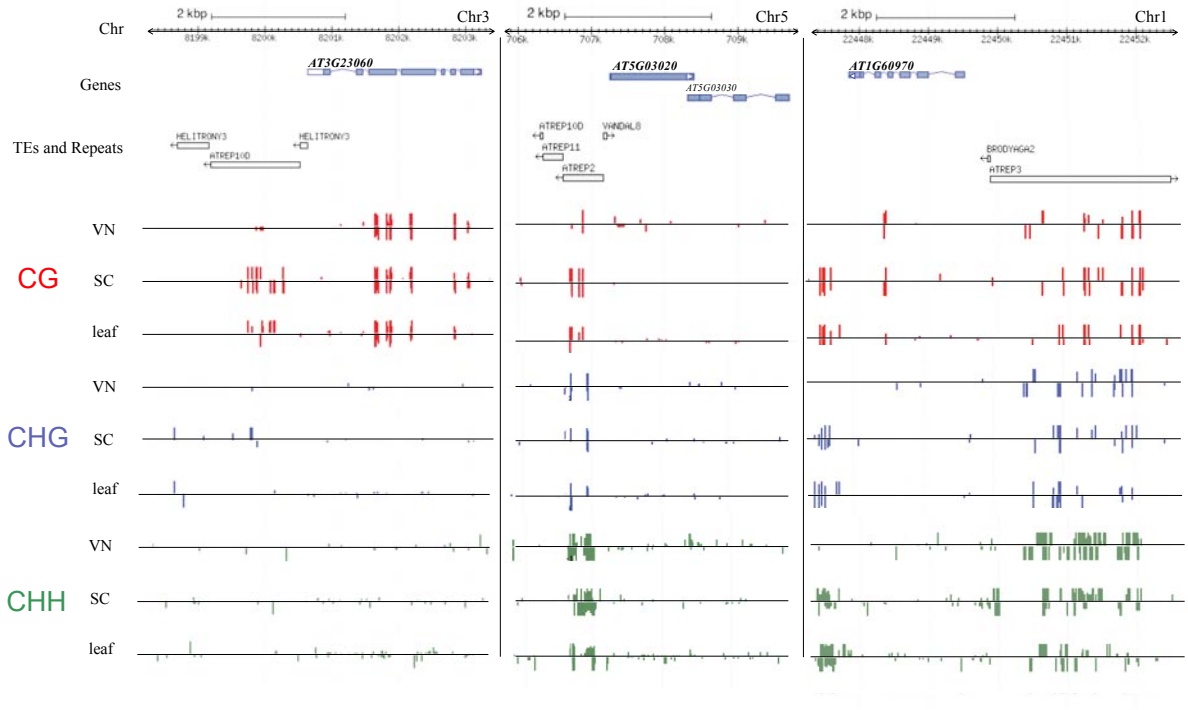


Figure 4.4 – DNA methylation at Maternally Expressed Genes (MEGs). Genome browser snapshot showing DNA methylation in CG, CHG and CHH contexts, between sperm cell (SC), vegetative nucleus (VN) and leaf DNA samples. *At3g23060*, *At5g03020* and *At1g60970* constitute previously described MEGs (WOLFF *et al.*, 2011), which are differentially regulated during seed development resulting from crosses between wild-type pistil and *met1* pollen. While *At5g03020* and *At1g60970* remain maternally expressed, *At3g23060* becomes

biparentally expressed, which correlates with the fact that *At3g23060* lacks DNA methylation characteristic of RdDM activity. This evidence could explain why repression is completely lost when this allele is inherited through *met1* pollen.

3. CONCLUSIONS

Profiling whole-genome DNA methylation of microspore, sperm cells and vegetative nucleus by bisulfite sequencing, could demonstrate that epigenetic reprogramming during male gametogenesis is markedly associated with differential DNA methylation occurring in the VN, which shows demethylation of symmetric CG and CHG, and hypermethylation at CHH for a specific class of transposable elements. As previously reported, CG demethylation in the VN is possibly the result of DME activity in pollen (SCHOFT *et al.*, 2011), but CHH hypermethylation in the VN, especially that observed at retrotransposons and pericentromeric regions, is still a matter of debate. Early observations of DNA methylation in pollen using an anti-5mC antibody, could already verify that in mature pollen of tobacco and *Lilium* sp. the VN has much higher levels of DNA methylation than the generative cell (JANOUSEK *et al.*, 2000; OAKELEY and JOST, 1996; OAKELEY *et al.*, 1997). Interestingly, one of these studies correlated DNA hypermethylation with H4 deacetylation in the VN (JANOUSEK *et al.*, 2000), which again does not fit with the decondensed state of the VN. However, this observation gives a further insight into the hypothesis that TE derepression and decondensed chromatin in the VN might also reflect the consequence of particular histone modifications that are coupled to RdDM. In contrast, analysis of DNA methylation in mature pollen of cork oak (*Quercus suber*), revealed an opposite trend with the VN being hypomethylated in comparison with the generative cell nucleus (RIBEIRO *et al*., 2009). As cork oak has a long progamic phase, this discrepancy may be reflecting a different spatial and temporal control of DNA methylation dynamics during pollen development.

The Arabidopsis microspore and SC show residual CHH methylation compared to the VN, indicating that components of the RdDM pathway are either down-regulated or more

selective throughout meiosis and SC specification. On the other hand, CG and CHG marks in the microspore and SC are comparable or even higher than in a leaf tissue, suggesting that throughout Arabidopsis development there is never a genome-wide erasure and reprogramming of CG methylation, as its stability seems to be required for proper transgenerational inheritance of epigenetic information (MATHIEU *et al.*, 2007). Moreover, it is possible that sperm cells are still going through epigenetic reprogramming while inside the pollen tube, as in mature pollen at anthesis they are still progressing through the S-phase (FRIEDMAN, 1999), the cell cycle stage where most epigenetic modifications are established (KLOC *et al.*, 2008).

Our analyses indicate that TEs surrounding paternally imprinted genes are somehow marked by lack of CHH methylation and targeted for DNA demethylation later during pollen development in the VN. On the other hand, RdDM actively targets TEs surrounding maternally expressed genes in pollen and sporophytic tissues. The entire resetting and reestablishment of DNA methylation at specific TEs in the VN could potentially be a means of reprogramming imprinted loci during pollen development, as illustrated in Figure 4.5. These observations provide a potential new avenue for the discovery of novel imprinted genes.

Figure 4.5

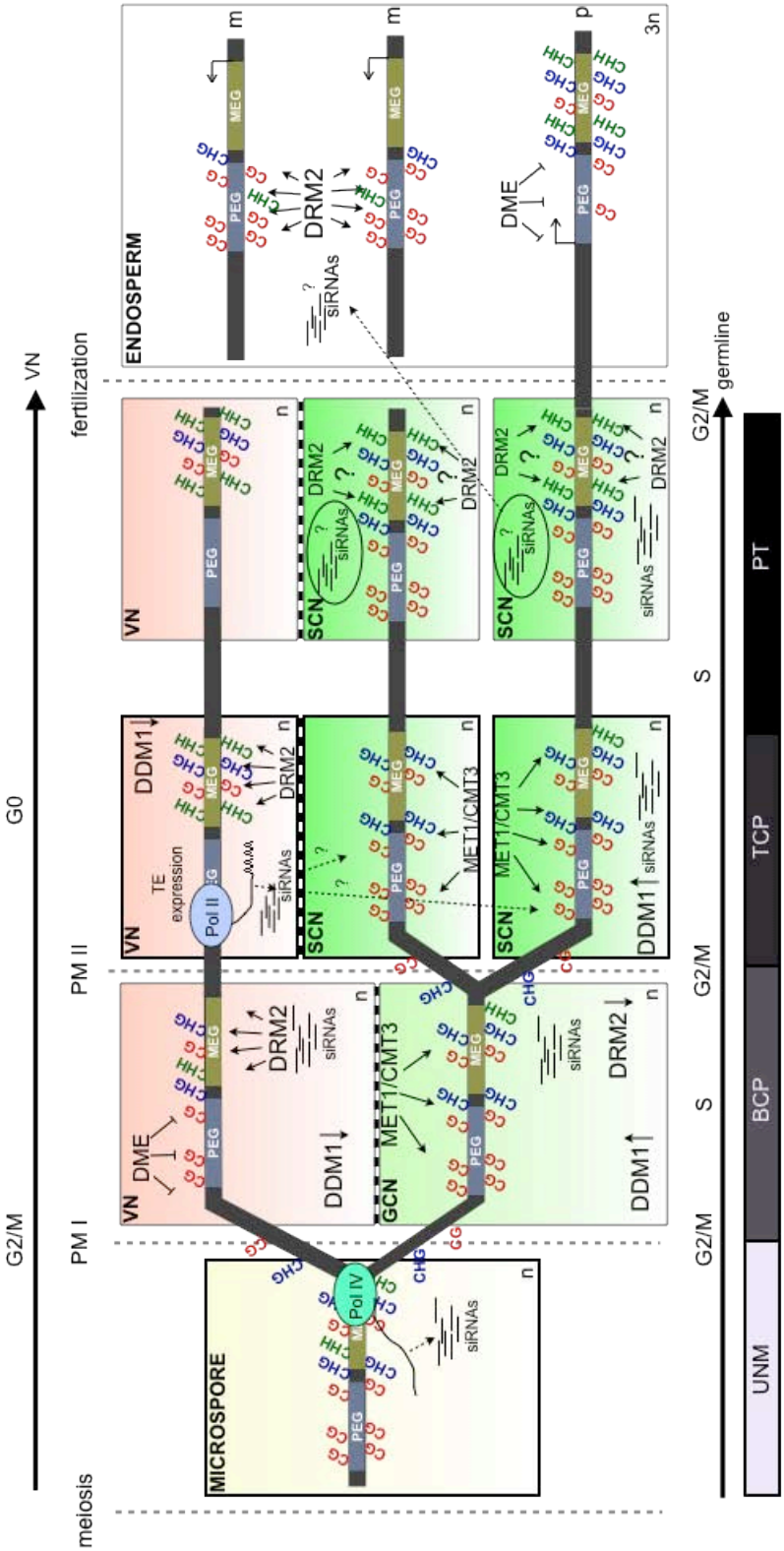

ENDOSPERM
3n
MICROSPORE
VN
SCN
GCN
DDM1
DRM2
MET1/CMT3
DME
Pol II
Pol IV
TE expression
siRNAs
MEG
PEG
fertilization
melosis
G0
G2/M
PM I
PM II
S
germline
UNM
BCP
TCP
PT

Figure 4.5 - Schematic representation of a model proposing DNA methylation and small RNA dynamics at imprinted loci throughout Arabidopsis pollen development. Post-meiotic haploid microspores (yellow gradient) divide initially on pollen mitosis I (PM I), to originate the vegetative nucleus (VN, red gradient) and a generative cell (light green gradient), which divides further on PM II giving rise to the sperm cells (green gradient). One sperm cell will fuse with the central cell in the female gametophyte to originate the triploid endosperm (white), where parent-of-origin expression is known to occur in plants. In the microspore, TEs neighboring Maternally Expressed Genes (MEG) produce siRNAs by Pol IV-derived transcription, maintaining CHH methylation after meiosis. After PM I, DRM2 is expressed preferentially in the VN (Fred Berger, unpublished), re-establishing CHH methylation by RdDM while Pol IV produces more siRNA from these regions. Paternally Expressed Genes (PEG) are not targeted by RdDM in the VN and consequently lose remnants of CG methylation by DME activity. In the generative cell nucleus (GCN) and subsequent sperm cell nuclei (SCN) after PM II, MET1 and CMT3 maintain CG and CHG methylation. As DRM2 is not expressed in the germline at this point, CHH methylation is passively lost after pollen mitosis, even if siRNAs derived from Pol IV transcription in the microspores and VN are present in the germline. DME activity combined with down-regulation of DDM1 and chromatin decondensation in the VN, results in reactivation of Pol II-derived expression of TEs surrounding PEGs. siRNAs produced from these loci may accumulate later in the germline, but were not detectable in mature pollen or sperm cells. During pollen tube growth, *DRM2* expression might be activated in the sperm cells, reinforcing CHH methylation at MEGs, which remains after fertilization in the endosperm. siRNAs derived from the expression of PEGs in the VN, should also accumulate in the sperm cells but not routed into the RdDM pathway, being delivered instead at fertilization to target DNA methylation at the maternal alleles of PEGs. The paternal alleles of PEGs have only CG methylation, which can be erased by DME leading to expression of the gene. The maternal allele of MEGs might be completely demethylated before fertilization, because of DME activity in the central cell, and as such is not targeted by the RdDM machinery. Active protection of CGs from remethylation once they have completely lost DNA methylation has been previously observed (MATHIEU *et al*., 2007). UNM, uninucleate microspore; BCP, bicellular pollen, TCP, tricellular pollen; PT, pollen tube. Upper and lower solid lines show cell cycle transitions in the VN and germline.

4. MATERIALS AND METHODS

Library preparation from bisulfite treated DNA

Genomic DNA was isolated from approximately 600.000 sperm cells, 300.000 vegetative nuclei and 1.000.000 microspores, and fragmented by Covaris in 10mM Tris-HCl, pH 8.0. Fragments were end repaired, A-tailed and ligated to methylated Illumina adaptors. Ligated fragments were bisulfite treated using the EZ DNA Methylation-Gold Kit (Zymo), and PCR enriched with Expand High-Fidelity Polymerase (Roche). Amplified fragments of 340-360bp were size selected by gel extraction, and sequenced on a HiSeq platform.

Analysis of Differentially Methylated Regions

For each individual data set the number of methylated cytosines and unmethylated cytosines was calculated in moving windows of fifteen cytosines across the genome. This method of analysis excluded single methylation site artifacts, and enabled us to detect differentially methylated regions (DMR) between two samples in each window by performing a two-sided fisher's exact test of independence. DMRs were defined as statistically significant bins with a p-value of 0.01, and were filtered for number of calls made for each C. The percentage of methylation was determined for each region, and was subsequently used to compute a fold change of methylation within DMRs.

5. REFERENCES

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

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

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

Friedman, W.E. (1999). Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development *126*, 1065-1075.

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

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

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

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

Janousek, B., Zluvova, J., and Vyskot, B. (2000). Histone H4 acetylation and DNA methylation dynamics during pollen development. Protoplasma *211*, 116-122.

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

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

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. (2008). RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol *18*, 490-495.

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

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

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

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

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Oakeley, E.J., and Jost, J.P. (1996). Non-symmetrical cytosine methylation in tobacco pollen DNA. Plant Mol Biol *31*, 927-930.

Oakeley, E.J., Podesta, A., and Jost, J.P. (1997). Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. Proc Natl Acad Sci U S A *94*, 11721-11725.

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

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

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

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*, 6752-6757.

Ribeiro, T., Viegas, W., and Morais-Cecilio, L. (2009). Epigenetic marks in the mature pollen of *Quercus suber* L. (Fagaceae). Sex Plant Reprod *22*, 1-7.

Roudier, F., Teixeira, F.K., and Colot, V. (2009). Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends Genet *25*, 511-517.

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

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

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell *136*, 461-472.

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

Tompa, R., McCallum, C.M., Delrow, J., Henikoff, J.G., van Steensel, B., and Henikoff, S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. Curr Biol *12*, 65-68.

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 Arabidopsis genes. Curr Biol *15*, 154-159.

Villar, C.B., Erilova, A., Makarevich, G., Trosch, R., and Kohler, C. (2009). Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. Mol Plant *2*, 654-660.

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

CHAPTER 5

MicroRNA Activity in the Arabidopsis Male Germline

Chapter section partially published as original article in peer-review journal:

Borges F, Pereira PA, Slotkin RK, Martienssen RA, Becker JD (2011) MicroRNA activity in the Arabidopsis male germline. J Exp Bot **62:** 1611-1620

1. INTRODUCTION

Post-transcriptional gene silencing was first observed over a decade ago and since then numerous discoveries unveiled a fascinating and unexpectedly conserved cellular mechanism. In addition, several studies from plants to animals underlined distinct regulatory mechanisms that evolved throughout different cell lineages. Germ cells developed specific mechanisms associated to chromatin remodeling and resetting of epigenetic marks that must be properly established before being transmitted to the next generation. For example, animals have a germ cell-specific class of small RNAs termed piRNAs (Piwi-interacting RNAs), associated with TE silencing and heterochromatin formation (CARMELL *et al.*, 2007; VAGIN *et al.*, 2006). Plants lack piRNAs and Piwi proteins, but alternatively, particular TE silencing mechanisms are also specifically established in the germline (OLMEDO-MONFIL *et al.*, 2010; SLOTKIN *et al.*, 2009). Thus, it is likely that certain cellular mechanisms controlling miRNA activity are also differentially regulated between the somatic and germ cell lineages.

The miRNA pathway is an essential mechanism to regulate different biological processes. In plants it is involved in several fundamental processes such as development, response to biotic and abiotic stresses and hormone responses. These small non-coding RNAs act through cleavage of highly complementary target mRNAs, but also through translational repression (BRODERSEN *et al.*, 2008; CHEN, 2004). More indirectly, miRNA-guided cleavage of *TAS* transcripts gives rise to trans-acting siRNAs (ta-siRNA) that regulate expression of other genes (ALLEN *et al.*, 2005; MONTGOMERY *et al.*, 2008). miRNA biogenesis depends on processing of stem-loop precursors by DICER-Like 1 (DCL1), and subsequent cleavage of target mRNAs requires binding of mature miRNAs to ARGONAUTE 1 (AGO1) (reviewed in VOINNET, 2009). Additionally, some miRNAs may also associate with AGO10 promoting

translational repression of specific target transcripts (BRODERSEN *et al.*, 2008). Based on overlapping data generated by numerous high-throughput sRNA sequencing studies, miRNA-like RNAs as well as other types of small RNAs resulting from known precursor stem loops were identified (VAZQUEZ *et al.*, 2008; ZHANG *et al.*, 2010). While processing of most miRNA-like RNAs described is dependent on the regular DCL1-based processing pathway (ZHANG *et al.*, 2010), it was additionally shown that the production of some long miRNAs (23-27nt) is dependent on the DCL3/RDR2/Pol IV pathway (ZHANG *et al.*, 2010).

Although miRNA activity is well characterized in the Arabidopsis sporophyte, little is known about the role of miRNAs during male gametophyte development. In contrast, the genetic programs driving male gametogenesis and pollen development have been intensively studied and reported throughout the last decade mostly by means of microarray-based studies, but also with the characterization of loss-of-function mutations impairing gametophyte development (BECKER and FEIJO, 2007; BORG *et al.*, 2009). These studies hold great promise towards the identification of a core transcriptional program driving plant germline differentiation and specification. However, understanding post-transcriptional regulation of gene expression in pollen and the gametes requires additional efforts.

2. RESULTS AND DISCUSSION

Comparative analyses of microRNAs accumulated in pollen and sperm cells

Small RNAs of 19-28nt from purified sperm cells and pollen were cloned and sequenced on Illumina 1G sequencer as previously described (SLOTKIN *et al.*, 2009). From the 2,540,585 signatures sequenced in pollen and 1,925,202 in sperm cells, 283,561 and 256,787 sequences matched to known Arabidopsis miRNAs in pollen and sperm, respectively. The list of most known miRNAs in these two datasets is publicly available in the SBS database (NAKANO *et al.*, 2006). We compared our data with the recently available dataset obtained by 454 sequencing of pollen small RNAs from Col ecotype (GRANT-DOWNTON *et al.*, 2009b). Their study could identify 31 different miRNA families in mature pollen, from which only miR776 was not detected in our pollen sample. In total, we could identify 75 known miRNA families expressed in pollen, and 83 in sperm cells (Figure 5.1A), 49 of which with higher expression than in whole pollen. A summary of normalized reads matching to the annotated form of known miRNAs in sperm cells and pollen datasets in comparison with that of inflorescence is presented in Table S5.1.

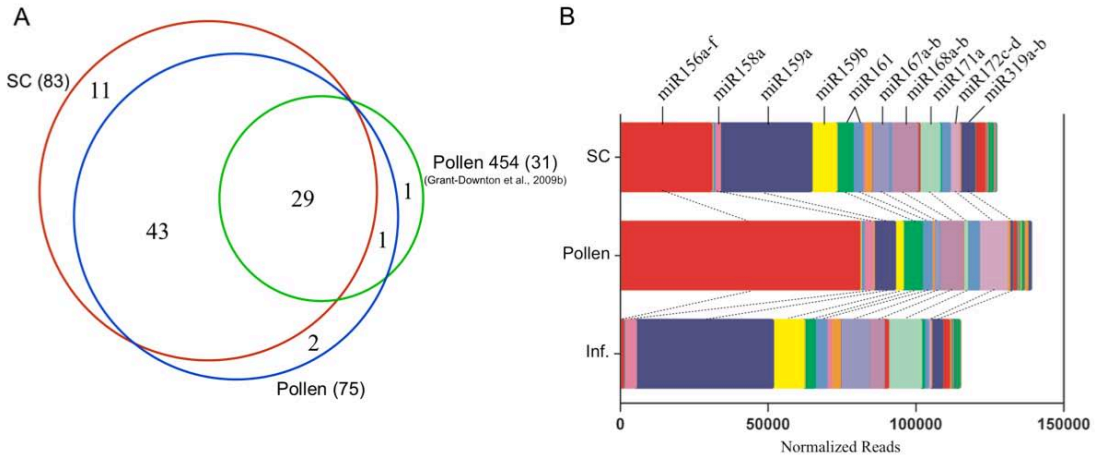


Figure 5.1 - MicroRNA families detected in sperm cells and pollen. (A) Venn diagram illustrating overlap between known miRNA families detected in sperm cells (83) and pollen (75) by Illumina sequencing (SLOTKIN *et al*., 2009), and a 454 sequencing data set of pollen small RNAs (31) reported by GRANT-DOWNTON *et al*. (2009b). Numbers in parentheses represent total miRNA families identified in each data set. **(B)** Relative abundance of known miRNAs detected in sperm cell (SC), pollen, and inflorescence data sets.

Expression of most miRNA families seems to be distinct between total pollen and purified sperm cells (Figure 5.1B), which was expected considering their different transcriptomes and cell fate (BORGES *et al.*, 2008). miR159a is particularly interesting as it is very abundant in sperm cells. Keeping in mind that miRNAs highly expressed in sperm cells should be identified in pollen samples despite a dilution effect, the significantly lower abundance of miR159a in pollen suggests that it could be sperm-specific. miR159 is involved in the regulation of several transcripts belonging to the MYB family of transcription factors, including *DUO1* (PALATNIK *et al.*, 2007), a germ cell-specific transcription factor that is responsible for expression of several germline-specific genes (BROWNFIELD *et al.*, 2009). miR159-guided cleavage of *DUO1* was detected in mature pollen (GRANT-DOWNTON *et al.*,

2009a), but is not clear if *DUO1* expression is modulated post-transcriptionally in the sperm cells or in the vegetative cell, where it should not be expressed. Initially, it was proposed that *DUO1* expression in somatic cells could be repressed at the transcriptional level by a repressor protein (GRSF) that binds to a consensus motif located in the promoters of germline genes (HAERIZADEH *et al.*, 2006). That silencer sequence was found in the promoters of Arabidopsis genes specifically expressed in the sperm cells such as *DUO1* and *MGH3*, but it seems that it is not required for repression in non-germline cells (BROWNFIELD *et al.*, 2009). As miR159a was already shown to direct cleavage of *DUO1* transcripts (GRANT-DOWNTON *et al.*, 2009a; PALATNIK *et al.*, 2007), is important to understand now whether certain cell lineages in plants have particular RNA-binding factors that could prohibit miRNA association to target transcripts, as it was demonstrated in animals (KEDDE *et al.*, 2007).

MicroRNA families such as miR156 and miR158 have isoforms in which the 5' terminal nucleotide is a Cytosine (miR156g and miR158b, respectively). Such isoforms were predicted to be associated preferentially with AGO5 (MI *et al.*, 2008; TAKEDA *et al.*, 2008), which correlates with the fact that miR156g and miR158b levels are higher in pollen and sperm than in sporophytic tissues (Table S5.1). However, it is intriguing that miR156a-f is so highly enriched in pollen and sperm (Figure 5.1B), as the entire family is predicted to target *SPL* genes (RHOADES *et al.*, 2002). This possibly indicates that the same miRNA family may have distinct activity by association with specific AGO complexes. Another example is miR845a that is highly abundant in pollen and sperm, but its targets are still unknown. Both miR845a and 845b are almost undetectable in inflorescence tissue (Table S5.1), which suggest that they must be preferentially processed in sperm cells and pollen.

Analysis of potentially novel microRNAs and natural variations in sequence length of known miRNA families

To screen for potentially new microRNAs in our sperm cell and pollen sRNA datasets, we used the miRCat tool available online within the UEA plant sRNA toolkit (MOXON *et al.*, 2008). Based on correspondence with genomic loci that could encode stem-loop precursors, we could identify 25 small RNA sequences as potential novel miRNAs. These results, including a number of predicted target transcripts for the novel miRNAs found, are presented in Table 5.1 and extended in Table S5.2. Recently, another computational method was developed for discovering novel plant miRNAs in high-throughout data (PIPmiR), and it was also able to detect miR5014, miR5020a, miR5024, and miR5026 (Table 5.1) as high-confident novel miRNAs in Arabidopsis (BREAKFIELD *et al*., 2012).

Some miRNA sequences discovered by miRCat are isoforms of known miRNA families with variations in sequence length (Table 5.1). This is the case of miR158b that was detected with an additional Uracil at the 5' nucleotide terminus. Interestingly, this isoform is more abundant in pollen and sperm cells than the normal annotated version (Figure 5.3A). The biological meaning for such variations in microRNA processing is not yet fully understood, but it was recently reported that single nucleotide extensions at the 5' terminus of known miRNAs can lead to incorporation into different Argonaute complexes (EBHARDT *et al.*, 2010). For this reason a sperm-enriched miR156h isoform with an extension at the 5' end and higher abundance in our datasets than the known 20-nt isoform deserves a closer attention (Figure 5.3A). This variation in natural sequence length of miR156h was analyzed by Ebhardt et al. (2010), and curiously, they observed that the known 20-nt version of miR156h is loaded into AGO1, while the miR156h plus 5' U isoform seems to associate mainly with AGO5.

Table 5.1 – Normalized reads for potentially novel miRNAs and isoforms with variations in the sequence length of known miRNA/miRNA* detected by miRCat tool, in sRNA datasets of Col pollen and sperm cells (SC) in comparison with inflorescence tissue (Inf). Underlined sequences indicate the mature form annotated in ASRP[(a)], including missing nucleotides (lower case). These variations were preferentially detected by miRCat since they are more abundant than the annotated isoforms of known miRNAs. The activity of two novel miRNA was tested by 5' RACE for targets marked in bold.

microRNA	Sequence	SC	Pollen	Inf	psRNA target(s)
Variations of known miRNA/miRNA[(a)]*					
miR156h	UUGACAGAAGAAAGAGAGCAC	107	24	162	SQUAMOSA PROMOTER BINDING PROTEIN-LIKE
miR158b	UCCCCAAAUGUAGACAAAGCA	248	1099	12	AT2G46590
miR161a.1	UUGAAAGUGACUACAUCGGGGu	5605	7146	2066	PPR AT5G41170
miR162	AUCGAUAAACCUCUGCAUCCAGg	4	58	1	DICER-LIKE 1
miR165*	gGAAUGUUGUCUGGAUCGAGGA	46	102	9	-
miR167c	uUAAGCUGCCAGCAUGAUCUUGU	10	0	2	-
miR408*	ACAGGGAACAAGCAGAGCATGG	589	29	12	AT2G47020, AT4G02940, AT1G04210, AT1G17180, AT4G03950
miR773a	TTTGCTTCCAGCTTTTGTCTCC	82	524	22	MET2 AT4G14140, AT4G05390, AT4G08990
miR778*	ACAAACUCGGUGUACAUAGACcaaaccaag	116	316	2	AT5G22380, AT1G69610
miR840*	UUGUUUAGGUCCCUUAGUUUCu	21	27	9	-
miR844a	AAUGGUAAGAUUGCUUAUAAGcu	11	22	11	AT2G13720, AT5G44120
miR852	AAGAUAAGCGCCUUAGUUCUGA	42	63	15	AT5G56650, AT5G56660
miR852a	AAGAUAAGCGCCUUAGUUCUGA	32	107	7	-
miR860	ucaAUAGAUUGGACUAUGUAUAUU	17	9	10	AT5G26030, AT3G12640
miR863a	uugAGAGCAACAAGACAUAAUAAAGAG	2	11	0	-
miR870	AAUCUAAUUUGGUGUUUCUUCGaug	2	13	1	AT3G55370
Potentially novel miRNAs [(b)]					
miR2934[(c)]	CAUCCAAGGUGUUUGUAGAAA*	101	392	6	-
miR4240.2[(d)]	AUGGCUAGAGUGACUAGACCCG	23	0	10	-
miR447a.2	UAUGGAAGAAAUUGUAGUAUU	84	42	152	AT1G42630, AT1G54710
miR447c.2	CCCCUUACAAUGUCGAGUAAA	0	13	3	-
miR868.2	UCAUGUCGUAAUAGUAGUCAC	32	92	5	CMT1 AT1G80740, CMT2 AT4G19020
miR5012	UUUUACUGCUACUUGUGUUCC	6	0	3	AT1G53700, AT2G37678
miR5013	UUUGUGACAUCUAGGUGCUUU	86	317	1	**AT3G60580**
miR5014	UUGUACAAAUUUAAGUGUACG	15	4	9	-
miR5015.1	UUGGUGUUAUGUGUAGUCUUC	17	20	0	-
miR5015.2	UCUGUUGUUGUUGGUGUUAUG	17	38	2	AT2G38320, AT1G12860, AT2G01760, AT5G38740
miR5016	UUCUUGUGGAUUCCUUGGAAA	2	5	0	-
miR5017	UUAUACCAAAUUAAUAGCAAA	4	5	67	-
miR5018	UUAAAGCUCCACCAUGAGUCCAAU	0	5	0	-
miR5019	UGUUGGGAAAGAAAAACUCUU	6	40	2	AT3G58810, AT1G14510, AT4G19550
miR5020a	UGGAAGAAGGUGAGACUUGCA	15	53	0	-
miR5020b	AUGGCAUGAAAGAAGGUGAGA	158	711	1	-
miR5021	UGAGAAGAAGAAGAAGAAAA	0	7	0	-
miR5022	GUCAUGGGGUAUGAUCGAAUG*	17	56	0	-
miR5023	AUUGGUAGUGGAUAAGGGGGC*	0	18	0	AT5G24950
miR5024	AUGACAAGGCCAAGAUAUAACA	4	7	0	-
miR5025	ACUGUAUAUAUGUAAGUGACA	4	9	0	AT2G48010
miR5026	ACUCAUAAGAUCGUGACACGU	10	16	2	-
miR5027	ACCGGUUGGAACUUGCCUUAA	10	27	3	-
miR5028	AAUUGGGUUUAUGCUAGAGUU	92	443	8	-
miR5029	AAUGAGAGAGAACACUGCAAA*	84	288	1	AT2G30070, **AT4G02900**

(a) – Annotation as in ASRP: The *Arabidopsis* Small RNA Project Database (http://asrp.cgrb.oregonstate.edu/).
(b) – No correspondence found in ASRP database. Predicted foldback structures and chromosome loci presented in Table S5.2.
(c) – Potentially mature miRNA sequence processed from ath-MIR2934, more abundant in our SC and Pollen datasets than ath-miR2934 UCUUUCUGCAAACGCCUUGGA reported in Grant-Downton *et al*. (2009b), but classified in our study as ath-miR2934* (Table S5.1).
(d) – Processed from ath-MIR4240 (Ma *et al*., 2010), shifted 10nt from ath-miR4240 that was not detected in our SC and Pollen datasets.
* – Correspondent miRNA* sequence also detected in the sRNA dataset.

MicroRNAs have been previously shown to be active in pollen, as miRNA-directed cleavage products were consistently detected in pollen RNA by 5'RACE (GRANT-DOWNTON *et al.*, 2009a). In order to confirm that the novel miRNAs detected in our analysis are active in pollen, we performed modified 5'RACE in pollen total RNA as described previously (LLAVE *et al.*, 2002). Cleavage products were detected for two new miRNAs, miR5012 and miR5029, which are predicted to target *At3g60580* and *At4g02900*, respectively. The majority of RACE products cloned showed a 20-50bp shift downstream of the predicted cleavage site (Figure 5.2), which suggests transcript degradation possibly derived from miRNA-directed cleavage occurring earlier in pollen development. These observations had been reported before in pollen 5'RACE experiments for *DUO1* transcripts (GRANT-DOWNTON *et al.*, 2009a). As a control, we performed 5'RACE for *At5g55020*, *MYB120*, a well-known target of miR159, and surprisingly this time we found the cloned products shifted upstream of the predicted cleavage site by 10nt. Interestingly, *At3g60580*, *At4g02900* and *MYB120* are highly expressed in microspores and their levels remain consistently high throughout pollen development, but transcripts were not detected in the sperm cells of mature pollen. This suggests that at least some miRNAs might be active only in the germline upon PMI, to degrade highly abundant transcripts previously accumulated at the microspore stage.

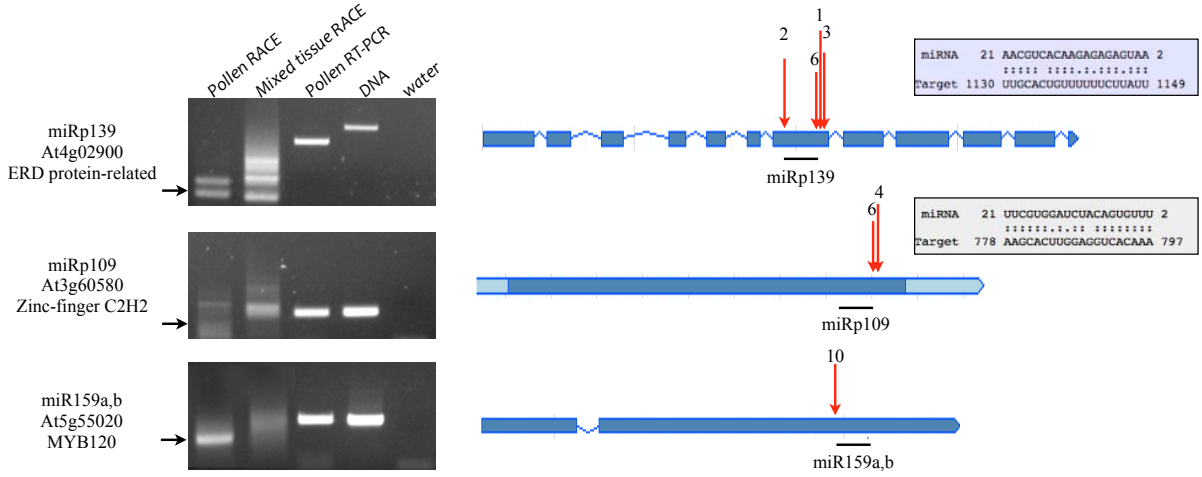


Figure 5.2 - MicroRNA-directed cleavage of target transcripts in pollen. 5' RACE PCR of the gene *At4g02900* demonstrates that this transcript is cleaved in pollen. Sequencing of 5' RACE PCR products show that miR5029 directs the cleavage of *At4g02900* in pollen and in a mixture of sporophytic tissues, which contains RNA from root, leaf, meristem and floral tissues. The cleavage products are mainly upstream of the miRNA target site (black bar). Cleavage of the miR5013 target gene *At3g60580* is also detected in both pollen and the mixed tissue sample, again occurring downstream of the miRNA target site, as well as 3' to this site. For the miR159 target gene *MYB120* (*At5g55020*), cleavage is detected in pollen, but not detected in the mixed tissue sample. Cleavage was consistently detected in the 5' of miRNA target site. Exons are indicated by boxes and introns are lines. The mRNA complementary site is shown below each gene model as a black bar. Each red arrow denotes a site of cleavage detected by 5' RACE PCR, with the number of times that cleavage product was sequenced indicated on top.

miRCat could also detect isoforms of known miRNA families with nucleotide extensions at the 3' terminus. For example, miR733 is annotated as a 21-nt miRNA and seems to be low abundant in both total pollen and sperm cells. However, a 22-nt miR773 with a C extension at the 3' terminus is more abundant (Table 5.1), suggesting in this case activation of the recently discovered mechanism for triggering siRNA production from target transcripts (CHEN *et al.*, 2010; CUPERUS *et al.*, 2010). In order to identify other putative siRNA triggers

in our datasets, we analyzed the abundance of all possible isoforms of known miRNA families from 19 to 24-nt in sequence length. Among the 22-nt miRNAs in *Arabidopsis* discussed previously (Chen *et al.*, 2010; Cuperus *et al.*, 2010), miR173, miR393, miR447, miR771, miR773, miR825*, miR167d and miR828 were detected in both pollen and sperm cell datasets (Figure 5.3B). Interestingly, miR773 targets the DNA methyltransferase *MET2* and is more abundant in pollen, while miR771 is enriched in sperm cells and was confirmed to function as siRNA trigger (CHEN *et al.*, 2010), but the targets remain unknown. We further analyzed the abundances of other miRNA families that could potentially function as siRNA triggers (i.e. more abundant 22-nt isoform) such as miR776, miR777, miR840, miR845b, miR848a, miR852, miR853 and miR2936 (Figure 5.3B). miRNA845 is highly enriched in pollen and sperm cells, and very low abundant in inflorescence tissue, suggesting preferential expression in the male gametophyte. This is a plant-specific miRNA family (BARAKAT *et al.*, 2007), for which a biological function remains obscure.

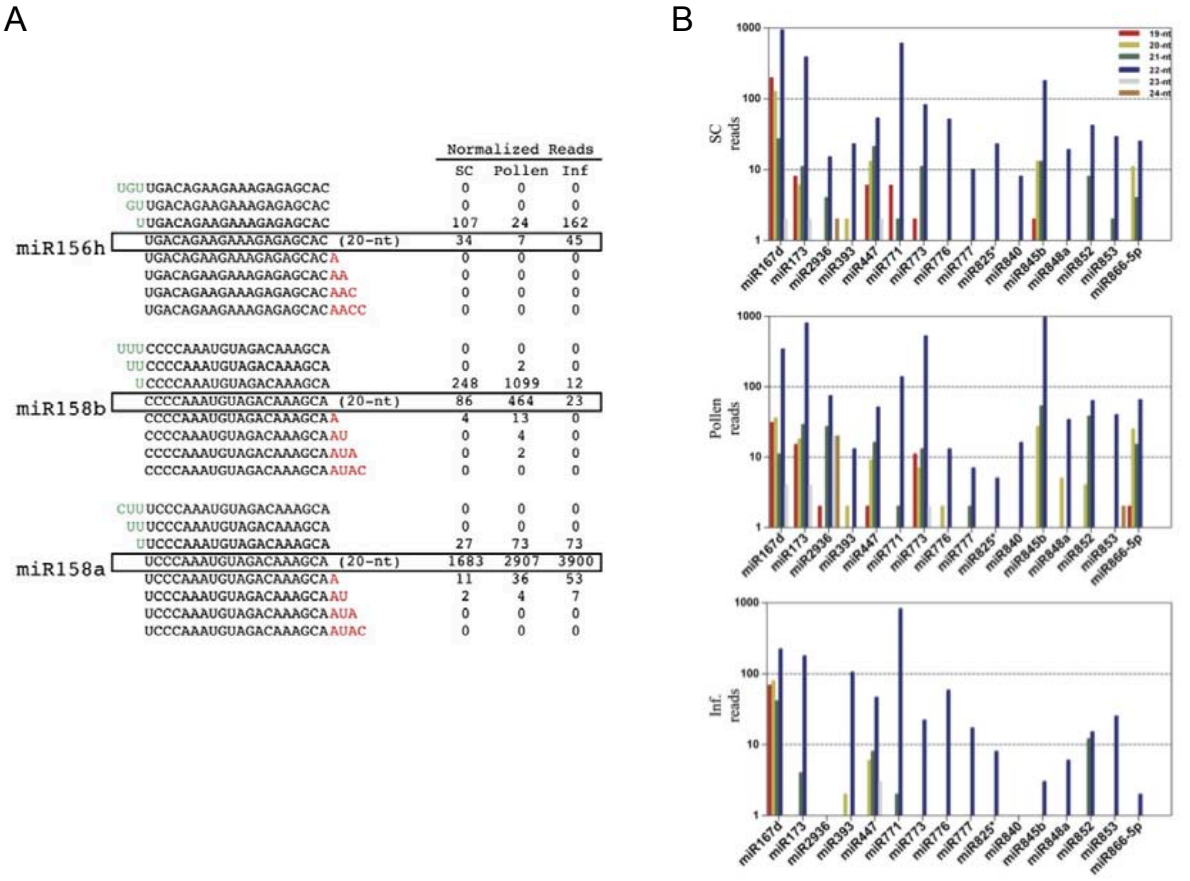


Figure 5.3 - Variations in sequence length of known miRNA families. (A) MicroRNAs 156h and 158b have isoforms of 21nt length with nucleotide extensions in the 5' terminus, for which the number of reads in SCs and pollen is higher than the annotated 20nt isoforms. miR158a reads are presented to exemplify that these variations do not accumulate significantly for the majority of other miRNAs. **(B)** 22nt miRNAs that may function as siRNA triggers are differentially accumulated in SC, pollen, and inflorescence tissues.

Novel small RNA pathways during germ cell differentiation and specification

Many of the genes involved in small RNA pathways in Arabidopsis are either not expressed or depleted in sperm cells and pollen. However, expression of particular genes involved in small RNA activity and DNA methylation such as *AGO9*, *MET1* and *DDM1*, as well as *AGO5*, is highly enriched in sperm cells (BORGES *et al.*, 2008; SLOTKIN *et al.*, 2009), suggesting that distinct genetic and epigenetic mechanisms might be established preferentially

in the germline. *AGO1* and to some extent *AGO10*, are the main regulators of miRNA-directed target cleavage in Arabidopsis (BRODERSEN *et al.*, 2008; VAUCHERET *et al.*, 2004), but both seem to be down-regulated in sperm cells (Figure 5.4). While AGO9 has been recently implicated in siRNA-mediated transposon inactivation in the female gametophyte (OLMEDO-MONFIL *et al.*, 2010), AGO5 is a closely related homolog of AGO1 and AGO10 (Figure 5.4) and could thus be involved in miRNA activity in the gametes.

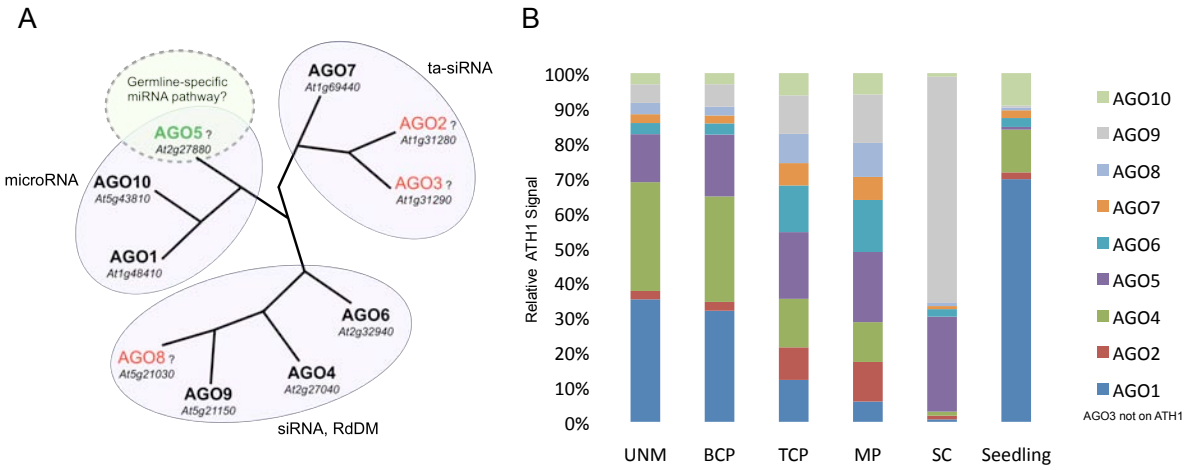


Figure 5.4 - Argonaute protein family in Arabidopsis. (A) Phylogenetic tree illustrating the 10 Argonaute proteins in Arabidopsis, subdivided into the three main functional classes based on sequence homology: miRNA-guided slicing and translational repression of target transcripts, trans-acting siRNA (ta-siRNA) activity, and chromatin remodelling by siRNA-directed DNA methylation (RdDM). AGO2, AGO3, AGO8 (red), and AGO5 (green) have unknown function. AGO5 is a close relative of AGO1 and AGO10, and could be involved in a novel miRNA pathway in the germline (see Figure 5.5). (Adapted from YIGIT *et al.*, 2006). **(B)** Relative expression of Arabidopsis Argonaute proteins represented on ATH1 array (100% reflects combined signal intensity only from probes representing Argonaute transcripts), throughout pollen development (HONYS AND TWELL, 2004, BORGES *et al.*, 2008). Expression of most Argonaute proteins is down-regulated during pollen development, while *AGO5* and *AGO9* are predominantly expressed in the sperm cells.

Analyzing relative expression of Arabidopsis Argonaute proteins throughout pollen development and in the sperm cells, reveals that *AGO1*, *AGO4*, and to some extent *AGO10*, lose the prominence as compared with their enriched expression in sporophytic tissues (Figure 5.4). This is confirmed by microarray analyses of gene expression in pollen, which shows that most Argonautes are depleted in mature pollen (PINA *et al.*, 2005), and only AGO5 and AGO9 are enriched in the sperm cells (BORGES *et al*., 2008). These evidences clearly suggest that the canonical small RNA pathways are down-regulated during microgametogenesis.

Functional characterization of Arabidopsis *ARGONAUTE 5*

As predicted by the sperm transcriptome data, we were able to confirm by expression of a translational fusion of AGO5 with GFP that in mature pollen and growing pollen tubes, AGO5 is preferentially accumulated in the sperm cell cytoplasm (Figure 5.5A-C). During pollen development, AGO5 seems to follow an interesting pattern, as it localizes mainly in the vegetative nucleus at the bicellular pollen stage, shifting to the sperm cell cytoplasm after the second pollen mitosis (Figure 5.5D). The closest homolog in rice (*MEL1*) was shown to be germline-specific and involved in progression through meiosis (NONOMURA *et al.*, 2007), but the biological function of both AtAGO5 and OsMEL1 and their role in the small RNAs pathways remain to be elucidated.

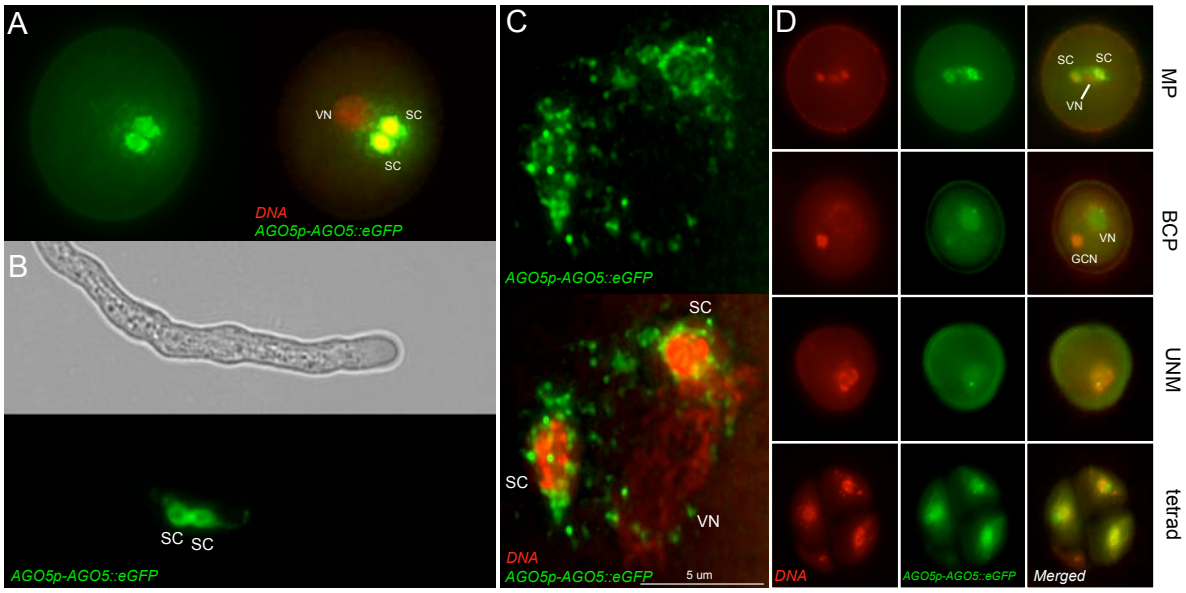
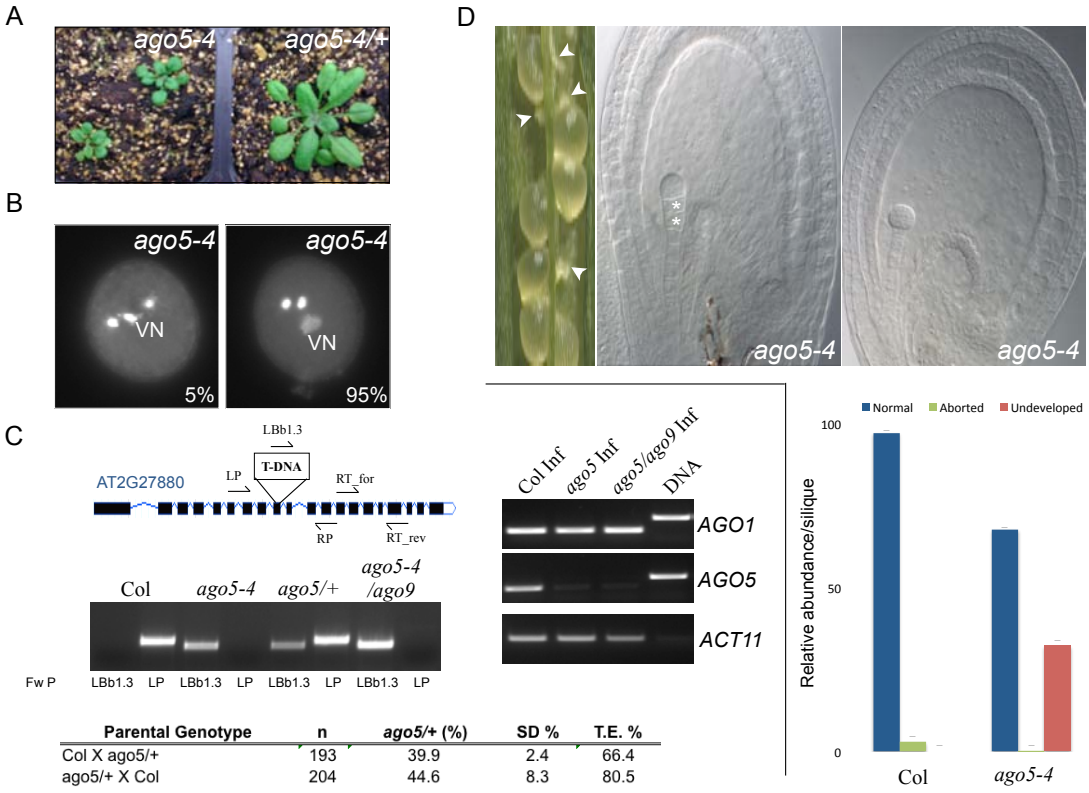


Figure 5.5 - *AGO5* expression in Arabidopsis pollen. (A) Transgene expression of AGO5 protein using the native promoter region (1000 bp upstream of the 5'-untranslated region) and the genomic coding sequence translationally fused to eGFP. *AGO5p-AGO5::eGFP* expression in mature pollen localizes preferentially in the sperm cell (SC) cytoplasm, remaining during pollen tube growth **(B)**. **(C)** A magnification of the male germ unit shows that AGO5–GFP localizes in the SC cytoplasm and not in the nucleus, extending through the cytoplasmic connection that links the SCs with the vegetative nucleus (VN). DAPI-stained DNA shows SCs and VNs in A and C. **(D)** AGO5–GFP expression during pollen development. GFP signal was detected in the nucleus of uninucleate microspores (UNM), and in bicellular pollen (BCP) it remains in both the generative cell (GCN) and sperm cell nuclei (SCN). In mature pollen (MP) AGO5-GFP localizes preferentially in the sperm cell (SC) cytoplasm.

In order to understand the role of *AGO5* during pollen development and germline specification, we characterized a mutant allele of *AGO5* (Salk_050483), from now on referred as *ago5-4*. Independent alleles were previously analyzed, such as *ago5-1* (KATIYAR-AGARWAL *et al.*, 2007), *ago5-2* (TAKEDA *et al.*, 2008) and *ago5-3* (D. Baulcombe, unpublished), but phenotypic analyses were not reported. We verified that seedling development is dramatically delayed in *ago5-4* homozygous mutants as compared to the heterozygous mutant (Figure 5.6A), and rosette leaves are dwarfed and dark green.

Heterozygous seedlings were phenotypically indistinguishable from wild-type plants. The high expression levels of AGO5 in the male gametes prompted us to analyze whether male gametophyte development is also compromised in *ago5-4*. Mature pollen stained with 4',6-diamidino-2-phenylindole (DAPI) revealed that the majority was phenotypically indistinguishable from wild-type pollen, i.e. tricellular (two condensed sperm nuclei and one vegetative nucleus with decondensed chromatin), however, there were a few visible (~5%) exceptions (~5%) showing a VN with condensed chromatin resembling a sperm cell nucleus (Figure 5.6A). In addition, *ago5-4* plants exhibit problems with set seed under normal growth conditions. Dissecting mature siliques showed approximately 30% of randomly distributed undeveloped ovules (Figure 5.6C). By analyzing the 70% of seeds that could develop, we noticed also a few cases of embryos showing patterning defects (Figure 5.6C). The seed set phenotype suggested that fertilization, early seed development or pollen tube targeting is impaired, but to clarify if this phenotype is correlated with a problem in the male or female gametophyte, or both, we performed reciprocal crosses between wild type and heterozygous *ago5-4/+* plants. These analyses revealed that the mutant allele is poorly transmitted through both gametophytes (Figure 5.6B).



Parental Genotype	n	ago5/+ (%)	SD %	T.E. %
Col X ago5/+	193	39.9	2.4	66.4
ago5/+ X Col	204	44.6	8.3	80.5

Figure 5.6 – Functional characterization of AGO5. (A) Homozygous plants with a mutant allele of *AGO5* are phenotypically characterized by a slow growth and dwarf leaves. Heterozygous *ago5-4/+* plants are phenotypically undistinguishable from wild-type. **(B)** Mature pollen is normal as compared to wild-type, but DAPI staining revealed rare events of vegetative nucleus (VN) condensation. **(C)** Genotyping and RT-PCR of *ago5-4* mutants confirmed to be a strong loss-of-function mutation, although a faint amplification product was sometimes detectable, possible derived from aberrant transcription of the T-DNA element. Reciprocal crosses with wild-type showed that the mutant allele is poorly transmitted through both male and female gametophytes. **(D)** Fertilization and seed development is compromised in *ago5-4*, showing 30% of undeveloped ovules (arrowheads). Apparently normal seeds also appear to have some embryo patterning defects (*).

Considering that AGO1 is down-regulated in sperm cells (Figure 5.4B), we hypothesized that its close homolog AGO5 could participate in miRNA-directed post-transcriptional gene silencing in the male germline. In order to test this hypothesis, we

performed microarray experiments using total RNA isolated from wild-type and *ago5-4* mutant pollen. We found that out of the 168 known miRNA targets represented on the ATH1 array, only one is slightly up-regulated in *ago5-4* pollen, thus indicating that AGO5 is either not involved in miRNA-directed gene silencing during pollen development, or its possible function in the sperm cells was not detectable by analyzing gene expression in total pollen. However, we found that genes activated by the germline-specific transcription factor DUO1, as well as transcripts normally abundant in wild-type sperm cells, are down-regulated in *ago5-4* pollen. This is the case for *MGH3*, *OPT8*, *TIP5;1* and *GEX2* (Table S5.4). As *DUO1* is not represented on ATH1, we confirmed by RT-PCR that *DUO1* transcripts are also depleted in the mutant (Figure 5.7A). Surprisingly, *DUO1* expression in the double *ago5/9* mutant is restored to wild-type levels, which correlates with the fact that *ago5/9* mutant also does not show the sporophytic defects observed in *ago5-4*. This result suggested an AGO5 and AGO9 interplay during pollen development. As *DUO1* is a miR159 target, our result might indicate AGO5 having a similar function as AGO10, that is involved in sequestering miR166 to prevent its incorporation into the functional AGO1 silencing complex during apical meristem formation (ZHU *et al.*, 2011). As such, a possible mechanism would be that loss of AGO5 in the sperm cells increases the incorporation of miR159 into AGO9, which in contrast to AGO5, is able to down-regulate *DUO1* post-transcriptionally.

Alternatively, the transcriptome profile in the mutant could be the result of a significant delay in pollen development. Interestingly, by analyzing the list of the 100 most up- and down-regulated genes in *ago5-4* pollen, we could observe that up-regulated genes are normally low abundant in wild-type sperm cells, while down-regulated genes in the mutant tend to be more expressed in the sperm cells than in pollen (Figure 5.7B). This strongly suggests that transcriptome changes between *ago5-4* and wild-type pollen might be reflecting

differential gene expression in the gametes, and not in the companion vegetative cell. We analyzed further the expression levels for these genes throughout pollen development, and here we found a striking tendency for genes down-regulated in the mutant, to be up-regulated during pollen development, and vice-versa (Figure 5.7C).

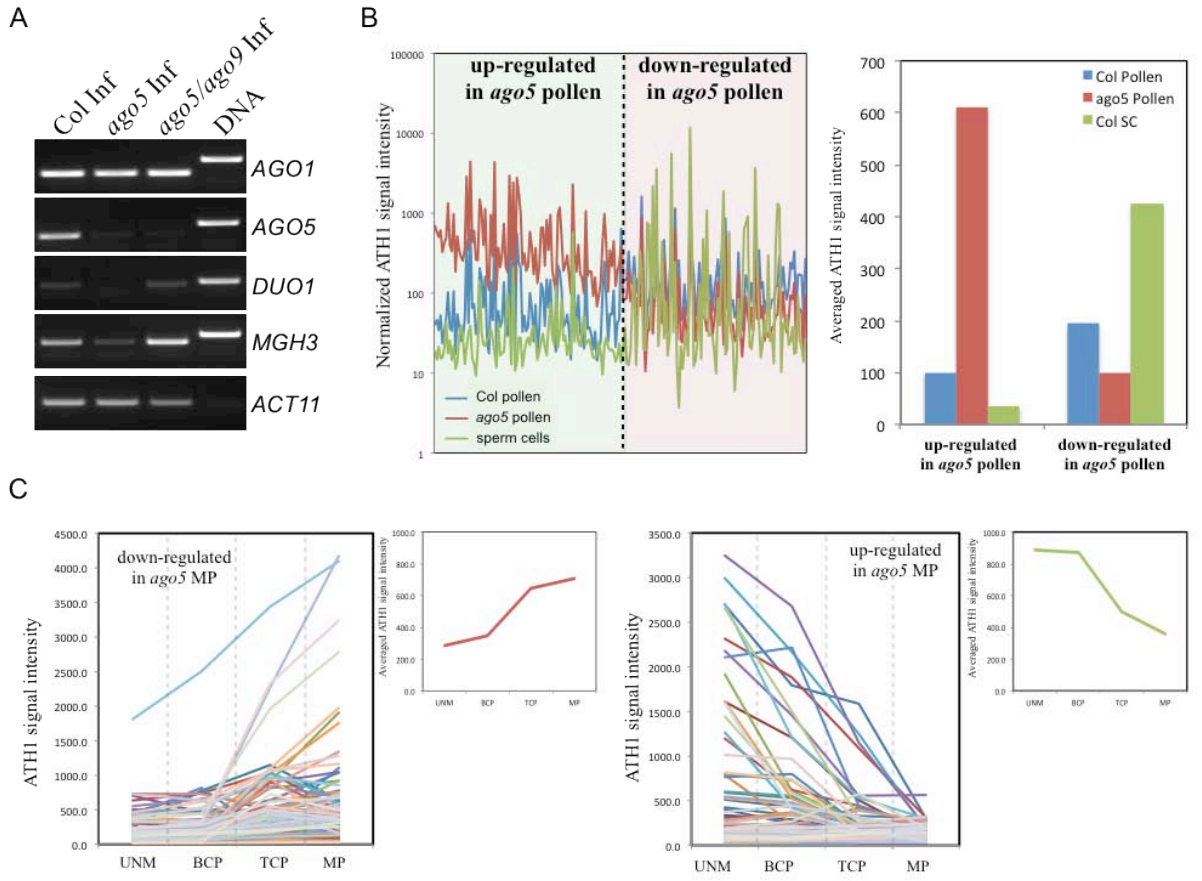


Figure 5.7 - Expression analysis throughout pollen development of genes differentially regulated in *ago5-4* pollen. (A) Confirmatory RT-PCR analysis of wild-type, *ago5-4* and *ago5/9* inflorescence tissue, showing that *AGO5* is strongly depleted in the mutant, as well as the germline-specific genes *DUO1* and *MGH3*. Expression of *DUO1* and *MGH3* is restored in the double mutant *ago5/9*. *ACT11* was used as amplification control. **(B)** Expression of the 100 most up- and down-regulated genes in *ago5-4* pollen coincide with striking differences of gene expression in the sperm cells, as up-regulated genes in the mutant reflect genes that are either not expressed or depleted in the sperm cells, and down-regulated transcripts in the mutant are normally highly abundant in sperm. **(C)** These groups of genes also reflect inverse trends along pollen development, as down-regulated genes in the mutant appear to be up-regulated during microsporogenesis, and vice-versa. Small adjacent graphs show averaged

gene expression as detected from the ATH1 array throughout pollen development (HONYS AND TWELL, 2004) of all genes statistically down- (left) and up-regulated (right) *in ago5-4* pollen.

These results correlate with the idea of a developmental delay in the mutant, but it is intriguing that it affects only a restricted number of genes and not the whole pollen transcriptome. As *DUO1* is down-regulated, maybe other key transcription factors involved in sperm cell specification are also depleted, which would explain the group of down-regulated genes. Interestingly, we noticed that *ago5-4* up-regulated genes are usually flanked by TEs and repeats targeted by RNA-directed DNA Methylation (RdDM) (Figure 5.8). If AGO5 is involved in RdDM and transcriptional gene silencing by targeting these regions, it would explain why the adjacent genes are up-regulated in the mutant. To test this hypothesis, we performed bisulfite sequencing on DNA isolated from wild-type and mutant pollen, for *At3g47470* and *At3g17120*, which are up-regulated in *ago5-4* pollen (Table S5.4). These preliminary analyses could not detect any significant change in DNA methylation in the mutant compared with wild-type pollen or leaf tissue (Figure 5.8), although it might require analyzing DNA from purified SC and VN, instead of total pollen.

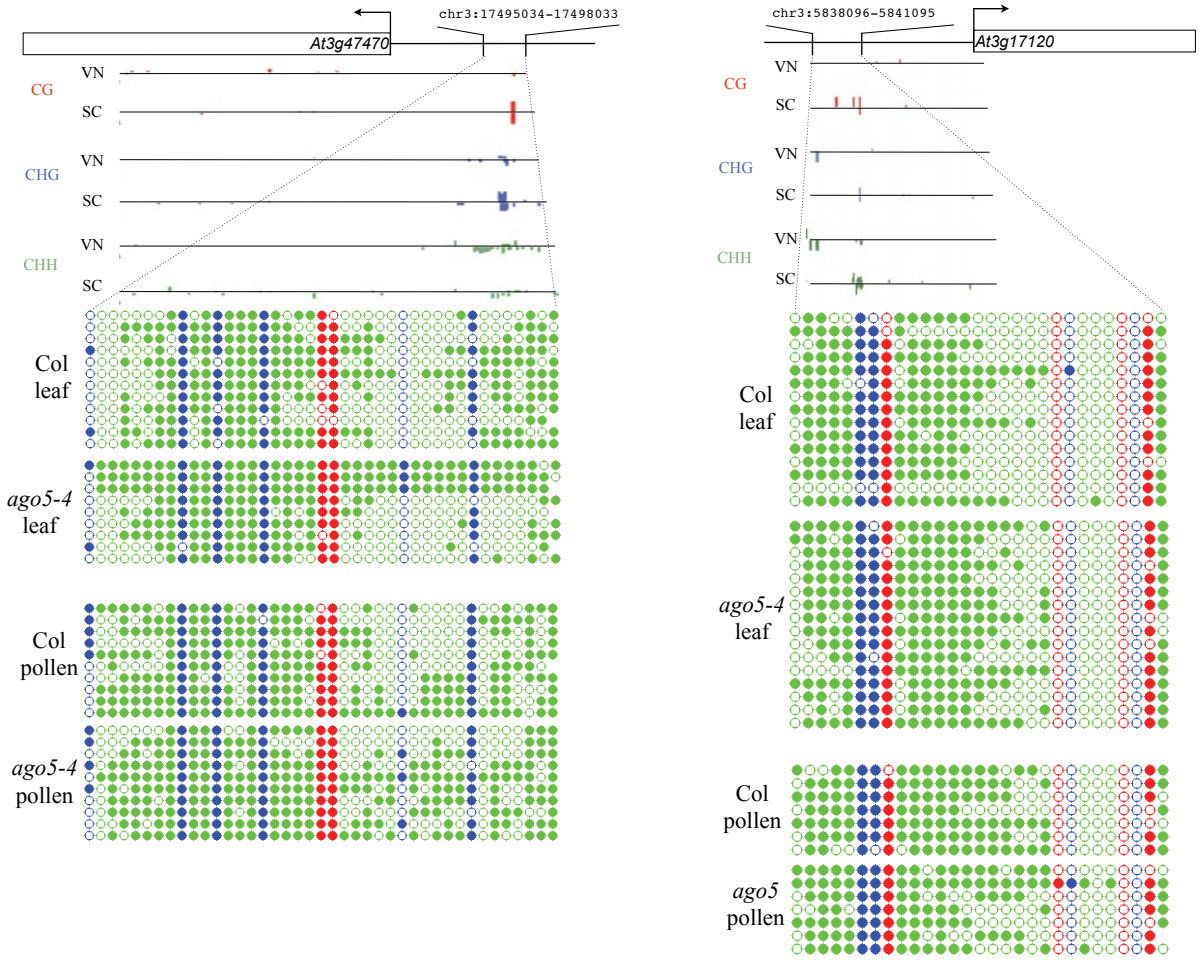


Figure 5.8 – Bisulfite sequencing of adjacent genomic loci of genes up-regulated in *ago5-4* pollen. *At3g47470* and *At3g17120* were amplified from bisulfite converted DNA and sequenced from leaf tissue and mature pollen from wild-type Col and *ago5-4* mutant. The location of the coding region is shown as a rectangle, whereas the 5' DNA is shown as a line. Genome browser snapshots show CG, CHG and CHH methylation profiling of wild-type sperm cell (SC) and vegetative nucleus (VN), according to the results presented in Chapter 4. Clones were sequenced for the sense strand of each sample. All cytosines in the DNA sequence are shown as circles. The methylation state of each cytosine is shown as an open circle (not methylated) or filled circle (methylated). The sequence context of the cytosine is denoted by the color of the circle (red, CG; blue, CHG; green, CHH).

Depleting microRNA activity during male germline specification

Viral suppressor of gene silencing P1/Hc-Pro

Loss-of-function alleles of *AGO5* and *AGO9* suggest that miRNAs are actively modulating the sperm cell transcriptome, but it seems that the absence of miRNA activity does not compromise male germ cell specification and fertilization. A recent study characterized a null allele of *DCL1* (*dcl1-5*), showing that certain miRNA families are essential to prevent precocious gene expression in the early embryo (NODINE and BARTEL, 2010). Importantly, this work showed that heterozygous plants segregate under normal mendelian ratios producing 25% of homozygous embryos that abort early in development. The *dcl1-5* allele is normally transmitted through both gametophytes, thus suggesting that miRNA activity is not essential in the gametophytes. However, it was not shown whether both *dcl1-5* male and female gametophytes lose miRNAs, so it is still reasonable to speculate that *dcl1-5* gametophytes could inherit miRNAs from their precursor pre-meiotic mother cells, or there is a DCL1-independent pathway for miRNA production acting in the Arabidopsis gametophytes.

In a first attempt to deplete the miRNA activity specifically in sperm cells, we used the viral suppressor of gene silencing P1:Hc-Pro, which was shown to interfere with the siRNA- and miRNA-directed gene silencing pathways (CHAPMAN *et al.*, 2004; DUNOYER *et al.*, 2004). In order to use this system in a germline-specific manner, expression was driven by the *MGH3* promoter. Several transgenic lines were isolated with different expression patterns of the transgene (Figure 5.9A), which not always correlated with the phenotype. We observed that in two independent lines, a significant proportion of pollen grains were still at bicellular stage at anthesis, indicating either failure of pollen mitosis II or developmental delay (A2 and

C3, Figure 5.9B). In contrast, one transgenic line showed normal pollen, but fertilization was compromised showing roughly 25% of undeveloped ovules (Figure 5.9C).

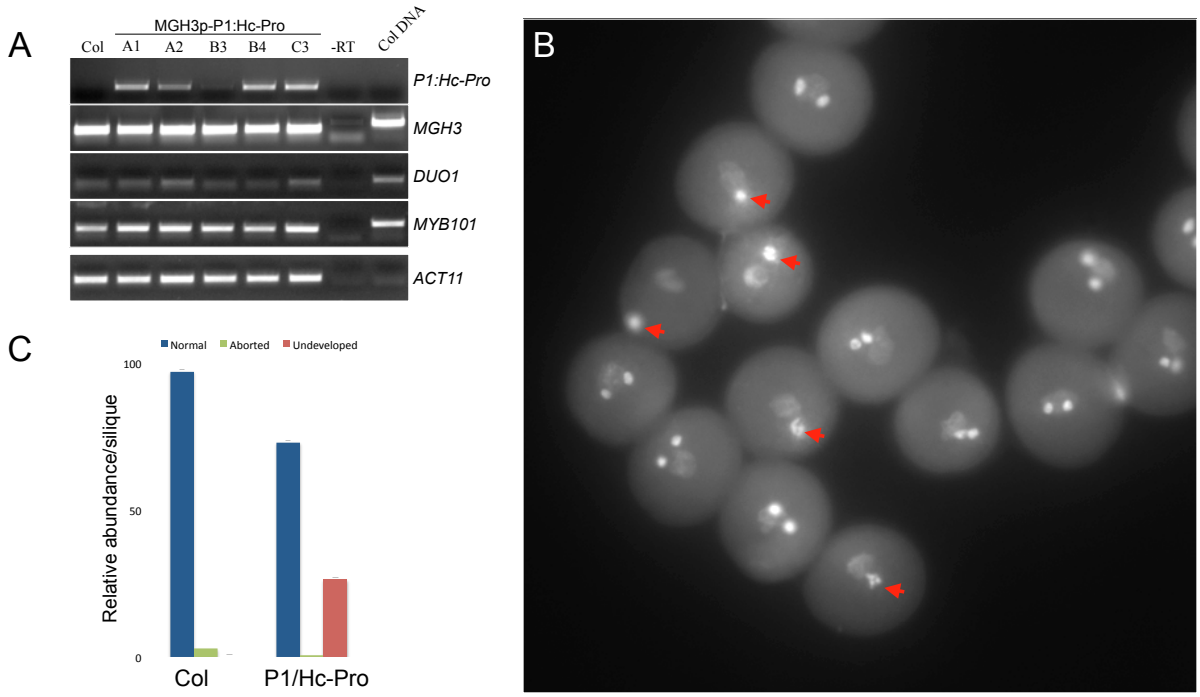


Figure 5.9 - Transgenic Arabidopsis plants expressing the viral suppressor of gene silencing P1:Hc-Pro. (A) RT-PCR analysis to confirm expression levels of the transgene, P1:HcPro, in inflorescence tissue of five independent transgenic lines. Differential gene expression was analyzed for well-characterized miRNA targets such as *DUO1* (sperm cell-expressed) and *MYB101* (vegetative cell-expressed), as well as for the DUO1-activated sperm cell gene *MGH3*. Significant variation was observed for the transgene expression, but was not negatively correlated with the expression of the two miRNA targets or the DUO1-activated *MGH3*, as initially expected. ACT11 was used as control for linear PCR amplification in every cDNA library. ACT11 primers span introns over exon-exon junction, and as such cannot amplify genomic DNA. **(B)** Imaging DAPI-stained mature pollen grains of transgenic line C3, as represented in the panel A, shows a significant number of pollen that fails to undergo pollen mitosis II, and are still bicellular at this point. **(C)** For a different transgenic line identified as A1 in panel A, mature pollen is tricellular and seem to develop normally, but fertilization seems to be compromised as observed by the 25% of undeveloped ovules.

Target mimicry

A particular case of endogenous regulation of miRNA activity has been shown to depend on non-coding transcripts capable of sequestering specific miRNA families, thus routing them out of their bona-fide pathway for target cleavage (FRANCO-ZORRILLA *et al.*, 2007). More recently, a collection of mimicry constructs targeting several miRNA families in Arabidopsis was developed (TODESCO *et al.*, 2010), and we made use of this mechanism to try to interfere with the activity of miRNAs expressed in sperm cells. We selected candidates based on their higher enrichment in the sRNA sequencing data for sperm cells compared with that of mature pollen. We also hypothesized that some miRNAs accumulated in SCs could be delivered at fertilization to trigger and/or control gene expression during early zygotic patterning, therefore we also considered expression of miRNA targets in the female gametes (WUEST *et al.*, 2010) to select candidate miRNA families (Table 2). This idea comes from the observation that transcripts accumulated exclusively in sperm cells can be important to control early zygotic development (BAYER *et al.*, 2009) (see Chapter 1).

The mimicry constructs were expressed under the control of the *MGH3* promoter, in order to deplete miRNA activity specifically during germline development. Surprisingly, for all transgenic lines of each independent mimicry construct analyzed, we could not detect any consistent phenotype indicating impaired germline development, fertilization or seed development. However, it was equally not possible to detect a strong expression of mimicry transgenes in mature pollen, suggesting that these transcripts may not be stable in the sperm cells nuclei.

Table 5.2 - Summary of miRNA families targeted for depletion by mimicry constructs in the sperm cells. These specific families were chosen considering their enrichment in sperm cells, as compared with their detection levels in total pollen. The functional category of the main targets of each miRNA family is represented, as well as their expression in both the male (MG) and female (FG) gametes (BORGES *et al*., 2008 and WUEST *et al*., 2010). (+) Present, (-) Absent, (?) Not on ATH1.

MicroRNA	Average enrichment sperm/pollen	Target functional class	Expr. target FG	Expr. target MG
miR169a-n	14.6	transcription factor	+	+
miR394a,b	9.5	F-box family protein	+	-
miR399e	6.3	ubiquitin-protein ligase	+	+
miR319a,b,c	6	transcription factor	+	-
miR159a,b,c	5.3	MYB TFs	+	-
miR164a,b,c	4.9	transcription factor	+	-
miR163	4.8	methyltransferase	+	-
miR167a,b,c	3.9	ARF TFs	+	+
miR391/390a,b	3	trans-acting siRNA	?	?
miR393a,b	2.5	auxin binding / ubiquitin-protein ligase	+	-
miR172b	2.4	transcription factor	+	+

3. CONCLUSIONS

Post-transcriptional gene silencing involves miRNA activity to eliminate transcripts of previous developmental processes, while modulating expression of active genes. In the recent years, small RNA sequencing technologies allowed a robust temporal and spatial profiling of miRNAs processed in several species, tissues and cell types, leading to the discovery of an ancient and widespread mechanism to control gene expression. Consequently, we can assume that we now know the basis of miRNA processing and activity in eukaryotes; however, we are just taking the first steps towards understanding its origins and evolution.

Post-meiotic pollen development is an exciting biological process in plants for its structural simplicity and fast transitions over two mitotic divisions, which are coupled with germline differentiation and specification. It is a unique system to analyze specialized small RNA pathways and small RNA trafficking with cellular resolution. We found that miRNAs are abundant and diverse in Arabidopsis sperm cells, and that after germ cell differentiation the sperm cells and neighboring vegetative cell accumulate different miRNAs, which correlate with different cell fate and gene expression profiles. However, there are still several open questions. The observation that AGO5 is likely to be part of a unique silencing complex established in the male germline, supports the idea that distinct and rather unknown cellular pathways might exist. Our results suggest that AGO5 could be both involved in sequestering miRNAs in pollen, thus prohibiting post-transcriptional gene silencing. In addition, it could also be part of the RdDM pathway targeting adjacent loci of genes whose expression is down-regulated during pollen development, but additional experiments are require to confirm these hypotheses.

We failed in identifying miRNAs whose activity modulates gene expression during

germline specification. Target mimicry became a well-established approach to study miRNA function in the Arabidopsis sporophyte, by mimicking phenotypes observed with null mutations in miRNA genes or plants expressing cleavage-resistant targets (TODESCO *et al.*, 2010), but we could not confirm that this method is functional in sperm cells. Instead, further efforts should focus on mutations in genes involved in miRNA processing in pollen and sperm cells, as a way of obtaining pollen depleted of miRNAs. However, such mutations might additionally compromise siRNA production. Expression of viral suppressors during sperm development provided some evidence showing a prominent role for small RNA activity during sperm cell specification and fertilization, but it could not discriminate between the miRNA and the siRNA pathways. One possible experiment is studying mutations that could uncouple the miRNA and siRNA pathways. This idea was explored in mouse oocytes, using a mutant allele of Dgcr8, a protein specifically involved in miRNA processing (SUH *et al.*, 2010). This work could show that miRNA function is suppressed during oocyte maturation, and that the defects initially observed in *Dicer* mutants where derived from the lack of endo-siRNAs that are also processed by Dicer (SUH *et al.*, 2010).

4. MATERIALS AND METHODS

Plant Material and Growth Conditions

Plants were grown on soil for 8 weeks in short-day conditions (8 h light at 21ºC-23ºC) and then transferred to long-day conditions (16 h light) to induce flowering. For microarray experiments, Arabidopsis wild-type Col-0 background and *ago5-4* mutant line (SALK_050483) were used for RNA isolation from pollen and inflorescence tissue. Pollen was isolated by vortexing open flowers in pollen extraction buffer (10 mM CaCl2, 1 mM KCl, 2 mM MES, pH 6.5), then filtered through Miracloth and concentrated by centrifugation (800g, 5min). All sample types were processed in triplicate from three individual RNA extractions with Trizol.

Small RNA sequencing and analysis

Small RNAs of 19-28nt were size selected by denaturing 15% PAGE and cloned as described in Slotkin et al. (2009). Read numbers of sequences matching to known Arabidopsis miRNAs in pollen and sperm are publicly available in the SBS database (NAKANO *et al.*, 2006, http://mpss.udel.edu/at_sbs). Potentially new microRNAs were found using the miRCat tool available online within the UEA plant sRNA toolkit (MOXON *et al.*, 2008), and manually filtered following submission information established by miRBase (http://www.mirbase.org).

Transgene construction and microscopy analysis

The genomic sequence including -1000bp upstream of *AGO5* (*At2g27880*), plus the coding region without stop codon, was amplified from *Arabidopsis thaliana* Col-0 genomic

DNA purified by phenol:chloroform. The amplicon containing AttB recombination sites was subcloned into pDONR201 (Invitrogen), and subsequently cloned into a modified promoter-less pB7FWG2 (Karimi *et al.*, 2002) by gateway technology (Invitrogen) (cloned by Patricia Pereira, IGC). The MGH3 promoter was used to express P1:HcPro and target mimicries. The genomic sequence 1222bp upstream of *MGH3* coding sequence was cloned into a gateway destination vector pK7WG2D within HindIII and SpeI unique restriction sites. P1:HcPro CDS was amplified from pRTL2 plasmid (KASSCHAU *et al.*, 2003) (gently provided by Cláudia Martinho and Elena Baena Gonzalez), and the target mimicries were constructed by Ignacio Rubio-Somoza (MPI Tübingen, Germany) as previously described (TODESCO *et al.*, 2010). These were subsequently cloned into an entry vector, and recombined into the modified MGH3p-pK7WG2D by gateway. Positive clones confirmed by sequencing were used to transform Agrobacterium strain GV3101::pMP90RK. Positive transformants were selected with correspondent antibiotics, and used to transform Arabidopsis Col-0 background by floral dipping (CLOUGH AND BENT, 1998). Primers used are listed in Table S5.5. Figures 5.5A and C, and Movie S5.2 were obtained by mounting DAPI-stained pollen grains on a glass slide, and imaging with a Deltavision microscope (Applied Precision). To image AGO5-GFP signal, pollen grains were exposed during 3.000ms of fluorescent light. Image was processed using SoftWorx (DV, Applied Precision). Time-lapse imaging of the growing pollen tube presented in Figure 5.5B was obtained using a Zeiss Axio Imager microscope, and processed in Micro-Manager1.3 (Vale Lab, UCSF). Pollen tube growth in liquid medium was performed under the conditions described elsewhere (BOAVIDA AND MCCORMICK, 2007) during 3,5 hour at room temperature.

GeneChip hybridizations

Total RNA from pollen and inflorescence tissue of wild-type and *ago5-4* mutant

plants was processed in triplicate for use on Affymetrix (Santa Clara, CA, USA) GeneChip Arabidopsis ATH1 Genome Arrays, according to the manufacturer's GeneChip 3' IVT Express kit user manual. Briefly, 100 ng of total RNA containing spiked-in poly-A+ RNA controls was used in a reverse transcription reaction (GeneChip 3' IVT Express Kit; Affymetrix, Santa Clara, CA, USA) to generate first-strand cDNA. After second- strand synthesis, double-stranded cDNA was used in a 16 h *in vitro* transcription (IVT) reaction to generate aRNA (GeneChip 3' IVT Express Kit; Affymetrix, Santa Clara, CA, USA). Size distribution of *in vitro* transcribed aRNA and fragmented aRNA, respectively, was assessed via an Agilent 2100 Bioanalyzer (Agilent, Böblingen, Germany), using an RNA 6000 Nano Assay. Fragmented aRNA was added to hybridization cocktail also containing hybridization controls. The mixture was hybridized on GeneChips for 16 h at 45°C. Standard post hybridization wash and double- stain protocols (FS450_0001; GeneChip HWS kit; Affymetrix, Santa Clara, CA, USA) were used on an Affymetrix GeneChip Fluidics Station 450. GeneChips were scanned on an Affymetrix GeneChip scanner 3000 7G. (This experiment was entirely performed by students attending the Genome Biology and Evolution Summer School 2011, IGC)

Data Analysis

Absent and Present calls were generated using Affymetrix Expression Console software and all subsequent analyses were conducted using dChip software as of January 2010 (http://www.dchip.org, Wong Lab, Harvard). For each sample type (Col pollen, *ago5-4* pollen, Col inflorescence, *ago5-4* inflorescence), a microarray experiment was performed with three biological replicates. However, only two replicates where considered for further analysis, as Principal Component Analysis could detect one array outlier each for *ago5-4* pollen and Col Inflorescence (data not shown). We used a sample wise normalization to the

median probe cell intensity (CEL) of all arrays, and for each array, the median CEL intensity was scaled to the median CEL intensity of the array defined as baseline, and normalized applying an Invariant Set Normalization Method (LI AND WONG, 2001). Model-based gene expression was obtained from normalized CEL intensities based on a Perfect Match-only model (LI AND HUNG WONG, 2001). Annotations for all genes represented on the Arabidopis ATH1 genome array were according to TAIR9.

5' RACE

Modified 5' RACE was performed using the GeneRacer Kit (Invitrogen) as described previously (LLAVE *et al.*, 2002), using total RNA extracted by Trizol from Col pollen and a control sample of mixed RNA that included seedling, leaf and floral tissue (Col background). To determine if un-cleaved target transcript was present in pollen, RT-PCR was performed of the same RACE cDNA using the same 3' gene-specific primer as used in the 5' RACE, but a different 5' gene-specific primer. PCR amplification products were gel-purified on column, and cloned into pGEM vector (Promega). The 5' RACE products were sequenced from the T7 promoter region to determine the precise location of the target transcript cleavage. Primers for 5' RACE PCR and RT-PCR can be found in Table S5.5.

RT-PCR

Semi-quantitative RT-PCR was performed with total RNA isolated with Trizol from inflorescence tissue (Col background). First-strand cDNA (Oligo-dT primed) was synthesized using the MLV reverse transcriptase - RNaseH minus (Promega) according to manufacturer instructions. 2ul of non-diluted cDNA was used as a template for 30x PCR reactions. Primers used are listed in Table S5.5.

Bisulfite sequencing

Genomic DNA was extracted from leaf tissue and pollen isolated from 150mg of open flowers, and purified by Phenol:Chloroform. 400ng were used for bisulfite treatment using the EZ DNA Methylation-Gold Kit (Zymo). 2ul of treated DNA were used for PCR with Expand High-Fidelity Polymerase (Roche), in a 50ul PCR reaction (10X PCR buffer + $MgCl_2$, 1.25ul 10mM dNTPs, 2.5ul 10uM primer 1, 2.5ul 10uM primer 2, 0.5ul Taq), and the following cycling condition: 94ºC, 3 min, 19x 94ºC, 30 sec; 54ºC, 30 sec; 72ºC, 45sec (each cycle decreasing 0.5ºC in the annealing temperature), 19x 94ºC, 30 sec; 44ºC, 30 sec, 72ºC, 45 sec, and 72ºC, 15min. Amplified products were run on a 2% agarose gel, purified by column, cloned into pGEM vector (Promega) and sequenced from the T7 promoter region. Primers used are listed in Table S5.5.

5. REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207-221.

Barakat, A., Wall, K., Leebens-Mack, J., Wang, Y.J., Carlson, J.E., and Depamphilis, C.W. (2007). Large-scale identification of microRNAs from a basal eudicot (Eschscholzia californica) and conservation in flowering plants. Plant J *51*, 991-1003.

Bayer, M., Nawy, T., Giglione, C., Galli, M., Meinnel, T., and Lukowitz, W. (2009). Paternal Control of Embryonic Patterning in *Arabidopsis thaliana*. Science *323*, 1485-1488.

Becker, J.D., and Feijo, J.A. (2007). How many genes are needed to make a pollen tube? Lessons from transcriptomics. Ann Botany *100*, 1117-1123.

Boavida, L.C., and McCormick, S. (2007). Temperature as a determinant factor for increased and reproducible in vitro pollen germination in *Arabidopsis thaliana*. Plant J *52*, 570-582.

Borg, M., Brownfield, L., and Twell, D. (2009). Male gametophyte development: a molecular perspective. J Exp Bot *60*, 1465-1478.

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

Breakfield, N.W., Corcoran, D.L., Petricka, J.J., Shen, J., Sae-Seaw, J., Rubio-Somoza, I., Weigel, D., Ohler, U., and Benfey, P.N. (2012). High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis. Genome Res *22*, 163-176.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y.Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. Science *320*, 1185-1190.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T., and Twell, D. (2009). A Plant Germline-Specific Integrator of Sperm Specification and Cell Cycle Progression. PloS Genet *5*, -.

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

Chapman, E.J., Prokhnevsky, A.I., Gopinath, K., Dolja, V.V., and Carrington, J.C. (2004). Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev *18*, 1179-1186.

Chen, H.M., Chen, L.T., Patel, K., Li, Y.H., Baulcombe, D.C., and Wu, S.H. (2010). 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci U S A *107*, 15269-15274.

Chen, X.M. (2004). A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science *303*, 2022-2025.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J *16*, 735-743.

Cuperus, J.T., Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Burke, R.T., Takeda, A., Sullivan, C.M., Gilbert, S.D., Montgomery, T.A., and Carrington, J.C. (2010). Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol *17*, 997-1003.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O. (2004). Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell *16*, 1235-1250.

Ebhardt, H.A., Fedynak, A., and Fahlman, R.P. (2010). Naturally occurring variations in sequence length creates microRNA isoforms that differ in argonaute effector complex specificity. Silence *1*, 12.

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet *39*, 1033-1037.

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

Grant-Downton, R., Le Trionnaire, G., Schmid, R., Rodriguez-Enriquez, J., Hafidh, S., Mehdi, S., Twell, D., and Dickinson, H. (2009b). MicroRNA and tasiRNA diversity in mature pollen of Arabidopsis thaliana. BMC Genomics *10*, 643.

Haerizadeh, F., Singh, M.B., and Bhalla, P.L. (2006). Transcriptional repression distinguishes somatic from germ cell lineages in a plant. Science *313*, 496-499.

Kasschau, K.D., Xie, Z.X., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function. Dev Cell *4*, 205-217.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007). A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev *21*, 3123-3134.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Orom, U.A.*, et al.* (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273-1286.

Li, C., and Hung Wong, W. (2001). Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol *2*, RESEARCH0032.

Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A *98*, 31-36.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C.*, et al.* (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell *133*, 128-141.

Moxon, S., Schwach, F., Dalmay, T., Maclean, D., Studholme, D.J., and Moulton, V. (2008). A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics *24*, 2252-2253.

Nakano, M., Nobuta, K., Vemaraju, K., Tej, S.S., Skogen, J.W., and Meyers, B.C. (2006). Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. Nucleic Acids Res *34*, D731-735.

Nodine, M.D., and Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev *24*, 2678-2692.

Nonomura, K., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurata, N. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell *19*, 2583-2594.

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

Palatnik, J.F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D.*, et al.* (2007). Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. Dev Cell *13*, 115-125.

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

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell *136*, 461-472.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. Curr Biol *20*, 271-277.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant Cell Physiol *49*, 493-500.

Todesco, M., Rubio-Somoza, I., Paz-Ares, J., and Weigel, D. (2010). A collection of target mimics for comprehensive analysis of microRNA function in *Arabidopsis thaliana*. PLoS Genet *6*, e1001031.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

Vaucheret, H., Vazquez, F., Crete, P., and Bartel, D.P. (2004). The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev *18*, 1187-1197.

Vazquez, F., Blevins, T., Ailhas, J., Boller, T., and Meins, F., Jr. (2008). Evolution of Arabidopsis MIR genes generates novel microRNA classes. Nucleic Acids Res *36*, 6429-6438.

Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. Cell *136*, 669-687.

Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenfuhrer, J., von Mering, C., and Grossniklaus, U. (2010). Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr Biol *20*, 506-512.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhang, X., and Jin, H. (2010). Multiple distinct small RNAs originate from the same microRNA precursors. Genome Biol *11*, R81.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.H., Liou, L.W., Barefoot, A., Dickman, M., and Zhang, X. (2011). Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell *145*, 242-256.

CHAPTER 6

Concluding Remarks

1. CONCLUDING REMARKS

High-throughput genomics were successfully used to shed light on the genetic and epigenetic aspects of male gametophyte development and sperm cell specification in Arabidopsis. I believe that the results obtained and presented in this thesis are forming a solid foundation for a faster advance in the fields of sexual reproduction and epigenetic reprogramming in plants, paving the way for further experimental hypothesis.

The study of the two differentiated cell types in mature Arabidopsis pollen has been hindered throughout the years, as there were no reliable methods available to isolate each nucleus in sufficient amounts suitable for molecular genetic studies. Therefore, our FACS methods constitute a major advance into this direction (Chapter 2). These methods provided high sorting rates and purity, which we considered essential for comprehensive downstream applications, such as the microarray-based transcriptional profile of Arabidopsis sperm cells (Chapter 3). Transcriptional activation in Arabidopsis sperm cells is at least partially controlled by the transcription factor DUO1, which is capable of activating many sperm cell-expressed genes when ectopically expressed in sporophytic tissues (BORG *et al.*, 2011). However, understanding the real impact of DUO1 in shaping the sperm cell transcriptome requires that further studies analyze mutant *duo1* pollen instead of studying DUO1 activity ectopically. It is possible that the activity of certain transcription factors depends on the cellular context where they are expressed, as specific chromatin factors may act as barriers for genetic reprogramming (TURSUN *et al.*, 2011). The methods developed for sperm cell sorting may now be applied to the isolation of *duo1* sperm.

Genes involved in small RNA silencing and RNA-directed DNA methylation (RdDM) are enriched in the sperm cells. In the DNA methylation pathway, genes like *MET1* and *DDM1* are germline-specific in mature pollen, which is consistent with the transient reactivation of transposable elements in wild-type mature pollen that occurs specifically in the vegetative nucleus (SLOTKIN *et al.*, 2009). As in *ddm1* mutant cells, 21nt siRNAs derived from activation of *Athila* retrotransposons are also produced in pollen, but the function of these epigenetically activated siRNAs is not yet understood. The idea that they can move from the vegetative nucleus to the sperm cells (SLOTKIN *et al.*, 2009), suggested a function in suppressing *Athila* activity in the germline. In addition, as these 21nt siRNAs were detected only in mature wild-type pollen, it is possible that they are male-specific, raising another interesting possibility that these siRNAs accumulate in the sperm cells to be delivered and play a role after fertilization, in the developing zygote (CALARCO and MARTIENSSEN, 2011). This hypothesis, together with the observation that 24nt siRNAs are produced in the female gametophyte through an AGO9-dependent pathway (MOSHER *et al.*, 2009; OLMEDO-MONFIL *et al.*, 2010), suggest a mechanism in which different classes of siRNAs are delivered at fertilization by both male and female gametes. These siRNAs would then contribute with epigenetic information to silence transposon activity on each parental genome, thus explaining the hybridization barrier that exists in Arabidopsis (MARTIENSSEN, 2010). In order to address these hypotheses, it would be necessary to isolate mutant alleles of genes specifically involved in the production of *Athila* 21nt siRNAs, without compromising pollen viability.

We tried to understand the mechanism of transposon de-repression in pollen by analyzing DNA methylation throughout pollen development, and found that the gametes are hypermethylated at CG and CHG context, but devoid of CHH methylation that is characteristic of TE silencing by RdDM in sporophytic cells (Chapter 4). Considering that in

mature pollen at anthesis the sperm cells are still actively replicating their genome (~1.4C at pollination) (FRIEDMAN, 1999), it is possible that RdDM activity in the germline occurs mainly during pollen tube growth, at the point when the sperm cells are finishing replicating their genome and preparing for fertilization. Future studies should address this issue by isolating sperm cells and VN from growing pollen tubes to analyze their epigenomes. In contrast, the VN loses CG methylation genome-wide at TE bodies, as a possible consequence of DEMETER (DME) activity (SCHOFT *et al.*, 2011). The extent of this demethylation is not as strong as observed in the Arabidopsis endosperm (GEHRING *et al.*, 2009; HSIEH *et al.*, 2009), which is enhanced by the fact that MET1 is absent in highly proliferative tissue. For this reason, it is not clear yet if DME activity is indeed responsible for TE reactivation in the VN. Hypermethylation at CHH context in the VN is difficult to understand, but previous reports suggested that this could be the consequence of losing centromere identity that leads to chromatin decondensation, thus making the genome available for RdDM (SCHOFT *et al.*, 2009). These observations would explain why, unlike *ddm1* mutant nuclei, the VN gains CHH methylation, but its biological meaning is not clear. It is possible that the CHH methylation profile in the VN reflects RNAi-mediated chromatin modifications, and further experiments should focus on the distribution of the various histone modifications related with chromatin decondensation and TE de-repression.

On the other hand, we provide evidence suggesting a double layer of epigenetic regulation during pollen development that is associated with 24nt siRNAs and RdDM targeting certain TEs like RC/Helitrons and DNA/MuDR, while other elements from the same families are globally demethylated (Chapter 4). In the light of genomic imprinting this is an interesting observation, as these types of TEs tend to locate around imprinted genes. In fact, imprinted loci expressed from the maternal allele are more prone to be targeted by RdDM in pollen, possibly compensating the indiscriminate activity of DME. Mutants impaired in

RdDM such as *drm2* should be used to understand if these particular TEs lose CG methylation, and if it affects parent-of-origin expression in the developing seed. We found that in pollen RdDM is less prominent at imprinted genes that are expressed exclusively from the paternal allele in the endosperm, and these loci lose CG methylation in the VN. In fact, imprinting of paternal alleles seems to follow different rules than those verified for maternally imprinted genes. This is the case of *PHERES1* (*PHE1*), which is expressed only from the paternal allele because a 3' downstream tandem repeat is methylated (MAKAREVICH *et al.*, 2008). Normally, maternally expressed genes need to be demethylated by DME in order to be expressed in the endosperm, while the corresponding paternal allele is thought to remain methylated being subsequently repressed by the FIS2 Polycomb group (PcG) complex (KOHLER and MAKAREVICH, 2006). However, it is possible that a hypermethylated allele inherited from the paternal side, is targeted by DME only after fertilization during endosperm development. We do not understand yet the mechanisms and biological meaning of TE re-activation in pollen, but our data provides evidence that among other things, it could also be essential for re-establishing imprinting marks in the paternal genome (see proposed model in Figure 4.5, Chapter 4).

This work presents the first characterization of miRNAs in isolated plant gametes (Chapter 5). Although most known miRNA families were detected in the sperm cells, we found enrichment for some families when compared with their abundance in total pollen, clearly reflecting a distinct cell fate and suggesting that the miRNA processing machinery is active during germ cell differentiation. However, we could not determine a role for miRNAs during sperm cell specification, both using the viral suppressor of gene silencing P1:HcPro or target mimicries designed for depleting specific miRNAs families. As AGO1 and AGO10 are down-regulated sharply during pollen development and not expressed in the germline

(Chapter 3), we characterized its close homolog AGO5 that is highly abundant in the sperm cells. We could observe fertilization defects, and reciprocal crosses demonstrated that the mutant allele is poorly transmitted through both male and female gametophytes, which was expected as it is also expressed in ovules. However, microarray experiments suggested that AGO5 might not be involved directly in the canonical miRNA pathway in pollen, as miRNA-targets are not up-regulated in *ago5-4* pollen. Nevertheless, we cannot rule out the possibility that slight changes occurring in the sperm cells are diluted when analyzing gene expression in pollen samples. It is very interesting that miR159 is one of the most abundant miRNA detected in sperm cells, and its target *DUO1* is down-regulated in *ago5-4* pollen, together with *MGH3* and other DUO1-activated genes such as *GEX2*, *GCS1* and *TIP5;1*. This result implies that instead of loading miR159 to induce cleavage of target transcripts, AGO5 could sequester miR159, preventing it from being loaded into another Argonaute complex involved in post-transcriptional gene silencing. The other Argonaute protein expressed in sperm cells is AGO9, and expression of *DUO1* and *MGH3* in double homozygous *ago5/ago9* is restored as compared with *ago5-4* pollen and up-regulated when compared with wild type pollen. Altogether, this mechanism resembles the recently described pathway for miR166 activity in the shoot apical meristem of Arabidopsis embryos, where AGO10 is involved in sequestering this miRNA family and preventing that it gets incorporated into AGO1, which would cleave target transcripts that are required for SAM maintenance (ZHU *et al.*, 2011). One possible experiment should address this hypothesis by performing 5'RACE experiments with sperm cell RNAs, to clarify if *DUO1* transcripts are indeed cleaved in the mutant but not in the wild-type. In addition to the role of Argonaute proteins in small RNA-directed DNA methylation of heterochromatic regions, and miRNA-guided slicing and translational repression of target transcripts, it was recently shown that the catalytic activity of the mammalian Argonaute2 is responsible for a Dicer-independent pathway of miRNA processing (CIFUENTES *et al.*, 2010;

YANG *et al.*, 2010). As Argonaute proteins are widely conserved among eukaryotes, it might be worthwhile to investigate whether Argonaute proteins in plants could also process miRNA stem-loop precursors.

Taken together our results could only emphasize the possibility that miRNA pathways may be suppressed during male sperm cell development, even though their activity was detected in pollen (Chapter 5, GRANT-DOWNTON *et al.*, 2009). A recent study using a null allele of *DCL1* provided more evidence towards this hypothesis, as reciprocal crosses could show that the *dcl1-5* null allele is normally transmitted through both gametophytes (NODINE and BARTEL, 2010). It would be interesting now to analyze individual *dcl1-5* sperm cells, to confirm that they lose miRNAs normally detected in wild-type pollen. If miRNAs are lost in mature *dcl1-5* pollen, it would indicate that miRNAs are processed during pollen development and are not essential in the male gametophyte. Alternatively, normal miRNA levels in *dcl1-5* pollen would suggest that there is either a DCL1-independent pathway in pollen, or no need for new miRNA processing, as miRNAs or DCL1 protein itself could be inherited from the diploid pollen mother cell, remaining in the post-meiotic microspore and throughout male gametogenesis. In fact, our results suggest that AGO5 might be involved in suppressing post-transcriptional gene silencing by sequestering certain miRNAs, and additionally involved in repressing gene expression at the transcriptional level by RdDM, targeting TEs and repeats adjacent to protein-coding genes. If confirmed, these mechanisms would include AGO5 as having a prominent role in promoting epigenetic reprogramming during germline development. It is important to understand now which small RNAs classes bind to AGO5 and AGO9 in Arabidopsis pollen and isolated sperm cells. Previous studies performed AGO5 pull down from protein fractions isolated from inflorescence tissue, showing that AGO5-bound small RNAs are mostly siRNAs derived from intergenic regions

and have a Cytosine at the 5' end (MI *et al.*, 2008; TAKEDA *et al.*, 2008), and very few miRNAs where detected. However, it is possible that this profile is completely different in mature pollen, since AGO1 and AGO4 are not as prominent as in sporophytic tissues.

2. REFERENCES

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M., and Twell, D. (2011). The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. Plant Cell *23*, 534-549.

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

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D*., et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Friedman, W.E. (1999). Expression of the cell cycle in sperm of Arabidopsis: implications for understanding patterns of gametogenesis and fertilization in plants and other eukaryotes. Development *126*, 1065-1075.

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

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

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.

Kohler, C., and Makarevich, G. (2006). Epigenetic mechanisms governing seed development in plants. EMBO Rep *7*, 1223-1227.

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

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

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C.*, et al.* (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. Cell *133*, 116-127.

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

Nodine, M.D., and Bartel, D.P. (2010). MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. Genes Dev *24*, 2678-2692.

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

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

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

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijo, J.A., and Martienssen, R.A. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell *136*, 461-472.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. Plant & Cell Physiol *49*, 493-500.

Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct conversion of C. elegans germ cells into specific neuron types. Science *331*, 304-308.

Yang, J.S., Maurin, T., Robine, N., Rasmussen, K.D., Jeffrey, K.L., Chandwani, R., Papapetrou, E.P., Sadelain, M., O'Carroll, D., and Lai, E.C. (2010). Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A *107*, 15163-15168.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.H., Liou, L.W., Barefoot, A., Dickman, M., and Zhang, X. (2011). Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. Cell *145*, 242-256.

APPENDIX

Section I

Table S2.1 - Primers used in Chapter 2

Section II

Figure S3.1 - Quality parameters analyzed by Expression Console™ software (Affymetrix) for RMA (Robust Multichip Analysis) summarized data. **(A)** Pearson's Correlation of signal intensities. **(B)** MvA plots show the overall reproducibility of a single replicate within each different sample (sperm cells, pollen and seedling), according to the data variability 'M', as a function of the mean 'A'

Table S3.1 - Expression summary of all genes represented on the ATH1 array and sRNAs and Methylation Analysis. Average of signal values and Present/Absent calls in sperm cells, pollen and seedling samples, highlighting sperm-enriched, -depleted, -selective and preferentially expressed genes. The comparison of gene expression between all three samples is represented by the values of fold change. Gene families involved in the Small RNA and DNA methylation pathways were analysed separately in terms of enrichment in sperm cells relative to seedlings **(.xls file)**

Table S3.2 - Gene functional Analysis by DAVID tools. Functional Clustering of genes detected in sperm cells, pollen, seedlings and highlighting those with enriched expression in sperm, for which the genes representing the first fifteen clusters with a higher score are presented **(.xls file)**

Table S3.3 - Primers used for RT-PCR analysis in Chapter 3

Section III

Figure S4.1 - DNA methylation through pollen development. **(A)** A schematic representation of tissues profiled in this study. **(B)** Overall view of methylation of all TEs across chromosome 4. Height of bar and color represent Heat map presentation of methylation percentages of all TEs on chromosome 4 broken up into each methylation context

Table S4.1 - Summary of bisulfite sequencing for sperm cells (SC), vegetative nucleus (VN), microspore and leaf libraries, including total read count, methylated and unmethylated calls and percentages

Table S4.2 - List of TEs showing Differentially Methylated Regions (DMR). CHH DMRs between leaf and microspore, CHH DMRs between and SC and VN, and CHG DMRs between SC and VN **(.xls file)**

Section IV

Table S5.1 - Normalized reads for known miRNAs detected in sRNA data sets of SCs and pollen, in comparison with that of inflorescence tissue **(.xls file)**

Table S5.2 - Potentially novel miRNAs detected in the sRNA data sets of SCs and pollen. Predicted foldback structure of stem-loop precursors and chromosome locus

Table S5.3 - Normalized counts of known miRNA families and newly identified miRNAs, for the different sequence contexts from 19 nt to 24 nt and extensions in the 5'-terminal nucleotide, in the data sets of SCs, pollen, and inflorescence **(.xls file)**

Table S5.4 - Expression summary of genes differentially regulated in *ago5-4* represented on the ATH1 array. Comparison of gene expression between wild-type and mutant samples, filtered by a lower bound of fold change 1.2. **(.xls file)**

Table S5.5 - Primers used in Chapter 5

Table S5.6 - Genomic sequences of regions amplified by bisulfite PCR, adjacent to *ago5-4* up-regulated genes, *At3g17120* and *At3g47470*

Video S5.1 - Time-lapse imaging of AGO5-GFP in the SCs of growing pollen tube **(.mov file)**

Video S5.2 - 3D magnification of the male germ unit expressing AGO5-GFP in the cytoplasm

of SCs and the cytoplasmic tail connecting to the VN **(.mov file)**

Section I

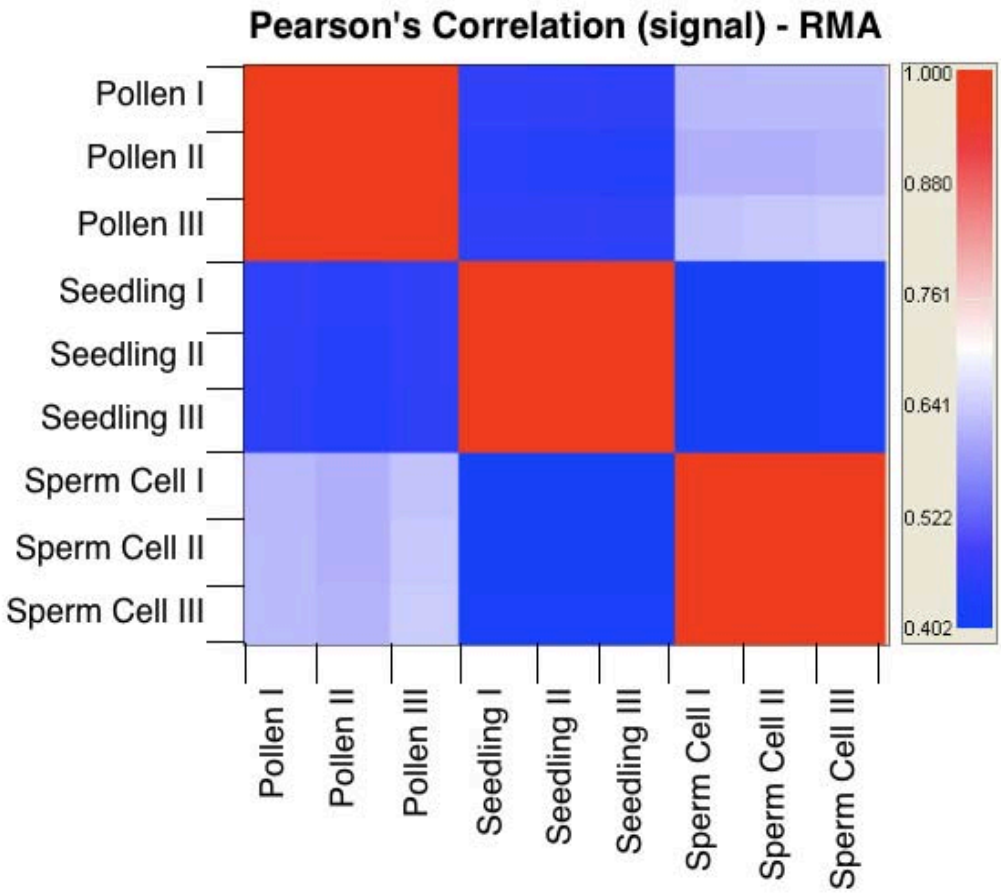
Table S2.1 - Primers used in Chapter 2

	Primer	Sequence (5' – 3')
RT-PCR	VEX1_for	CACGCACCTTCAGAGCATTT
	VEX1_rev	TCTTAGCCGTGGAGTTGGACTTG
	MGH3_for	ACTAGACGACCGTACCGTGGT
	MGH3_rev	CGATTCTATCTCACCCATCAA
	TUB4_rt_for	ATCCCAAACAACGTCAAGTCC
	TUB4_rt_rev	CTCTCCGGCTGTAGCATCTTGGTAC
GATEWAY Cloning	Mgh3p_TOPO _F	caccTACTTCTCCGACCAAAAACTT
	Mgh3 _TOPO_R	AGCACGTTCCCCACGAATGCGT

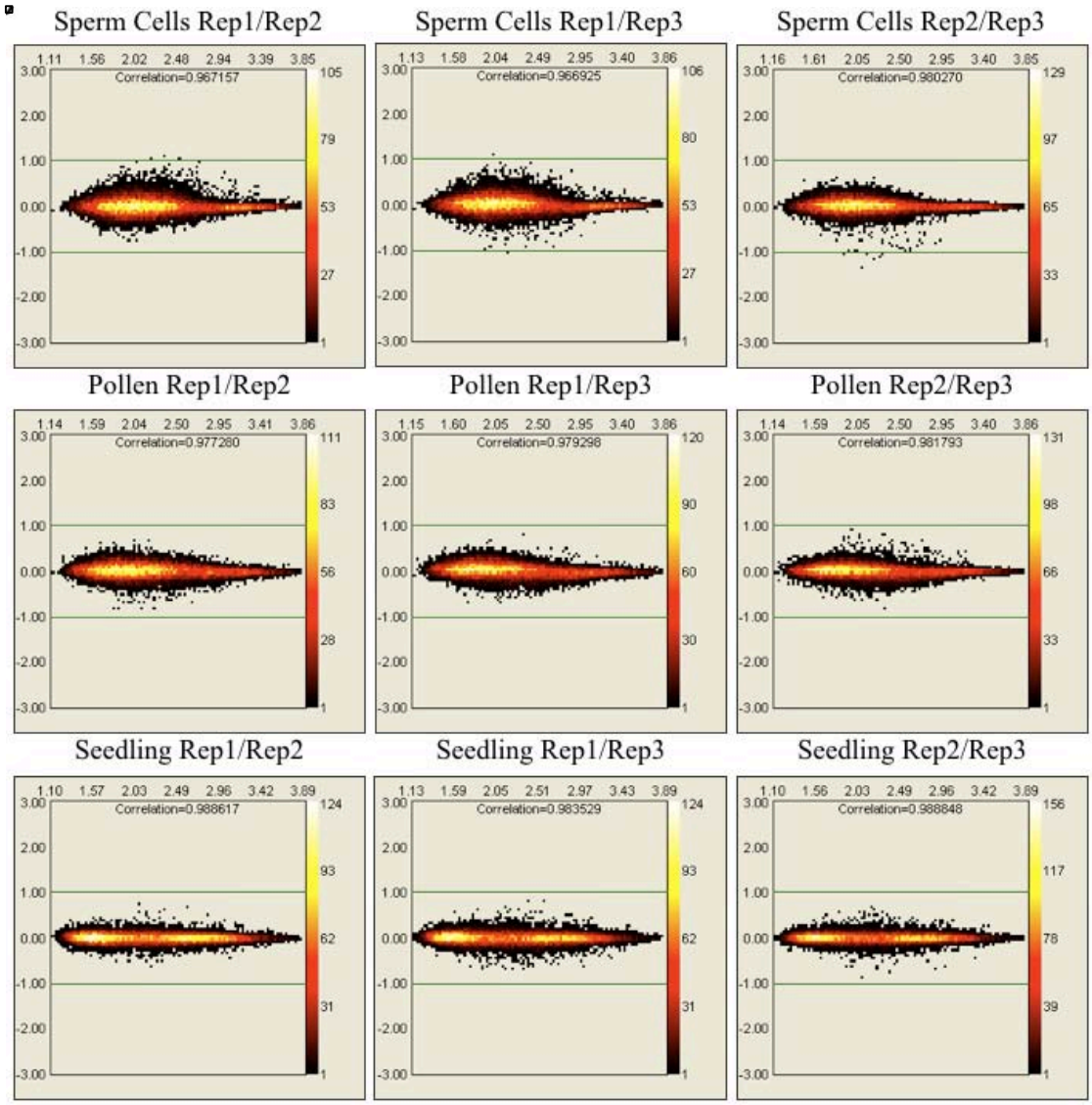
Section II

Figure S3.1

A



B



Quality parameters analyzed by Expression Console™ software (Affymetrix) for RMA (Robust Multichip Analysis) summarized data. **(A)** Pearson's Correlation of signal intensities. **(B)** MvA plots show the overall reproducibility of a single replicate within each different sample (sperm cells, pollen and seedling), according to the data variability (M), as a function of the mean (A).

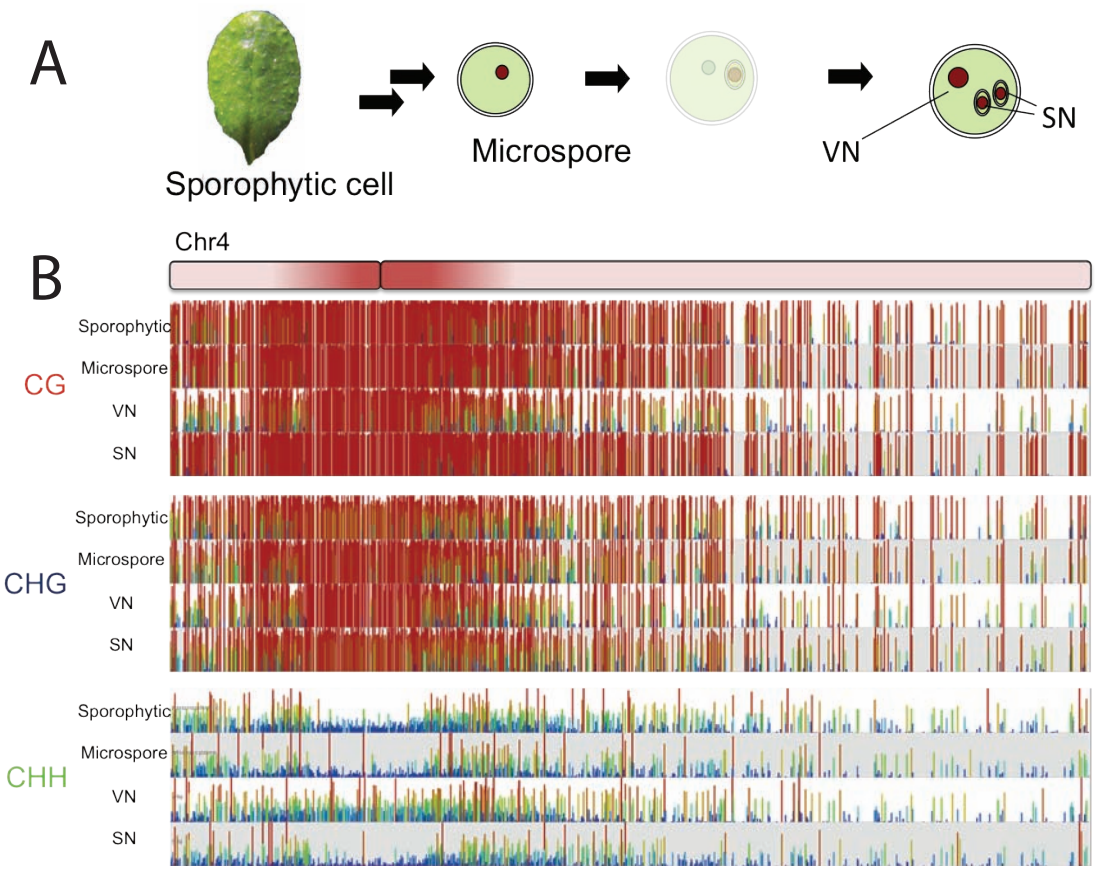
Table S3.3 - Primers used for RT-PCR analysis in Chapter 3

AGI	Primer	Sequence (5' – 3')
At5g014180	At5g014180_for	CACGCACCTTCAGAGCATTT
	At5g014180_rev	TCTTAGCCGTGGAGTTGGACTTG
At1g19890	Atmgh3_for	ACTAGACGACCGTACCGTGGT
	Atmgh3_rev	CGATTCTATCTCACCCATCAA
At5g62850	Atvex1_for	ATGGAAGATGAAATCGGTCTC
	Atvex1_rev	GCAGCATACATGATGACGTTG
At1g69770	Cmt3_for	ATATGATCATCACCCTCTTAA
	Cmt3_rev	GGAAGAGTTGGCTTCATGAAT
At3g43920	Dcl3_for	TGGTGTCAAAGCACAAGCCAC
	Dcl3_rev	TACATCACAGCCTCACGATTG
At4g11720	Hap2_for	GTTTAGTCGATTTCATCACCG
	Hap2_rev	TTGTTTCCTCTGTCTTTCCAC
At3g23780	Nrpd2a_for	GTAGCTCTGCTGCATACACCA
	Nrpd2a_rev	CCATGCTGAACAGCTCCTGAC
AT1G23210	AT1G23210_for	GGTGCCATGTGGCGAATTTAC
	AT1G23210_rev	AAGTAGGCCCAAAAGAGGTGC
AT1G56380	AT1G56380_for	GAAGTGTGGACTTGTCGAAGA
	AT1G56380_rev	AGAAACACCTCATCGGGACAT
AT1G60240	AT1G60240_for	CCACTTCCTCCGAAGCGAAGA
	AT1G60240_rev	GTTGTGGATACAGACGTCGAA
AT1G66610	AT1G66610_for	CAAAATGGCTGCACCGAGACA
	AT1G66610_rev	GTCCTTCTCCGGAGTTTGGAA
AT2G01920	AT2G01920_for	TTCGCGCTTATTTCATGTTTC
	AT2G01920_rev	ACTTTTGCGGCTTTCCAATCT
AT3G09620	AT3G09620_for	TCTAAGGCTCCTGTGACTGCT
	AT3G09620_rev	GCTCCTCCAGGAAGGGAGAAG
AT3G13820	AT3G13820_for	GAGGAAGATGACGAAGAGTAC
	AT3G13820_rev	TGGTTGCTTGACAGGAACTTT
AT3G19890	AT3G19890_for	GGAAGAAGATCACCTTCCAGG
	AT3G19890_rev	AGAGCTAGAACACACAAGTGC
AT3G28400	AT3G28400_for	ATGCTAGACCAAGCTCTGAAG
	AT3G28400_rev	AGACTGAGGCTGAGTAGATGA
AT3G49450	AT3G49450_for	ATGTTCTGTCGATCATTACCC
	AT3G49450_rev	GTGTAAACTGCTTGACCTTCA
AT3G58390	AT3G58390_for	GAATGTTGTGCAAGCCTTTCT
	AT3G58390_rev	TAGTTGTTCCCCTGAAGCATG
AT3G60760	AT3G60760_for	GTGGCTTTTGCATACTCAGAT
	AT3G60760_rev	ATTGATGGCGATGTCAAAACT
AT3G63020	AT3G63020_for	CGCTGGAGCACTCCTGTTGAT
	AT3G63020_rev	ATCTCCGGTGAATCTACTGCC
AT4G16820	AT4G16820_for	CGGTTGCGGTGTTTTCCTTTG
	AT4G16820_rev	GTTGCTCATCCAACAGTTTCC
AT4G35370	AT4G35370_for	CCAGAAGTACTCCTCAGTGGG

AGI	Primer	Sequence (5’ – 3’)
	AT4G35370_rev	AACATATGGCAGCAAAAGTTC
AT4G35700	AT4G35700_for	AAGTGCTAGTGGTACCGGTGA
	AT4G35700_rev	CACCATCGGAAGAAGCCGTAG
AT4G36070	AT4G36070_for	TCAGCCGTCTGAAGCAGATTG
	AT4G36070_rev	TGACTTGAGGCTTGCACTTCT
AT5G04210	AT5G04210_for	CTATACCGCCATGAGATCAAC
	AT5G04210_rev	GGAGCTAGAAGAACCAACCTC
AT5G46370	AT5G46370_for	GGTGTTGTGGTGTTATGCCTT
	AT5G46370_rev	TGTCAAACTGGAAACCGATCG
AT5G55680	AT5G55680_for	GATTCTATTGAGGTGGTGGGA
	AT5G55680_rev	TCCGGTGATCCAAAAATCTCC
AT5G59200	AT5G59200_for	GAATCAGGTGACTTTAGTTGC
	AT5G59200_rev	AGGGTGTGCTATATCTCCCAC
AT5G44340	Tub4_for	ATCCCAAACAACGTCAAGTCC
	Tub4_rev	CTCTCCGGCTGTAGCATCTTGGTAC

Section III

Figure S4.1



DNA methylation through pollen development. **(A)** A schematic representation of tissues profiled in this study. **(B)** Overall view of methylation of all TEs across chromosome 4. Height of bar and color represents heat map of methylation percentages for all TEs on chromosome 4, broken up into each methylation context

Table S4.1 - Summary of whole-genome bisulfite sequencing of sperm cells (SC), vegetative nucleus (VN), microspore and leaf DNA libraries, including total read count, methylated and unmethylated calls and percentages.

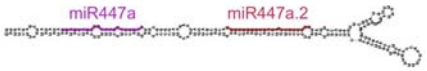
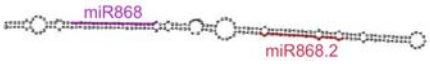
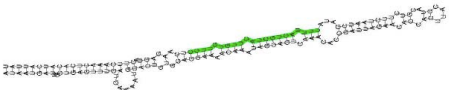
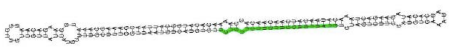
	Total Read Count	Methylated Call	Unmethylated Call	%methylation
VN CG	12173873	3814130	8359743	**31.33045663**
SC CG	15561600	5168950	10392650	**33.21605747**
Leaf CG	18591187	6714705	11876482	**36.11767769**
Microspore CG	22785543	6421822	16363721	**29.26769593**

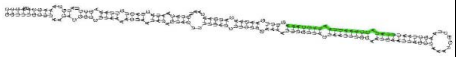
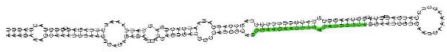
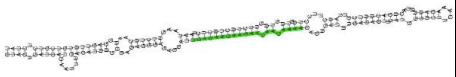
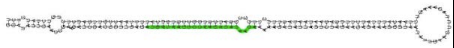
	Total Read Count	Methylated Call	Unmethylated Call	%methylation
VN CHG	12327702	1995156	10332546	**16.18433022**
SC CHG	15603102	2058724	13544378	**13.19432508**
Leaf CHG	17990924	2939506	15051418	**16.33882729**
Microspore CHG	15138185	1773623	13364562	**12.95601411**

	Total Read Count	Methylated Call	Unmethylated Call	%methylation
VN CHH	50444784	3502968	46941816	**6.944162949**
SC CHH	60479531	1062640	59416891	**1.757024207**
Leaf CHH	64316917	2950982	61365935	**4.588189449**
Microspore CHH	82923524	1486413	81437111	**1.202724799**

Section IV

Table S5.2 - Potentially novel miRNAs detected in the sRNA data sets of SCs and pollen, predicted targets, foldback structure of stem-loop precursors and Chr locus.

MicroRNA	Sequence	psRNA target prediction	Foldback Structure/Chr Loci
ath-miR2934	CAUCCAAGGUGUUUGUAGAAA	-	 Chr3: 5500049-5500069
ath-miR4240.2	AUGGCUAGAGUGACUAGACCCG	-	 Chr4: 10711731-10711752
ath-miR447a.2	UAUGGAAGAAAUUGUAGUAUU	AT1G42630, AT1G54710	 Chr4: 1528146-1528166
ath-miR447c.2	CCCCUUACAAUGUCGAGUAAA	-	 Chr4: 1523443-1523463
ath-miR868.2	UCAUGUCGUAAUAGUAGUCAC	CMT1 AT1G80740, CMT2 AT4G19020	 Chr3: 6488293-6488313
ath-miR5012	UUUUACUGCUACUUGUGUUCC	AT1G53700, AT2G37678	 Chr2: 8439829-8439849
ath-miR5013	UUUGUGACAUCUAGGUGCUUU	AT3G60580	 Chr2: 7173950-7173970
ath-miR5014	UUGUACAAAUUUAAGUGUACG	-	 Chr1: 24557674-24557694
ath-miR5015.1	UUGGUGUUAUGUGUAGUCUUC	-	 Chr5: 20572201-20572221
ath-miR5015.2	UCUGUUGUUGUUGGUGUUAUG	AT2G38320, AT1G12860, AT2G01760, AT5G38740	 Chr5: 20572211-20572231

MicroRNA	Sequence	psRNA target prediction	Foldback Structure/Chr Loci
ath-miR5016	UUCUUGUGGAUUCCUUGGAAA	-	 Chr2: 9927232-9927252
ath-miR5017	UUAUACCAAAUUAAUAGCAAA	-	 Chr4: 11962971-11962991
ath-miR5018	UUAAAGCUCCACCAUGAGUCCAAU	-	Chr1: 24547624-24547647
ath-miR5019	UGUUGGGAAAGAAAAACUCUU	AT3G58810, AT1G14510, AT4G19550	 Chr2: 10410415-10410435
ath-miR5020a	UGGAAGAAGGUGAGACUUGCA	-	 Chr4: 6791397-6791417
ath-miR5020b	AUGGCAUGAAAGAAGGUGAGA	-	 Chr4: 6787104-6787124
ath-miR5021	UGAGAAGAAGAAGAAGAAAA	-	 Chr2: 11981899-11981918
ath-miR5022	GUCAUGGGGUAUGAUCGAAUG*	-	 Chr1: 22376144-22376164
ath-miR5023	AUUGGUAGUGGAUAAGGGGGC*	AT5G24950	 Chr3: 5497076-5497096
ath-miR5024	AUGACAAGGCCAAGAUAUAACA	-	 Chr1: 16768828-16768849
ath-miR5025	ACUGUAUAUAUGUAAGUGACA	AT2G48010	 Chr1: 16996353-16996373

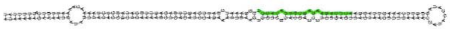
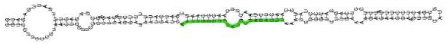
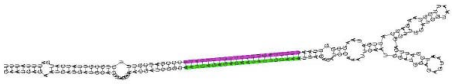
MicroRNA	Sequence	psRNA target prediction	Foldback Structure/Chr Loci
ath-miR5026	ACUCAUAAGAUCGUGACACGU	-	 Chr4: 7844608-7844628
ath-miR5027	ACCGGUUGGAACUUGCCUUAA	-	 Chr5: 10943927-10943947
ath-miR5028	AAUUGGGUUUAUGCUAGAGUU	-	 Chr4: 6546765-6546785
ath-miR5029	AAUGAGAGAGAACACUGCAAA*	AT2G30070, AT4G02900	 Chr5: 12285774-12285794

Table S5.5 - Primers used in Chapter 5

	Primer	Sequence (5' – 3')
RACE	AT4G02900_5RACE_2	CCAGTATCTAGCCCCGCTCTCATAT
	AT3G60580_5RACE_2	CGACCCGAATCACGCAAAGTCTGTTA
	AT4G02900_5Nstd	GCTGCTGAAACGCTGTCCCGGTAAT
	At3g19890_GSP_R	CACAAGTGCGCTCCGATACACGTCTTTT
	At3g19890_GSPNest_R	GGTTCTTGTCCTTGTCGTAACCGAGAGCAA
	At1g62910 F	GGGAAAATAGAGGCTGATGTTGTG
	At1g62910 R	TCTCTTAGCCTTACAAAATCCCTTT
	At5g55020 F	TTTCCCCTGTTTCAAGAGACTCC
	At5g55020 R	TGTTGGTGTTGTTGTTGTGGTTC
MGH3 promoter	mgh3p_hind3_F	aatAAGCTTTACTTCTCCGACCAAAAACTT
	mgh3p_Spe1_R	aatACTAGTGTGCGATTTCTTCGAGAGAAC
P1:HC-Pro	attb1_P1HcPro_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTcagcagttacattcgcatcag
	attb2_P1HcPro_R	GGGGACCACTTTGTACAAGAAAGCTGGGTCctatgcagctccgacgcggta
Bisulfite sequencing	At3g17120 _ss_F	GAAGATGAAAAAGAAAYAAAAGAAG
	At3g17120_ss_R	ACTTRTTACACTAACRCTCCTTACT
	At3g47470_ss_F	TATAATGGATTAGAAGATAYTGAAG
	At3g47470_ss_R	RCTCCAAAATTTTARTRTTAATCAA
	At2g20610 exon_F	GTTGYTGATTATATGAAYYGAGATYTT
	At2g20610 exon_R	TTAATTACAACCATARCCACARTRTTCTC
AGO5-GFP	attB1_ago5_1_for	GGGGACAAGTTTGTACAAAAAAGCAGGCTtccgtagataaaagcggtattat
	attB1_ago5_2_for	GGGGACAAGTTTGTACAAAAAAGCAGGCTtctttgttttctgaataatctgc
	attB2_ago5_rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCgcaataaaacataacctcctt
AGO5 Genotyping	ago5_geno_F	CCCGTCTACCTACCAATGGA
	ago5_geno_R	ACCACGGTCAATCCGAGTAG
	LBb1.3	ATTTTGCCGATTTCGGAAC
RT-PCR	ago5RT_F	TGTTCCTCGTGTTACTTTCGTG
	ago5RT_R	GTGGCACAATTGACACAGATTT
	ago1RT_F	CTGTTTGCTCAGAACCACAATG
	ago1RT_R	TAGAATCGAGCCCTAAAAGCTG
	ACT11_F	GTACCTTCCAACAGATGTGGATA
	ACT11_R	TTCGATTAATAGCAGAAACACC

	Primer	Sequence (5’ – 3’)
	Duo1_RT_for	CTACCGGGAAGAACTGATAACG
	Duo1_RT_rev	CGAAAGGCTGATGCTCTAACTT
	MGH3_for	ACTAGACGACCGTACCGTGGT
	MGH3_rev	CGATTCTATCTCACCCATCAA
	HcProRT_for	AGAAGTCGATTAAGGTCG
	HcProRT_rev	TTCGAAGTTCGTTGGGAC
	MYB101_RT_for	TAGGGTTTCCTCCAGTGACCTA
	MYB101_RT_rev	GAGGCCATAGGATCTACACTGG

Table S5.6 - Genomic sequences of regions amplified by bisulfite PCR, adjacent to *ago5-4* up-regulated genes, *At3g17120* and *At3g47470*.

***At3g17120* - 5' region methylated (chr3:5838096-5841095)**
AAAGAAGATGAAAAAGAAACAAAAGAAGGAAGAAGAAAGAGAAAGGCTAAAATAGACTAA
ATATTGCCAAAATTTCTGTAGCCGACAAATACTATTTGGTCCAAGGTTATTTTGTGTATT
CTTTTGAAGTCAAAAGTTATTTCTTACACATACTCAACTATAAAGGGGATAAAGTATAGT
GTGAATTGTTGACGGCAGAAGAGAGAAAGGTACATATTTATATCTCAATGAAAAGATAAT
AAACGTGTTAATAGAGACAGTAAGGAGCGTTAGTGTAACAAGTCAGATTCATGGGGCCAC

***At3g47470* - 5' region methylated (chr3:17495034-17498033)**
TTTTAAACTGTTAACTCAAATCTTATAATGGATTAGAAGATACTGAAGATCCTTATAAGA
TATTTTGGGAATATTCTTCATGCATTTTGTAGGAAATTACAAAATTTAGTGTGCTAACAG
ATGATCTAGCCAGACCCCAGAAAACAAGATACATAAGCCATCGATTCGGCCTCAAAATTT
GAAAGCCATTTTAAAAGAAAAAACTGTTACAAATATTTTTGAGTTTGTAGAATCCTACTA
TTTACATTGTAGTCTGTTATATAATCCCACATTTGATTAACACTAAAATTTTGGAGCAAA