DISS. ETH NO. 20165

Initiation of seed coat development is controlled by Polycomb group proteins and requires a signal from the sexual endosperm

A dissertation submitted to

ETH ZURICH

for the degree of
Doctor of Sciences

presented by

PAWEL JAN ROSZAK

Master of Science in Biotechnology
Poznan University of Life Sciences

Born on January 14, 1982
Citizen of Poland

accepted on the recommendation of

Prof. Dr. Claudia Köhler, examiner
Prof. Dr. Rita Groß-Hardt, co-examiner
Prof. Dr. Samuel C. Zeeman, co-examiner

2012

Acknowledgements

First of all, I would like to thank my supervisor Prof. Claudia Kohler for her constant support, interest and long hours of discussion on my projects. Looking back, I could not imagine a better place than Claudia's group for doing my PhD studies.

I am grateful to Prof. Sam Zeeman and Prof. Rita Groß-Hardt for participating in my thesis committee.

I am very thankful to all members from Prof. Kohler's and Prof. Gruissem's groups for the excellent working environment, for all the help and support I received from them whenever I was in need and for all the joyful moments that made me feel I am in the right place.

I would like to thank Sabrina Huber for the technical help with many experiments and Andre Imboden for taking care of my plants in Eschikon.

I am thankful to Prof. Gruissem for sharing laboratory facilities with our group and accepting me in his group during the last months of my PhD studies.

I would like to thank all my friends from outside of the institute, especially members of our dancing group for so much fun during the last few years of my stay in Zurich.

I would like to thank my parents and siblings for their permanent support and encouragement.

Finally I would like to thank Karolina for being with me every day, for giving me hope and bringing sunshine into my life.

TABLE OF CONTENT

ABBREVIATIONS

35S	Promoter derived from *Cauliflower mosaic virus* gene encoding for 35S ribosomal subunit
δVPE	δ VACUOLAR PROCESSING ENZYME
AG	AGAMOUS
AGO9	ARGONAUTE 9
AHK2, 3	ARABIDOPSIS HISTIDINE KINASE 2, 3
AP2	APETALA 2
BAN	BANYULS
CLF	CURLY LEAF
Col	Columbia
CRE1	CYTOKININ RESPONSE 1
cv	coefficient of variation
DAE	Days after emasculation (removal of anthers)
DAP	Days after pollination
DIC	Differential interference contrast
EGFP	Enhanced green fluorescent protein
EMS	Ethyl methanesulfonate
EMF2	EMBRYONIC FLOWER 2
Esc	Extra sexcombs
E(z)	Enhancer of zeste
FIE	FERTILIZATION INDEPENDENT ENDOSPERM
FG	Female gametophyte
FLC	FLOWERING LOCUS C
FM	Functional megaspore
FIS2	FERTILIZATION INDEPENDENT SEED 2
FT	FLOWERING TIME
GUS	β-GLUCURONIDASE
H3K27me3	Trimethylation of Lysine 27 on Histone 3
Hox	Homeotic box
IKU	HAIKU
KPL	KOKOPELLI
L*er*	Landsberg *erecta*
MADS	Acronym derived from four members of the family: <u>M</u>CM1 (*Saccharomyces cerevisiae*), <u>A</u>GAMOUS (*Arabidopsis thaliana*), <u>D</u>EFICIENS (*Antirrhinum majus*) and <u>S</u>RF (*Homo sapiens*)
MEA	MEDEA
MINI3	MINISEED 3
MMC	Megaspore mother cell
MS	Murashige and Skoog medium
MSI1	MULTICOPY SUPPRESOR OF IRA 1
NCD	Nuclear-cytoplasmic domain

Nurf55	Nucleosome remodeling factor subunit (55 kD)
PAs	Proanthocyanidins
Pc	Polycomb
PCD	Programmed cell death
PcG	Polycomb Group
Ph	Polyhomeotic
PHD	Homeodomain
PHE1	PHERES 1
Pho-RC	Pleiohomeotic Repressive Complex
PP2A	PROTEIN PHOSPHATASE 2A
PRC1, 2	Polycomb Repressive Complex 1, 2
PRE	Polycomb Responsive Element
Psc	Posterior sex combs
Sce	Sex combs extra
SEM	Standard error of the mean
SET	Acronym derived from three proteins containing this catalytic domain; Supressor of variegation, Enhancer of zeste and Trithorax
Sfmbt	Scm-related protein containing four malignant brain tumor domains
siRNA	Small interfering RNA
Su(z)12	Suppressor of zeste 12
SWN	SWINGER
T-DNA	Transfer DNA
Tris	Tris(hydroxymethyl)aminomethane
TT	TRANSPARENT TESTA
TTG1, 2	TRANSPAREN TESTA GLABRA 1, 2
TTN2	TITAN 2
VEFS	Acronym derived from four proteins containing this domain VERNALIZATION 2, EMBRYONIC FLOWER 2, FERTILIZATION INDEPENDENT SEED 2, Suppressor of zeste 12
VEL1	VERNALIZATION5/VIN3-LIKE 1
VIN3	VERNALIZATION INSENSITIVE 3
VRN2	VERNALIZATION 2
VRN5	VERNALIZATION 5
Ws	Wassilewskija
wt	Wild type
YFP	Yellow fluorescent protein

SUMMARY

Sexual reproduction and seed formation strongly contributed to the evolutionary success of flowering plants. These processes involve the generation of gametes by the female and male gametophytes and transport of male sperm cells to the female gametophyte. During double fertilization two sperm cells fuse with the egg and the central cell, leading to the initiation of seed formation. Developing seeds contain the embryo surrounded by nourishing endosperm tissue and surrounded by the protective seed coat. While the endosperm is utilized by the expanding embryo as a nutrient source and the seed coat cells undergo cell death during seed maturation, the embryo is a developing organism of the following generation. In the absence of fertilization, the female gametophyte with surrounding maternal tissues does not form a seed but degenerates within a few days. This ensures that no assimilates are invested in the formation of a seed that would not fulfill its role in reproduction. Development of the seed that consists of the maternal diploid seed coat, the ephemeral, sexually derived triploid endosperm and the diploid organism of the next generation, require tight regulation of gene activity during their coordinated growth.

Polycomb Group (PcG) proteins are involved in the transcriptional repression of target genes by specific chromatin modifications, changing accessibility of the locus to the transcriptional machinery. PcG proteins are conserved in animals and plants. In plants, they control an array of developmental processes like flowering time determination, vernalization, and flower organ development. They also actively participate in the development of the seed regulating divisions of the central cell and cellularization of the endosperm. PcG proteins in *Arabidopsis* form at least three Polycomb Repressive Complexes (PRC2) with partially redundant function between their homologous subunits. The activity of the Fertilization Independent Seed (FIS) PRC2 complex is restricted to the female gametophyte and developing endosperm. Lack of any FIS complex subunit causes defects in endosperm development, leading to seed abortion. Moreover, in the absence of fertilization *fis* mutants autonomously initiate development of seed like structures containing autonomous endosperm, but no embryo.

In my work I addressed the question whether differences in the penetrance of the Fertilization Independent Seed (FIS) phenotype between FIS complex subunits results from redundant action of their homologous genes. I could demonstrate that there is no

redundancy between the *FIS2* gene and its homologs but demonstrated that the differences in the phenotype between different *fis* mutants results from malfunction of FIS complex subunits that play an additional role in sporophytic tissues of the surrounding integuments.

I discovered that decreased level of PcG proteins in the integuments cause initiation of their differentiation into seed coat, revealing that PcG proteins not only repress divisions of the central cell but also development of the seed coat in the absence of fertilization. Subsequent analysis of autonomous seeds revealed that the difference in penetrance of the FIS phenotype between FIS complex subunits is additionally accompanied by a difference in seed coat development. While mutants of FIS complex components that have an additional function in sporophytic tissues developed a normal seed coat, mutants of FIS complex components that function specifically in the female gametophyte did not develop a seed coat and the integuments degenerated. Moreover, I could directly show that seed coat development is initiated as a response to the developing endosperm what most probably encompass mobile signal formation and transmission to the ovule integuments. Therefore, I propose that only the sexual, but not the autonomous endosperm forms a signal necessary for seed coat formation and the need for this signal is bypassed in mutants with lowered dosage of PcG proteins in the integuments.

Although the nature of the signal remains undiscovered, I found the endosperm-specific MADS box protein *AGAMOUS LIKE 62* (*AGL62*) being involved in the process of signal formation. Fertilized *agl62*/- mutant seeds develop endosperm and embryo but fail to develop a seed coat, implicating that the seed coat initiation signal is absent in *agl62*. Inevitably, future work will focus on the elucidation of the signal that drives seed coat formation.

ZUSAMMENFASSUNG

Sexuelle Reproduktion und Samenbildung waren entscheidend für den evolutionären Erfolg von blühenden Pflanzen. Sexuelle Reproduktion in Blütenpflanzen beinhaltet die Bildung von Gameten durch weibliche und männliche Gametophyten und den Transport von männlichen Spermazellen in den weiblichen Gametophyten. Während der doppelten Befruchtung fusionieren zwei Spermazellen mit der Eizelle und der Zentralzelle und initiieren den Beginn der Samenbildung. Der sich entwickelnde Samen beinhaltet den Embryo, umgeben vom nährenden Endospermgewebe und umhüllt von der schützenden Samenhülle. Während das Endosperm als Nahrungsquelle dient, um das Wachstum des Embryos zu ermöglichen, wird die Samenhülle während der Samenreifung dem Zelltod unterworfen. Somit ist nur der Embryo ein sich entwickelnder Organismus der nächsten Generation. Ohne Befruchtung entwickelt sich kein Samen aus dem weiblichen Gametophyten, und dieser degeneriert innerhalb von wenigen Tagen. Dadurch wird sichergestellt, dass keine Nährstoffe verschwendet werden für Samen, die ihre reproduktive Rolle nicht erfüllen können. Die koordinierte Entwicklung der Samenschale, des Endosperms und des Embryos verlangt eine streng kontrollierte Regulation der Genaktivität dieser drei Komponenten.

Polycomb Gruppen (PcG) Proteine regulieren die transkriptionelle Repression von Zielgenen durch spezifische Chromatinmodifikationen, welche die Zugänglichkeit der Transkriptionsmaschine verändert. PcG Proteine sind in Tieren und Pflanzen konserviert. In Pflanzen kontrollieren sie eine Reihe von Entwicklungsprozessen, wie Blühzeitdetermination, Vernalisation und die Entwicklung von Blütenorganen. Sie sind ebenfalls an der Entwicklung von Samen beteiligt, wie bei der Regulation der Teilung der Zentralzelle und der Zellularisierung des Endosperms. PcG Proteine in *Arabidopsis* bilden mindestens drei Polycomb Repressive Komplexe (PRC2) mit teilweiser redundanter Funktion zwischen den homologen Untereinheiten. Die Aktivität des befruchtungsunabhängigen Samen (FIS) PRC2 Komplex ist begrenzt auf den weiblichen Gametophyten und das sich entwickelnde Endosperm. Das Fehlen einer FIS Komplex Untereinheit führt zu Defekten in der Endospermentwicklung und zum Abbruch der Samenentwicklung. Zudem initiieren nicht befruchtete *fis* Mutanten die Entwicklung von samenartigen Strukturen, die ein autonomes Endosperm aber keinen Embryo enthalten.

In meiner Arbeit ging ich der Frage nach, ob der Unterschied in der befruchtungsunabhängigen Samenbildung in verschiedenen *fis* Mutanten von der

redundanten Wirkung der homologen Gene resultiert. Ich konnte zeigen, dass *FIS2* homologe Gene nicht redundant sind. Weiterhin konnte ich zeigen, dass der Unterschied im Phänotyp zwischen verschiedenen *fis* Mutanten dadurch erklärt werden kann, dass einige FIS Untereinheiten eine zusätzliche Rolle im sporophytischen Gewebe haben, das die Integumente und später die Samenschale bildet.

Ich konnte zeigen, dass eine verringerte Dosis von PcG Proteinen in den Integumenten zur Initiierung der befruchtungsunabhängigen Samenschalenbildung führt. Diese Daten zeigen, dass PcG Proteine nicht nur die autonome Teilung der Zentralzelle unterdrücken, sondern auch die befruchtungsunabhängige Entwicklung der Samenschale. Darüber hinaus konnte ich zeigen, dass die Entwicklung der Samenschale eine Antwort des sich entwickelnden Endosperms ist, welche höchstwahrscheinlich eine mobile Signalbildung und Transmission zu den Ovulintegumenten beinhaltet. Diese Daten bilden die Grundlage für ein Modell, in dem ich postuliere, dass nur das sexuelle, aber nicht das autonome Endosperm ein Signal bildet, welches für die Bildung der Samenschale nötig ist. Obwohl ich das Signal nicht identifizieren konnte, habe ich zeigen können, dass das endosperm-spezifische MADS-box Protein AGAMOUS Like 62 (AGL62) in den Prozess der Signalbildung involviert ist. Befruchtete *agl62/-* Mutante Samen entwickeln ein Endosperm und einen Embryo, bilden aber keine Samenschale. Dies impliziert, dass das Signal für die Bildung der Samenschale in *agl62* nicht vorhanden ist. Zukünftige Arbeiten werden sich auf die Aufklärung des Signals fokussieren, welches zur Bildung der Samenschale nötig ist.

1. INTRODUCTION

1.1. Ovule and Female Gametophyte Development

The plant life cycle is divided into two subsequent phases: the diploid sporophytic and haploid gametophytic phase. In flowering plants, the prominent sporophyte supports the discreet gametophytic generation developing after meiosis in male and female floral organs.

Arabidopsis ovules initiate as a finger-like structures emerging from the carpel margin meristems at around floral stage 9 (Smyth et al., 1990; Robinson-Beers et al., 1992). They quickly differentiate to three distinguishable zones along the proximal-distal axis. Most proximally located is the funiculus that connects the ovule to the mother plant, the centrally located chalaza gives rise to the ovule integuments and in the most distally located nucellus a single cell differentiates into a megaspore mother cell (MMC). In most flowering plants only one somatic cell differentiates into a MMC, although more nucellar cells have this potential. Restriction of MMC fate in the surrounding cells involves non-cell-autonomous signaling via the small interfering RNA (siRNA) pathway. Disruption of this pathway in sporophytic cells surrounding the gametic lineage by mutations in *AGO9* (*ARGONAUTE 9*) leads to the differentiation of multiple gametic cells that are able to initiate gametogenesis (Olmedo-Monfil et al., 2010).

The differentiated MMC undergoes meiosis giving rise to a tetrad of haploid megaspores. Only one of them, the functional megaspore (FM), survives, differentiates and develops into the female gametophyte (FG), while the three other meiotic products of MMC division undergo programed cell death (PCD) (Rodkiewicz, 1970; Papini et al., 2011). Selection of the functional megaspore is assumed to be based on the interaction with surrounding nucellar cells by existing plasmodesmata connections and positional information, as it is only the most proximal (chalazal) megaspore that does not undergo PCD.

The functional megaspore enlarges and its nucleus undergoes three rounds of mitotic divisions without cytokinesis, resulting in the formation of an eight-nucleated embryo sac. Four nuclei on both poles are separated by a large central vacuole. One nucleus from the distal pole migrates towards the proximal pole, where together with one nucleus from the chalazal pole it forms a pair of two nuclei, called polar nuclei. These nuclei fuse together into one diploid central cell nucleus.
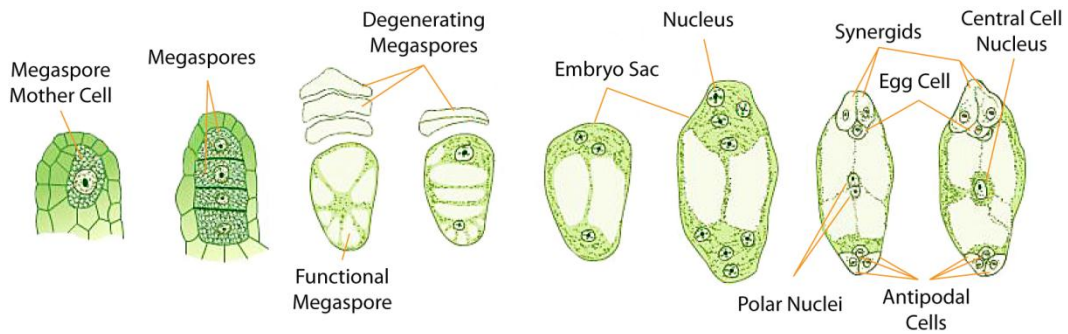
Figure 1-1. Female gametophyte development. Meiosis of the megaspore mother cell produces four haploid spores. Three of them degenerate and one develops into the functional megaspore. During development, three rounds of mitosis form two clusters of four nuclei at the two ends of the female gametophyte. Nuclei migration and cellularization generates seven cells: one egg cell and two synergid cells which form the egg apparatus at the micropylar end, and three antipodal cells at the chalazal end. In the large central cell, two polar nuclei migrate towards the center and fuse together forming the homodiploid central cell nucleus. Adapted from www.connect.in.com.

The syncytial phase of embryo sac development results in the formation of one big cell containing eight nuclei. After the process of their positioning, the syncytium cellularizes. Because features of cell identity are detectable immediately after cellularization, it was suggested that cell fates are programmed even before the process of cellularization starts and that the decision about cell fate is determined by positional information in the syncytium. An auxin gradient is formed along the proximal-distal axis in maturating female gametophyte with higher auxin concentrations on the proximal micropylar pole causing cells to acquire synergid and egg cell fate. Ectopic expression of genes responsible for auxin biosynthesis in the whole female gametophyte strongly affected wild-type hormone distribution causing many egg and antipodal cells to express synergid specific marker (Pagnussat et al., 2009).

The female gametophyte is the haploid generation with its main function to form female gametes and to attract the pollen tube. In *Arabidopsis* the mature FG is reduced to seven cells that are protected by the maternal sporophytic tissues of the ovule and thick walls of the ovary (carpels). Egg cell and central cell are two fertilization competent cells and therefore called gametes. The egg cell together with two synergid cells form the egg apparatus that is located at the micropylar end of the embryo sac. This region is penetrated by the pollen tube upon fertilization. The synergid cells produce an attractant that guides the elongating pollen tube to the FG. Laser ablation of synergid cells in the *in vitro Torenia fournieri* system caused failure of pollen tube attraction and fertilization

(Higashiyama et al., 2001), revealing an essential role of synergid cells in pollen tube attraction. The opposite pole of the embryo sac is occupied by antipodal cells. Whereas *Arabidopsis* and many other flowering plants contain three antipodals, in several grasses the antipodals proliferate resulting in up to several hundred cells. However, their function is not well understood. In *Arabidopsis* antipodal cells undergo programmed cell death prior FG maturation and do not participate in the fertilization process. In contrast, in some other species (e.g. grasses) they persist throughout seed development (Williams and Friedman, 2004).
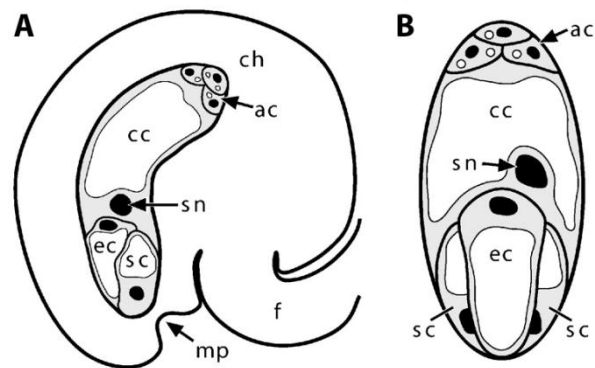
Figure 1-2. Schematic representation of the *Arabidopsis* ovule (A) and female gametophyte (B), adapted from (Drews and Yadegari, 2002). View in (B) is perpendicular to that in (A). Within the female gametophyte, the gray areas represent cytoplasm, the white areas vacuoles, and the black areas nuclei. Abbreviations: ac, antipodal cells; cc, central cell; ch, chalazal region; ec, egg cell; f, funiculus; mp, micropyle; sc, synergid cell; sn, secondary nucleus.

The middle part of the embryo sac is occupied by the homodiploid central cell. The large central vacuole is a specific feature of this cell. The central cell is highly polarized and its nucleus is localized at the micropylar pole, in the proximity to the egg cell, where it meets one of the sperm cells upon fertilization.

1.2. Integument Development

The developing female gametophyte is protected by two ovule integuments. They initiate their growth in a sequential manner from the chalaza, a part of ovule primordium localized between funiculus and nucellus. The inner integument initiates through a series

of cell divisions in the epidermal layer of the chalaza. It forms a ring-like rim that delimits the nucellus, the distal part of the ovule primordium. Subsequently growth of the outer integument is initiated in a similar manner. The outer integument overtakes and covers the inner integument and the nucellus containing the developing FG (Robinson-Beers et al., 1992). Unequal growth of the outer integument causes the ovule to bent, since mutants that lack or are affected in its development (e.g. *inner no outer*) do not bend like wild type ovules (Baker et al., 1997). The resulting anatropous ovule morphology is the most common type among angiosperms, in which the micropyle is adjacent to the funiculus.
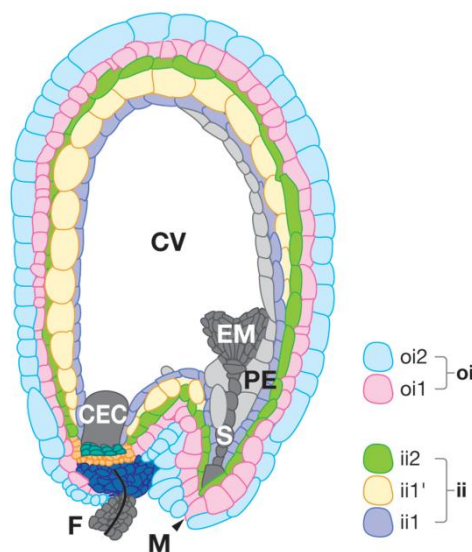
Figure 1-3. Depiction of integuments/seed coat layers adapted from (Lepiniec et al., 2006). Abbreviations: ii1, ii1', ii2, three layers of the inner integument; oi1, oi2, two layers of the outer integument; EM, embryo; CV, central vacuole; F, funiculus; M, micropyle; S, suspensor; CEC, chalazal endosperm cyst; PE, peripheral endosperm; VB, vascular bundle.

Both, inner and outer integuments consist of two cell layers, with the most inner layer of the inner integument being capable of one additional periclinal division. Division of this layer produces another, third layer of the inner integument, so that mature *Arabidopsis* integuments have five cell layers. Their growth continues until the whole nucellus harboring FG becomes enclosed. The growth terminates in the apical end where expanding integuments meet and leave a small opening called micropyle. The micropyle is used during fertilization by the elongating pollen tube to enter the ovule and to reach the

female gametophyte. The bent morphology of the ovule ensures positioning of the micropyle close to the funiculus and placenta where the pollen tube leaves the transmitting track.

The FG does not develop in the *aintegumenta* mutant that fails to form integuments, implicating that integuments are necessary for FG development (Elliott et al., 1996). Conversely, integument growth seems not to require a functional FG as the *sporocyteless* mutant (Yang et al., 1999) and triple mutants of cytokinin receptors *cre1-12* (*CYTOKININ RESPONSE 1*) *ahk2-2tk* (*ARABIDOPSIS HISTIDINE KINASE 2*) *ahk3-3* fail to form a FG but develop integuments (Kinoshita-Tsujimura and Kakimoto, 2011).

Once integument outgrowth and gametophyte maturation are completed, the ovule remains unchanged until fertilization occurs. If fertilization does not occur, *Arabidopsis* ovules will undergo cell death after a few days.

1.3. Double Fertilization

Flowering plants are the most prevalent group of plants and their success is mainly attributed to their mode of reproduction. In angiosperms fertilization occurs in the female gametophyte, which is hidden within the sporophytic maternal tissues. Enclosing the FG by integuments and carpels made it well protected but generated a requirement for the evolution of a special mechanism delivering the two sperm cells to the female gametes. The process of double fertilization has been independently discovered by S. Nawashin and L. Guignard more than hundred years ago (Nawaschin, 1898; Guignard, 1899).

In angiosperms two sperm cells are produced by the male gametophyte, a pollen grain which develops in anthers. The mature male gametophyte is a haploid structure reduced to only three cells of two types: a single vegetative cell and two sperm cells. Both male gametes do not have cell walls and are localized inside the vegetative cell which supports and protects their maturation and also plays a crucial role during fertilization process. Sperm cells of flowering plants are not motile and require a special mechanism that delivers them to the FG. This is ensured by the elongating pollen tube produced by the vegetative cell.

Upon pollination the pollen grain is delivered to the pistil's stigma, the female receptive organ. After a re-hydration process the pollen grain is activated and the vegetative cell starts formation of the pollen tube. Pollen tube elongation is characterized by a fast tip growth guided by maternal factors and directed to the FG. It starts its growth

in the stigma, passes the style and enters the transmitting track inside the placenta. Leaving the transmitting track, the pollen tube enters the ovary chamber (Johnson and Lord, 2006). It grows along the funiculus directly towards the micropyle, where it enters the ovule. There are multiple steps of pollen tube guidance by the female tissues reviewed in (Higashiyama and Hamamura, 2008), but the best understood is the last one that involves a role of the female gametophyte. The pollen tube will arrest growth or lose its orientation in mutant plants affected in synergid cell development (Hulskamp et al., 1995). Consistently, a FG with physically damaged synergids will not attract pollen tubes, suggesting the formation of a short distance signal by synergid cells (Higashiyama et al., 2001).
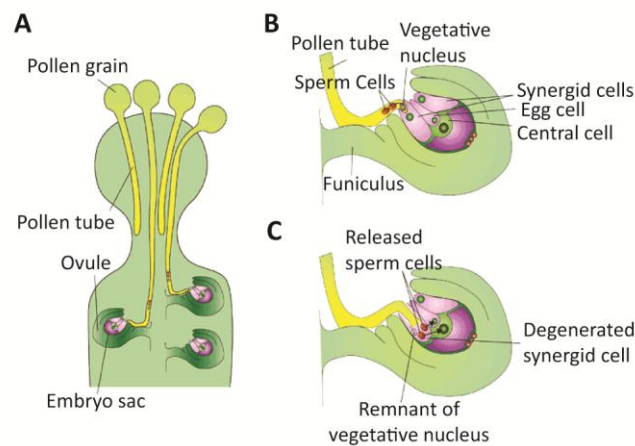
Figure 1-4. Schematic representation of the double fertilization process in *Arabidopsis* adapted from (Twell, 2006). (A) Pollen grains germinate to produce pollen tubes that grow within the pistil. (B) Each pollen tube, containing two sperm cells and a vegetative nucleus, is attracted towards the ovule and is precisely guided to the micropyle. (C) The pollen tube discharges the two sperm cells into one of the two synergid cells of the embryo sac. Double fertilization follows: one sperm cell fuses with the egg cell and the second with the central cell to produce the embryo and the endosperm of the seed.

The growth of the pollen tube is a complex and highly controlled process. Whereas some species allow self-fertilization, in others, growth of the pollen tube can be inhibited when it is derived from the same species. Inhibition can occur either at the very beginning at the recognition step on the stigma or later during its growth in the style. Self-incompatibility prevents self-fertilization what ensures creation of genetically more diverse populations. The evolution of pollen tube growth, a system delivering non-motile

sperm cells to the FG, made the fertilization process independent from the presence of water and allowed angiosperms for wider spreading and colonization of new environments (Chapman and Goring, 2010).

While the pollen tube elongates, the vegetative cell nucleus as well as the two sperm cells are transported towards its tip. During growth of the pollen tube the sperm cells overtake the vegetative nucleus. The pollen tube enters the ovule through the micropyle and penetrates one of the two synergid cells, which degenerates upon pollen tube discharge. When the pollen tube bursts, its content together with the two sperm cells is released inside the FG. The two sperm cells migrate to the area between central cell and egg cell and after a short time (5-7 minutes) they fuse with the female gametes. Male gametes have no preferential targets and one of them fertilizes the egg cell initiating development of the embryo and a second one fertilizes the central cell that develops into the endosperm (Hamamura et al., 2011).

1.4. Seed Development

Reproduction through seeds conferred an enormous advantage to seed plants that contributed to the evolutionary success of angiosperms. Seeds can be easily transported on long distances, their specialized components protect and support the developing embryo and germinating seedling and allow the embryo to remain dormant to ensure the optimal timing for germination (Linkies et al., 2010).

Seed development is a complex process that encompasses simultaneous growth of two fertilization products of different genomic composition and the surrounding maternal tissues of sporophytic origin. Seed development starts after fusion of two genetically equal sperm cells with two female gametes. The fertilized egg cell develops into the embryo, the only component of the seed that contributes to the next generation. Fertilization of the second female gamete, the central cell, leads to development of the endosperm, an ephemeral tissue responsible for transmission of nutrients from the mother plant to the embryo. Double fertilization indirectly also leads to the development of the seed coat that envelopes the two fertilization products and protects them from various hazards (Nowack et al., 2010).
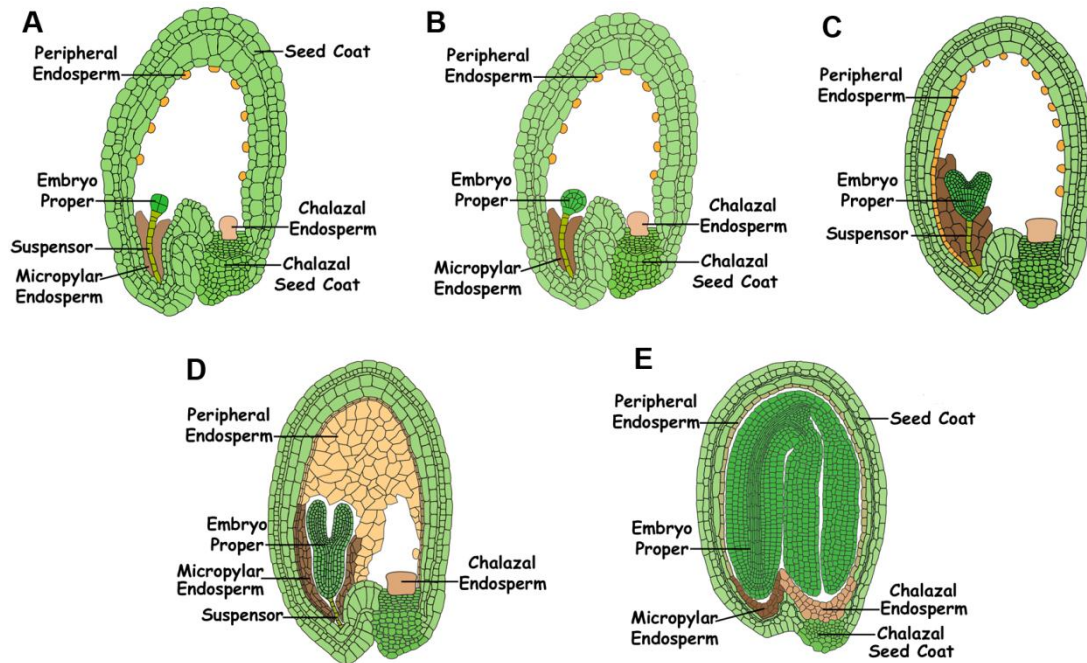
Figure 1-5. Schematic representation of *Arabidopsis* seed development adapted from www.seedgenenetwork.net. Depicted are seed developmental stages that are named based on characteristic steps of embryo development: (A) pre-globular stage; (B) globular stage; (C) heart stage; (D) linear cotyledons stage; (E) bent cotyledons stage.

1.5. Embryo Development

The process of embryogenesis starts when the zygote elongates and undergoes an asymmetrical division generating a small apical cell and a larger basal cell. This first division determines the two cell lineages that follow different developing programs. Divisions of the basal cell result in the formation of a transient suspensor, while the apical cell differentiates into the proembryo.

After two subsequent divisions of the apical cell the two regions are specified again, the apical part will give rise to cotyledon primordia and the shoot apical meristem, whereas the basal part will develop into the hypocotyl. At the globular stage, the uppermost cell of the suspensor is specified to become the hypophysis. Through an asymmetrical division, the hypophysis gives rise to a lens-shaped cell from which the quiescent center will be generated and a basal cell from which stem cells of the root apical meristem and columella cells will be derived (Lau et al., 2010). The outgrowth of cotyledons' primordia causes the embryo to attain a heart shape. Further cell divisions will

lead to growth of the hypocotyl and cotyledons that will bend and eventually fill whole cavity of the seed at its maturity (Peris et al., 2010).
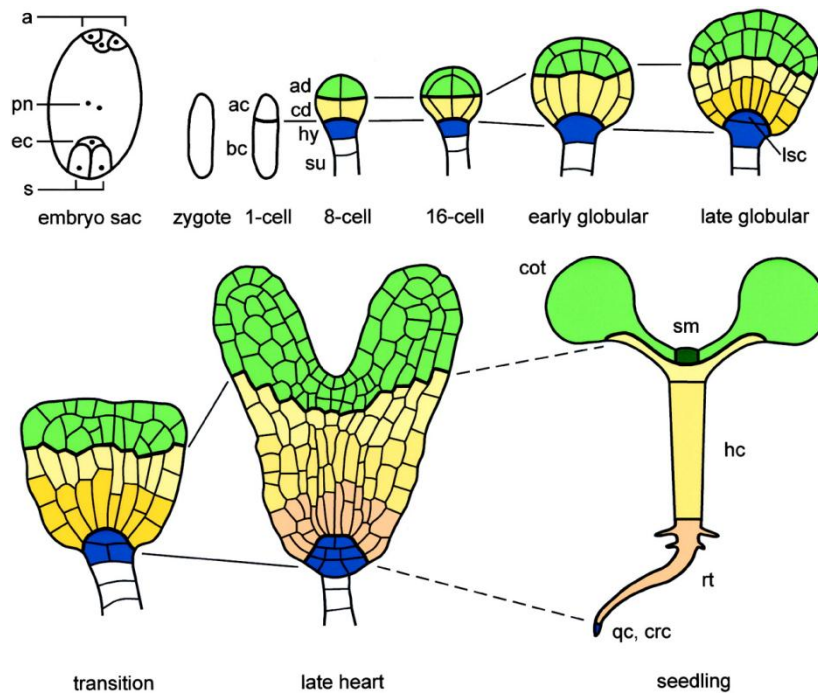
Figure 1-6. Schematic representation of embryo development adapted from (Laux et al., 2004). The upper and lower thick lines represent clonal boundaries between the descendants of the apical and basal daughter cells of the zygote and between the apical and central embryo domains, respectively. Abbreviations: a, antipodals; ac, apical daughter cell; ad, apical embryo domain; bc, basal daughter cell; cd, central embryo domain; cot, cotyledons; crc, central root cap; ec, egg cell; hc, hypocotyl; hy, hypophysis; lsc, lens-shaped cell; pn, polar nuclei; qc, quiescent center; rt, root; s, synergids; sm, shoot meristem; su, suspensor.

Growth of the embryo in most dicots is accompanied by vanishing of the endosperm tissue that is used as a nutritious material necessary for the fast growth of the embryonic tissues and synthesis of storage material like proteins and lipids. Mature seeds of most monocots contain relatively small embryos and the endosperm is used up later during germination. At this stage the mature embryo enters its last phase, the desiccation process that prepares the embryo for dormancy.

1.6. Endosperm Development

The endosperm is the second fertilization product that develops after the fusion of one of the two sperm cells with the central cell. The central cell nucleus originates from the fusion of the two haploid polar nuclei and therefore, contains a double number of chromosomes already before fertilization. Fertilization of the diploid central cell with the haploid male gamete initiates development of the triploid endosperm with two copies of maternal and one copy of paternal genomes.

Development of the endosperm is divided into two distinct phases. It starts with several rounds of nuclear mitotic divisions that are not accompanied by cytokinesis. The enlarging central cell becomes a multinucleated syncytium containing large amounts of free nuclei surrounded by a sphere of denser cytoplasm. In the second phase cell walls are formed giving rise to the cellularized endosperm that for a short time continues to undergo periclinal cell divisions. The cellularized endosperm is a terminal ephemeral tissue most of which is eventually consumed as a nutrient source by the growing embryo. The syncytial phase, specific for nuclear-type endosperm development, lasts for about eight mitotic cycles (Costa et al., 2004). The primary endosperm nucleus starts dividing almost immediately after double fertilization resulting in two daughter nuclei that become separated as one of them migrates towards posterior (chalazal) pole. Each nucleus of the developing endosperm is surrounded by a portion of condensed cytoplasm with a radial microtubular system emanating from the nucleus. These nuclear-cytoplasmic domains (NCDs) are involved in nuclear positioning and control of cell wall formation at the time of cellularization (Brown et al., 1999). Only the first three nuclear divisions occur in a synchronized way. The fourth round passes as a wave of successive mitoses that begins at the micropylar pole and progresses to the chalazal-most nucleus. This divides the syncytium into regions that continue development in a domain-specific manner. Differences in nuclear shape, cytoskeletal arrays and cytoplasmic characteristics mark the differentiation of the syncytium into three developmental domains, the micropylar endosperm at the anterior pole, the chalazal endosperm at the posterior pole and the peripheral endosperm domain in between those two (Boisnard-Lorig et al., 2001; Brown et al., 2003). While cytoplasm in the micropylar and chalazal domains is dense and nearly non-vacuolated, the syncytium in the central chamber consists of a single layer of evenly spaced NCDs surrounding a large central vacuole. Nuclei surrounding the zygote are fusiform, while those in the central and chalazal domains are spherical. While micropylar

and peripheral domains continue with nuclear division, nuclei from the chalazal domain do not divide after the fourth mitosis but become larger as a result of endoreduplication. Additionally, some NDCs from the part of the peripheral domain localized in proximity to the chalazal domain fuse together forming multinucleated NCD structures called nodules. These in turn fuse in the big nuclei of chalazal cyst. The chalazal domain is localized in the proximity of the vascular bundle of the maternal tissue that delivers assimilates to the seed. Since the chalazal endosperm forms multiple specialized extensions that penetrate the adjacent sporophytic tissue it is assumed that it plays an important role in the loading of storage compounds from surrounding maternal tissues into the developing seed (Nguyen et al., 2000).
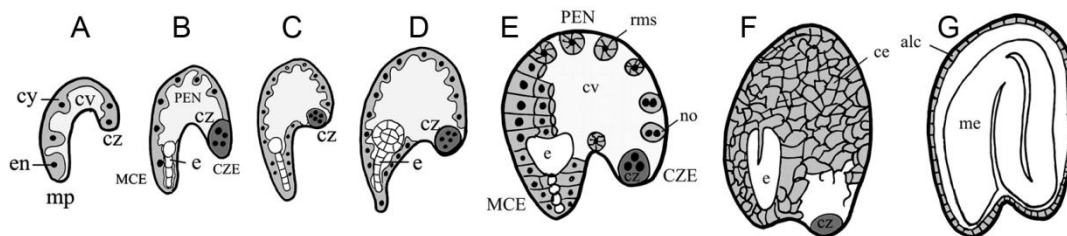

Figure 1-7. Schematic representation of *Arabidopsis* endosperm development adapted from (Olsen, 2004). (A) Dividing endosperm nuclei migrate from the micropylar zone (mp) toward the chalazal end (cz). (B, C) As development progresses, the endosperm develops three distinct regions: the micropylar endosperm (MCE) surrounding the embryo, the central or peripheral endosperm (PEN), and the region of the chalazal endosperm (CZE), which contains the chalazal cyst (cz). (D) At the end of the globular embryo stage, the embryo becomes completely surrounded by cytoplasm. (E) At the embryo heart stage cellularization of MCE occurs, which propagates to the PEN, while endosperm nodules (no) as well as the chalazal cyst (cz) are formed in CZE. (F) Completely cellular endosperm (ce). (G) The endosperm is consumed during seed maturation, leaving only the peripheral aleurone-like cell (alc) layer surrounding the mature embryo (me).

The cellularization process starts after the eighth mitotic cycle. It begins in the area surrounding the embryo at the early heart stage and progresses through the peripheral domain. Cellularization does not affect the chalazal cyst that persists as a syncytium throughout the seed development. At the beginning of the process NCDs elongate and polarize. The microtubular system is rearranged and phragmoplast formation is initiated at the place where two neighboring NCDs interact with each other (Otegui and Staehelin, 2000). Cell wall formation produce open-ended tube-like cells orientated towards the central vacuole. In the last step periclinal cell walls are formed. The cellularizing

endosperm still undergoes periclinal cell divisions to completely fill the peripheral zone. After this complicated process the endosperm tissue is degraded by the growing embryo that takes its place in the mature seed. Only one, the most outer layer of the endosperm, called aleurone layer persists till seed maturity (Brown et al., 1999).

1.7. Seed Coat Development

Following fertilization and divisions in the endosperm the seed coat (testa) starts growth and differentiation from the cells of the ovule integuments. The integuments undergo intensive proliferation during FG maturation before fertilization, resulting in a high number of small densely packed cells. Around anthesis cell divisions decelerate and initial growth of the seed coat after fertilization is attained predominantly due to cell expansion. Seed coat growth is observed already about 1,5 DAP, corresponding to the 16 nuclei stage of endosperm development while the zygote only starts its first division (Brown et al., 2003).

The seed coat consists of five cell layers that follow different developmental programs. Cells of the innermost layer (the endothelium) differentiate at first, marked by the biosynthesis of proanthocyanidins (PA), which are flavanol compounds also known as condensed tannins (Dixon et al., 2005). Their name reflects the fact that, on acid hydrolysis, the extension units are converted to colored anthocyanidins, and this forms the basis of the classical vanillin-HCl assay for the detection of these compounds. PAs accumulate in the central vacuole of endothelium cells that in the later stages of seed growth are crushed by the growing embryo. After release PAs oxidize, turn brown and impregnate the other seed coat layers (Pourcel et al., 2005). The PAs biosynthesis pathway is very well described, which is mainly attributed to the ease in screening for mutants with colorless seeds. Many mutants affected in this trait were isolated and collectively called *transparent testa* (*tt*) mutants. Genes that are affected in these mutants can be generally divided into two groups. One that encodes enzymes and proteins involved in compartmentation of the flavonoid compounds (*TT3, TT4, TT5, TT6, TT7, BAN, TT12, TT18* and *TT19*) and a second group that consists of transcription factors regulating these processes (*TT1, TT2, TT8, TT16, TTG1* and *TTG2*) (Winkel-Shirley, 2001). Endothelium specific TT2 belongs to the MYB transcription factors family and TT8 is a basic helix-loop-helix (bHLH) transcription factor important for PAs biosynthesis but also operating in the epidermal layer of the seed coat. TRANSPARENT TESTA GLABRA 1, a WD40 repeat protein,

forms an endothelium specific complex together with TT2 and TT8 that activates expression of target genes (Baudry et al., 2004). TT16 is another endothelium specific factor. This MADS-box protein is required to initiate PA biosynthesis most probably through its role in specification of the endothelium cells fate, since the most inner seed coat layer in the *tt16* mutant does not differentiate properly (Nesi et al., 2002).

The two outer layers of the inner integument do not follow the endothelium developmental program but undergo programed cell death. Cell death in these two cell layers is delayed upon loss of δVPE function. The δVPE gene encodes a Caspase-1-like protein that belongs to the family of vacuolar processing enzymes and is specifically expressed in the two outer layers of the inner integument (Nakaune et al., 2005).

The two layers of the outer integument start their differentiation with accumulation of starch containing amyloplasts during the initial growth period. During further development they acquire different fates. The inner layer produces a thick secondary cell wall making it resistant to the growing embryo that presses against the seed coat. The most outer, epidermal layer produces large amounts of a pectinaceous carbohydrate, the mucilage. The synthesis of mucilage takes place in the Golgi apparatus and is later secreted into the apoplast. The cytoplasm moves apart from the cell walls as more mucilage is accumulated and forms a column in the center of the cell. A secondary cell wall separating the cytoplasm from the mucilage is produced and the whole structure with a central cell wall elevation is called columella (Western et al., 2000). Similar to endothelium, the cell fate of the epidermis is under control of a transcription factor complex that consists of the WD40 repeat protein TTG1, the MYB family transcription factor MYB5 and the basic helix-loop-helix protein TT8 (Li et al., 2009). APETALA 2 is another transcription factor that is involved in epidermis development. Lack of AP2 affects the whole outer integument that does not form columella structures and neither produces mucilage (Jofuku et al., 1994). It also affects expression of many transcriptional regulators involved in epidermal development (Western et al., 2004), suggesting that AP2 plays a similar role in the outer integument as TT16 in the endothelium layer.

The developing seed coat consists of five cell layers that during development differentiate and change their structure dramatically. While the seed matures also the embryo progresses in its development, effectively filling the whole seed and eventually pressing on the seed coat. The structure of the epidermal cells is preserved by mucilage and the columella while other layers are crushed and the released tannins color the seed

brown. The seed coat physically protects the embryo, provides impermeability to water and oxygen, promotes stable dormancy and supports seed dispersal. When the seed gets into contact with water seed imbibition takes place, mucilage is extracted and a gel capsule is formed around the seed that is thought to aid hydration.

1.8. Cross-talk of the Endosperm with the Seed Coat

Co-ordinated growth of the three seed components embryo, endosperm and seed coat is necessary for the formation of viable seeds in angiosperms. These three genetically different components interact together to attain a seed size that ensures best chances for successful germination. Endosperm growth starts almost immediately after double fertilization and is correlated with enlargement of the seed coat. These two components during the first few days after fertilization determine the final size of the mature seed (Sundaresan, 2005). During this time the embryo has reached only the heart stage, indicating that rather the endosperm together with the seed coat influence the size of the embryo than opposite.

Seed coat growth takes place in two steps: an initial phase of cell proliferation, as a continuation of pre-fertilization integuments growth and strong cell elongation after fertilization (Western et al., 2000; Garcia et al., 2005). In the *haiku2* (*iku2*) and *miniseed3* (*mini3*) mutants elongation of the seed coat cells is strongly affected, causing seeds of these mutants being much smaller than wild-type seeds. Investigation of the mutants demonstrated that mutations do not affect seed coat development directly. *MINI3* and *IKU2* genes are expressed in the endosperm and embryo but not in the cells of the seed coat. The mutations are recessive and strongly delay growth of the endosperm in its early developmental stages causing its precocious cellularization and reduced embryo proliferation (Luo et al., 2005). The reduced endosperm is surrounded by a smaller seed coat. Since the number of cells in the integuments does not differ in mutants and wild-type plants, the effect is attributed to a reduced expansion of seed coat cells (Garcia et al., 2003; Luo et al., 2005). Based on these studies it has been suggested that seed coat cell elongation is a mechanism for seed coat size adjustments. Moreover, it implicates the existence of a signal from the endosperm that triggers cell elongation in the integument.

The *transparent testa glabra 2* (*ttg2*) mutant is another mutant that influences seed size. The *ttg2* mutation affects differentiation of the seed coat and directly affects its

size (Johnson et al., 2002; Garcia et al., 2005), implicating TTG2 acting downstream of an endosperm-mediated effect on seed coat growth. Similar to the *mini3* and *iku2* mutation, *ttg2* does not affect cell proliferation but cell elongation. While in *mini3* and *iku2* the abnormally developing endosperm does not induce expansion of the seed coat, in the *ttg2* mutant endosperm growth is restricted by reduced growth of the seed coat. This suggests that the maternally derived seed coat directly impacts on the decision about seed final size.

1.9. Polycomb Group Proteins

Polycomb group (PcG) proteins are chromatin associated factors involved in the regulation of transcriptional activity of the targeted loci. Originally, PcG proteins have been characterized in *Drosophila melanogaster* as factors controlling proper spatial expression of homeotic (*Hox*) genes. The *Hox* genes encode for transcription factors involved in cell fate determination along the anterior-posterior body axis. Their early activation in specific body segments establishes expression pattern that are maintained in later developmental stages. Stable repression of *Hox* genes outside their expression domains is executed by PcG proteins (Jürgens, 1985). Mutants deficient in PcG proteins express *Hox* genes outside their assigned domains, causing severe alterations of the fly body plan. The Polycomb (Pc) gene was the first identified PcG gene, and the *pc* mutant is characterized by the appearance of ectopic sex combs on the second and third pair of legs. Recent studies have revealed that PcG proteins bind many more genes in addition to the *Hox* genes. Among these target genes are mainly transcription factors involved in diverse cellular functions and developmental pathways like cell cycle control and cellular senescence, genomic imprinting, stem cell plasticity and cancer (Schuettengruber and Cavalli, 2009). In *Drosophila,* PcG genes form two main complexes with distinct functions: Polycomb Repressive Complex 1 (PRC1) and PRC2. PRC2 consists of the four core components Enhancer of zeste [E(z)], Extra sexcombs (Esc), Suppressor of zeste 12 [Su(z)12] and nucleosome remodeling factor (Nurf55). The catalytic subunit, E(z), is a SET (Supressor of variegation, Enhancer of zeste and Trithorax) domain-containing methyltransferase that catalyzes the di- and trimethylation of lysine 27 on histone H3 (H3K27me3). H3K27me3, the hallmark of PcG-dependent gene silencing, is specifically recognized by the chromodomain of Pc, one of subunits of the PRC1 complex (Cao and

Zhang, 2004). The *Drosophila* PRC1 complex is composed of Pc, Polyhomeotic (Ph), Posterior sex combs (Psc) and dRing, also called Sex combs extra (Sce) (Levine et al., 2002). The Sce is a ubiquitin E3 ligase that monoubiquitinates lysine 119 of histone H2A. This modification is in addition to H3K27me3 also recognized as a mark for transcriptional silencing (Wang et al., 2004) and the whole PRC1 is believed to mediate the silencing through chromatin compaction (Francis et al., 2004). A third important complex Pleiohomeotic Repressive Complex (Pho-RC) was identified only in *Drosophila*. It consists of the DNA-binding protein Pho and Scm-related protein containing four malignant brain tumor domains (Sfmbt) that bind mono- and dimethylated H3K9 and H4K20 (Klymenko et al., 2006). The PcG silencing machinery is brought to the target loci through the interaction with Polycomb Responsive Elements (PREs). A PRE is a non-conserved DNA sequence that consists of binding sites for several different DNA-binding proteins like Pho, Pleiohomeotic-like (Phol), GAGA factor (GAF), Pipsqueak (Psq), Zeste and Grainy head (Grh) (Müller and Kassis, 2006). These factors mediate the interaction of specific DNA sequences with PRC1 and PRC2 complexes that do not directly interact with DNA and therefore do not recognize PREs. It is still not clear how the targeting occurs, because mutants in PRE binding proteins do not show PcG mutant phenotypes and many of them are involved in transcriptional activation as well as repression (Müller and Kassis, 2006). Genome-wide studies of PcG proteins and H3K27me3 localization revealed that while PcG proteins bind chromatin very localized to PREs, H3K27me3 is more broadly distributed surrounding targeted loci (Schuettengruber and Cavalli, 2009). The "looping model" suggests binding of the PcG machinery to confined PRE sequences and modification of histones in its close proximity. As a result of changes in the three dimensional chromatin structure (looping) nucleosomes being distantly located from PREs become localized into the operating zone of the PRC2 SET domain subunit (Kahn et al., 2006).

PcG proteins and their function in transcriptional gene repression are conserved between insects, vertebrates and plants. Table 1-1 lists *Drosophila* PcG genes and their homologs in mammals and *Arabidopsis*.

Table 1-1. PRC2 core subunits in *Drosophila*, mammals and *Arabidopsis*

Drosophila protein	**Mammalian protein**	**Protein motif**	*Arabidopsis* protein
E(z)	EZH1/EZH2	SET domain	CLF SWN MEA
Su(z)12	SUZ12	Zn finger VEFS domain	EMF2 VRN2 FIS2
Esc	EED	WD repeats	FIE
Nurf55	RbAp46/48	WD repeats	MSI1

1.10. PRC2 in *Arabidopsis*

As a result of genome duplications the number of genes encoding subunits of the *Arabidopsis* PRC2 complex strongly increased. Homologs of PRC2 subunits acquired roles in distinct expression domains. Therefore, different PRC2 homologs control different, although sometimes overlapping target genes. So far, three different PRC2 complexes were identified in *Arabidopsis*: the VERNALIZATION (VRN) complex, the EMBRYONIC FLOWER (EMF) complex and the FERTILIZATION INDEPENDENT SEED (FIS) complex. Complexes are named after Su(z)12 homologs VERNALIZATION2 (VRN2), EMBRYONIC FLOWER2 (EMF2) and FERTILIZATION INDEPENDENT SEED2 (FIS2) that are PRC2 subunits unique in the given complex. Following the *Drosophila* and mammalian model the core of each *Arabidopsis* PRC2 complex comprises four proteins. MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) and FERTILIZATION INDEPENDENT ENDOSPERM (FIE) are two subunits with no other homologs in the *Arabidopsis* genome, involved in PRC2 related function. They therefore participate in the formation of every of the three PRC2 complexes. Although *MSI1* is a member of a small family of five homologous genes, there are no reports showing MSI2-5 proteins being involved in PRC2 complex formation. MSI1 and FIE are

homologs of *Drosophila* Esc and Nurf55, respectively. Homologs of the *Drosophila* E(z) methyltransferase in *Arabidopsis* are CURLY LEAF (CLF), MEDEA (MEA) and SWINGER (SWN).

The FIS and the VRN complexes have been biochemically characterized (Köhler et al., 2003a; De Lucia et al., 2008). Like in *Drosophila*, core components of the FIS and VRN PRC2 were part of 600 kDa protein complexes (Tie et al., 2001). De Lucia and colleagues identified three plant homeodomain (PHD) finger proteins being part of a common PHD-PRC2 complex. Lack of PHD proteins VERNALIZATION 5 (VRN5), VERNALIZATION5/VIN3-LIKE 1 (VEL1) and VERNALIZATION INSENSITIVE (VIN3) leads to reduced H3K27me3 levels, suggesting that PHD proteins are required for high levels of H3K27 trimethylation that in turn are required for stable epigenetic repression of target genes. Similarly, in *Drosophila* and mammals PRC2 complexes interact with Polycomblike (Pcl) and human PHF1 PHD proteins, respectively, and this interaction is required for the generation of high levels of H3K27me3 (Nekrasov et al., 2007; Cao et al., 2008).

Both, the EMF and VRN complexes are important regulators of flowering time in *Arabidopsis*. This developmental transition is an important step in the plant life cycle and involves a switch of the vegetative shoot apical meristem (SAM) to an inflorescence SAM. The decision to flower is taken when the mature plant has reached the developmental competence to react to environmental conditions that ensure successful reproduction. The EMF complex prevents precocious flowering of juvenile plants. Strong *emf2* mutant alleles start flowering right after germination, producing minute plants that do not produce seeds (Yoshida et al., 2001). Consistently, flower meristem identity genes *APETALA 2* (*AP2*) and *AGAMOUS* (*AG*) are deregulated and ectopically expressed in *emf2* mutant seedlings shortly after germination (Chen et al., 1997). The histone methyltransferase CLF is most likely the catalytic subunit of the EMF complex. Although the whole complex has not yet been purified, phenotypic similarities of *clf* and *emf2* mutants as well as the shared subset of deregulated genes and interaction of CLF and EMF2 proteins in the yeast two-hybrid system strongly support this conclusion (Chanvivattana et al., 2004).

PcG proteins are furthermore required to maintain cell identity in plants. Loss of CLF and EMF2 function causes homeotic transformations of floral organs, e.g. fused sepals with stigmatic papillae cells along their margins, stamenoid petals and anther-like tissue on the surface of the petal blades. Complete loss of PcG function in the *clf swn* double

mutant causes formation of callus-like structures producing somatic embryos, highlighting the requirement of PcG function to maintain cell identity (Goodrich et al., 1997; Chanvivattana et al., 2004).

A second PRC2 complex controlling flowering time is the VRN complex that enables *Arabidopsis* plants to flower after vernalization. Vernalization is the process of acquiring the competence to flower after extended periods of exposure to cold. In climatic zones with seasonal changes vernalization will cause plants to flower after winter, ensuring high chances to set the seeds on time, before the next cold season comes. In the *Arabidopsis* genus there are winter-annual accessions that require vernalization to flower, as well as the summer-annual accessions that start the reproductive program independently of vernalization (Shindo et al., 2006). The molecular mechanism of vernalization relies on stable epigenetic silencing of the flowering repressor *FLOWERING LOCUS C* (*FLC*). The H3K27me3mark in the coding region of *FLC* is deposited by VRN complex and correlates with its repressed state. FLC is a repressor of the flowering promoter *FLOWERING TIME* (*FT*). Therefore, repression of *FLC* will cause *FT* activation and induction of flowering. Exposure to cold initiates methylation only in a very restricted region of the first intron of *FLC.* Extended periods of cold exposure are a requirement for stable *FLC* silencing, and will enable a spreading of H3K27me3 over the *FLC* locus after return to warm conditions. Surprisingly, binding of VRN2 to the *FLC* locus seems independent of vernalization, as VRN2 binding has been detected all over the *FLC* locus before and after vernalization. Activity of the VRN complex depends on PHD proteins VRN5, VEL1 and VIN3. Expression of *VIN3* gradually increases during cold exposure, suggesting that *VIN3* could be the output of a temperature sensing mechanism. The core of the VRN complex consists of VRN2, SWN or CLF, FIE and MSI1 but formation of a bigger and likely active PHD-PRC2 complex depends on the presence of PHD proteins (Wood et al., 2006; De Lucia et al., 2008). Additionally, also the CLF-EMF2 PRC2 complex has been shown to act on the *FLC* locus to repress its activity under warm conditions (Jiang et al., 2008).

The third PRC2 complex in *Arabidopsis*, the FIS complex, differs from the two others not only in composition but also in the domains of its activity as it is not expressed in sporophytic but only in gametophytic tissues.

1.11. The FIS Complex

The first components of the FERTILIZATION INDEPENDENT SEED (FIS) complex have been discovered in genetic screens aiming to identify mutants that are able to form seeds in the absence of fertilization (Ohad et al., 1996; Chaudhury et al., 1997). The production of seeds uncoupled from the fertilization process is a characteristic feature of apomixis. However, whether loss of FIS function is a requirement for apomictic seed formation is not clear. The FIS complex comprises the four core subunits FIS2, MEA, FIE and MSI1. Mutants with a defect in any subunit show characteristic phenotypes before as well as after the fertilization, indicating that the FIS complex plays important roles in the gametophyte and in the developing seed. Indeed, FIS subunits are expressed in the central cell of the female gametophyte and in the developing endosperm after the fertilization. In the absence of fertilization the central cell of *fis* mutants starts dividing and develops into an asexual endosperm. This autonomously developing endosperm differentiates and forms a chalazal cyst, peripheral endosperm occasionally cellularize but autonomous *fis* seeds eventually collapse (Chaudhury et al., 1997). Autonomous seeds of *fis2*, *mea* and *fie* mutants do not contain an embryo, suggesting that the FIS complex does not repress autonomous egg cell development. The *msi1* mutant is exceptional in its ability to form autonomous embryo like structures. However, MSI1 is also a component of other chromatin associated complexes, suggesting that autonomous embryo formation is rather a consequence of deregulated processes unrelated to FIS function (Köhler et al., 2003a). Together, these data strongly suggest that the FIS PRC2 complex acts in the central cell nucleus to suppress endosperm development in the unfertilized female gametophyte.

Impaired FIS function also causes developmental abnormalities after fertilization. While wild-type seeds initiate endosperm cellularization at about 5 DAP containing heart stage embryos, endosperm of *fis* mutants does not cellularize but overproliferates and *fis* embryo development arrests at heart stage. Since *FIS* genes are not expressed in the embryo, embryo arrest is likely a consequence of a failure in endosperm function (Chaudhury et al., 1997). *Fis* mutants are maternal effect gametophytic mutants; thus, every seed that inherits a *fis* mutant allele from the mother will arrest development, independent of the presence of a wild-type paternal allele. The female gametophytic effect of *fis* mutants is unlikely a consequence of maternal-specific expression of *FIS* genes. While *FIS* genes *MEA* and *FIS2* are indeed only expressed from maternally inherited alleles

(Luo et al., 2000; Vielle-Calzada et al., 2000; Kinoshita et al., 2001), *FIE* and *MSI1* are biallelically expressed but still reveal a maternal effect (Leroy et al., 2007). Therefore, it seems likely that abnormal endosperm development in *fis* mutants is a consequence of impaired FIS function in the central cell.

Transcription profiling of the *fis* mutants *mea* and *fie* led to the identification of *PHERES1* (*PHE1*), the first direct target gene of the FIS complex (Köhler et al., 2003b). Recently, H3K27me3 profiling performed on isolated endosperm nuclei identified 1800 FIS PRC2 target genes, among them 240 genes being specifically targeted only in the endosperm. FIS PRC2 targets were highly enriched for MADS-box transcription factors as well as genes involved in cell wall formation, suggesting a specific function of the FIS PRC2 in endosperm cellularization (Weinhofer et al., 2010).

Interestingly, the repressive action of the FIS PRC2 complex can be very selective. Target genes of FIS PRC2 *PHE1* and *MEA* are among the first examples of imprinted genes in plants. Genomic imprinting is an epigenetic phenomenon causing allele-specific gene expression dependent on the parent-of-origin. The paternal allele of *PHE1* and the maternal alleles of *MEA* are active, while repression of the maternal *PHE1* alleles and the paternal *MEA* allele is mediated by the FIS complex (Köhler et al., 2005). Genomic imprinting occurs in flowering plants and mammals and is considered to be a consequence of the "placental habit". The progeny of flowering plants and mammals is nourished exclusively by the mother through specialized endosperm and placenta tissues, respectively. Based on the parental conflict theory there is a discrepancy in the maternal and paternal interests over the distribution of maternal resources to the offspring. Whereas it is advantageous for the paternal parent to maximize available resources for his own progeny, it is advantageous for the maternal parent to distribute resources equally among her offspring (Haig and Westoby, 1989). This suggests that paternally expressed genes will promote embryo growth, while maternally active genes will cause the opposite effect. Analysis of many imprinted genes in mammals support this theory (Shirley M, 1999). Experiments with interploidy crosses in plants also support the parental conflict theory. Increased dosage of the paternal genome in diploid (2n) x tetraploid (4n) crosses results in enhanced growth of the endosperm and in turn enlarged seeds, whereas the reciprocal crosses produce seeds of smaller size (Scott et al., 1998). Seeds resulting from crosses with increased paternal genome contribution phenocopy fis-class mutant seeds, implicating a role of FIS PRC2 in balancing parental genome dosages (Erilova et al., 2009).

1.12. Partial Redundancy of PcG Proteins in *Arabidopsis*

Several *Arabidopsis PcG* genes can be grouped into families based on sequence homology, such as *CLF*, *MEA*, *SWN* and *EMF2*, *VRN2*, and *FIS2*. It is likely that these gene families are the result of gene duplications and subsequent diversification from ancestral sequences. The analysis of homologous sequences in plants indicates the existence of *EMF2*-like gene in early diverging lineages of land plants, suggesting that the *EMF2* gene is the ancestral gene, whereas *VRN2* arose in angiosperms after *EMF2* duplication. *EMF2* shares higher homology with *VRN2* than with *FIS2. FIS2* has only been found in few plant genomes and only in those that contain *VRN2,* suggesting that *FIS2* has evolved after *VRN2* duplication (Chen et al., 2009). A characteristic feature of this gene family is the VEFS (VRN2, EMF2, FIS2, Su(Z)12) domain that is responsible for the interaction with the catalytic subunit (CLF, SWN, MEA or E(Z)) of a given complex (Chanvivattana et al., 2004). Yeast two-hybrid studies revealed that *Arabidopsis* homologs of Su(Z)12 can interact with more than one SET domain protein. FIS2 interacts in yeast with MEA and SWN (Wang et al., 2006), and EMF2, VRN2, and FIS2 interact with SWN and CLF (Chanvivattana et al., 2004), implicating partial redundancy between PRC2 family members. Indeed, genetic interactions confirm this view. Whereas the *clf* mutant phenotype can be strongly enhanced by the *swn* mutant, a similarly strong mutant phenotype effect can be observed in a *emf2 vrn2* double mutant, revealing partial redundancy between CLF and SWN and EMF2 and VRN2 (Chanvivattana et al., 2004; Schubert et al., 2005). Wang and colleagues reported functional redundancy of MEA and SWN. While *MEA* is expressed only in the reproductive tissues of the mature female gametophyte and developing endosperm, *SWN* is expressed in both reproductive and vegetative tissues of the plant. This temporal co-expression, the enhancement of the FIS phenotype in the *mea swn* double mutant and a biochemical proof for the interaction of FIS2 with MEA and SWN supported the hypothesis of redundancy between MEA and SWN (Wang et al., 2006).

If indeed plant PRC2 subunits act redundantly raises the question how specificity of the complexes is determined. One possibility is that specificity is a consequence of differences in the spatial and temporal expression of PcG genes. Alternatively, it is possible that redundancy of PcG genes is not complete. In agreement with this view, most of the PcG single mutants show a phenotype in the presence of the predicted redundant protein.

Furthermore, complementation of the *clf* mutant with 35S driven *CLF*, *SWN* or *MEA* revealed that despite genetic redundancy of CLF and SWN only the CLF sequence could complement the *clf* mutation (Chanvivattana et al., 2004). However, a detailed understanding of this question requires further investigations.

2. AIM OF THE THESIS

The process of seed development encompasses coordinated growth of the three seed components: embryo, endosperm and seed coat. While development of the two fertilization products embryo and the endosperm has been widely studied, the process of seed coat initiation remains poorly understood. The aim of this thesis was to investigate whether the endosperm acts as a driving force for seed coat development. To study the molecular bases of this process I took an advantage of the FIS complex mutants that form autonomous endosperm and display different seed coat phenotypes.

3. MATERIALS AND METHODS

3.1. Plant Material and Growth Conditions

Seeds were surface-sterilized in 5 % (v/v) sodium hypochlorite (Sigma-Aldrich), 0.01 % (v/v) Triton X-100 (Fluka) for 10 min and plated to MS medium (0.4 % (w/v) MS Basal Salt Mixture (Duchefa, cat. M0221), 1 % (w/v) sucrose, 0.8 % (w/v) Bacto Agar (Difco), pH 5.6). After stratification at 4 °C in the dark for two days, seedlings were grown in a growth room under daily cycles of 16 h light and 8 h dark at 21 °C. Ten-day-old seedlings were transferred to soil and plants were grown in a growth chamber at 60 % humidity and a 16 h light and 8 h dark cycle at 21 °C.

Depending on the corresponding mutant line, either Landsberg *erecta* (L*er*) or Columbia (Col-0) plants have been used as wild-type accessions.

All the mutant lines used in this study are listed in Table 3-1.

Table 3-1. List of mutant lines used in this study

Gene	Locus ID	Allele	Accession	Type	Collection	Published
FIS2	AT2G35670	*fis2-1*	L*er*	EMS	-	(Chaudhury et al., 1997)
FIS2	AT2G35670	*fis2-5*	Col	T-DNA	SALK_009910	(Alonso et al., 2003)
FIE	AT3G20740	*fie-2*	L*er*	EMS	-	(Chaudhury et al., 1997)
FIE	AT3G20740	*fie-12*	Col	T-DNA	GK_362D08	(Wolff et al., 2011)
MEA	AT1G02580	*mea-1*	L*er*	Ds	-	(Grossniklaus et al., 1998)
MEA	AT1G02580	*mea-8*	Col	T-DNA	SAIL_55_C04	(Shirzadi et al., 2011)
MSI1	AT5G58230	*msi1-1*	Col	T-DNA	GARLIC 429 B8	(Köhler et al., 2003a)
EMF2	AT5G51230	*emf2-5*	L*er*	EMS	-	(Yang et al., 1995)
VRN2	AT4G16845	*vrn2-1*	L*er*	EMS	-	(Chandler et al., 1996)
SWN	AT4G02020	*swn-3*	Col	T-DNA	SALK_050195	(Chanvivattana et al., 2004)
KPL	AT5G63720	*kpl-2*	L*er*	Ds	GT_5_38351	(Ron et al., 2010)
AGL62	AT5G60440	*agl62-2*	Col	T-DNA	SALK_022148	(Kang et al., 2008)
TTN2	unknown	*ttn2-1*	Ws	T-DNA	CS84718	(Liu and Meinke, 1998)
PHE1	AT1G65330	*phe1-1*	L*er*	Ds	ET189	(Köhler et al., 2005)

Primers used to genotype T-DNA insertion mutants were designed using the T-DNA primer design tool at http://signal.salk.edu/tdnaprimers.2.html (Table 3-2). Mutant combinations were obtained by crossing the appropriate alleles and screening F2 and F3 plants for the desired heterozygous and homozygous mutant combinations.

Table 3-2. Primers used for genotyping of mutant alleles. For T-DNA-specific PCR the T-DNA primer and the reverse primer were used. Sequence from 5' to 3'.

Locus ID	Allele	Primers
AT2G35670	*fis2-1*	fwd: ATGATGAAAATGTATCATCGACACCAAG rev: ACCGCTCTGCATGTAACTCTTTTCT mutation removes Bsu36I site in amplicon
AT2G35670	*fis2-5*	fwd: TGTTGTTTCCATGATTTCTTTTTC rev: AAACCGAACCAGTTTTCATACC T-DNA: ATTTTGCCGATTTCGGAAC
AT3G20740	*fie-2*	fwd: ATGTAGATTACTCTGTACAATTGTCTCG rev: TGCGAGCACAGATTATACAA mutation removes BseRI site in amplicon
AT3G20740	*fie-12*	CACGGTAAGTTGGGCGTGTGGCG T-DNA: CCCATTTGGACGTGAATGTAGACAC
AT1G02580	*mea-1*	fwd: GTGGAGATGGCACTCTTGG rev: CGCATGTTCTGGTCCATAG DS: CCGTTTACCGTTTTGTATATCCCG
AT5G51230	*emf2-5*	fwd: CGTTTCTCCTAAGCCTGTCG rev: TGGTAATGCCCAAAGATACATAAC mutation creates additional MseI site in amplicon
AT4G16845	*vrn2-1*	fwd: TGCGTTCATTAAGTAGGCAACAGAAAATGG rev: GAGAAGTAGTTACCTTTGTTTTCTTACAGAAGAGT mutation creates XmnI site in amplicon
AT4G02020	*swn-3*	fwd: CGTTTCCGAGGATGTCATTGTG rev: TGGAACTTTTGAGTGGCTAGAGGTG T-DNA: GCGTGGACCGCTTGCTGCAACT
AT5G63720	*kpl-2*	rev: GTTACCGACCGTTTTCATCC DS: CCAACTTCACACGTCCTCTTTTGGG
AT5G60440	*agl62-2*	fwd: TGAGCTTTGCACACTTTGTGGTG rev: AAGCATTGTTTCCAAAGGGTGG T-DNA: ATTTTGCCGATTTCGGAAC

3.2. Generation of Transgenic Lines

The *PHE1pro:PHE1-EGFP* construct was prepared by ligating three separate DNA fragments: *PHE1* promoter and coding sequence, *EGFP* reporter gene and 3'UTR region of the *PHE1* gene, together with modified pCAMBIA1300 (http://www.cambia.org) plasmid. All the restriction enzymes and necessary buffers have been purchased from New England Biolabs or Fermentas and used as recommended by the manufacturer. EcoRI and XmaI enzymes were used to obtain 3.8 kb fragment of *PHE1* promoter and coding sequence from already existing plasmid CK297. XmaI and BstEII were used to obtain 740 bp long fragment of *EGFP* gene from existing CK312 plasmid. BstEII and PstI enzymes were used to digest CK297 plasmid to obtain 3.5 kb long fragment of *PHE1* gene 3'UTR region. Original 35S promoter from the pCAMBIA1300 plasmid was removed with EcoRI and PstI restriction enzymes and resulting ends were dephosphorylated. All fragments were ligated with the plasmid in a single reaction. Resulting construct was tested for proper orientation of the fragments by restriction digestion tests and sequencing. Correct clone was used for *Agrobacterium tumefaciens* (GV3103 strain) mediated transformation of *Arabidopsis* plants. Primary transformants were selected on MS plates supplemented with 25mg/L Hygromycin B.

The *PHE1pro:PHE1-EGFP* line has been widely used in our group. E.g. Dr. Isabelle Weinhofer with colleagues have successfully used this reporter line for FACS-based sorting of *Arabidospsis* endosperm nuclei (Weinhofer et al., 2010).

Binary destination vector pB7WG2 was used to generate *FIS2pro:FIS2*, *FIS2pro:VRN2* and *FIS2pro:EMF2* plasmids. The 1850bp of the *FIS2* promoter were amplified with primers containing SacI and XbaI sites (5' GAGCTCCGAATTCGCTGAGAGTTGGTC 3' and 5' TCTAGACTGCTTGATTAATCTATAAGCTGT 3', respectively) and cloned into pB7WG2 replacing the 35S promoter. The *FIS2* coding sequence was amplified (fwd 5' <u>CACC</u>ATGGCTAGGAAGTCCATACG 3' and rev 5' TTATTCATCAACTTCCATAGATTGTTGATTG 3') from L*er* seed cDNA and inserted into pENTR-D TOPO vector (Invitrogen). Entry clones for *VRN2* and *EMF2* (kindly provided by Dr. Lars Hennig) were generated by cDNA amplification and insertion into pENTR-D TOPO vector. Expression clones were generated by LR reaction following the Gateway cloning protocol (Invitrogen). Transgenic lines were generated by *Agrobacterium tumefaciens*–mediated transformation into *fis2-1*.

The *AGL62pro:AGL62-GFP* reporter line was kindly provided by Dr. Elisabeth Hehenberger. It was generated by inserting 2500 bp of amplified *AGL62* promoter and coding region fused to *EGFP* into pCAMBIA1300. Construct was transformed into the heterozygous *agl62-2* mutant line and independent, single transgene locus insertion T2 lines in *agl62* and wild type background were analyzed. The *AGL62pro:AGL62-GFP* in wild type was crossed with *fis2-5*.

The *VRN2pro:VRN2:GUS* reporter line was kindly provided by Caroline Dean. The translational fusion is on a 7.75 kb EcoRI-SpeI fragment. This includes the genomic sequence from the 5' UTR to the 3' UTR of the *VRN2* locus. The *VRN2* stop codon was mutated to a BamHI site and the GUS coding sequence was inserted at this position as a BamHI fragment (Choi et al., 2009).

The *SWNpro:SWN-YFP* translational fusion was constructed by amplifying a 7.2-kb genomic fragment containing 2.2 kb of 5'-flanking sequence and the coding region of *SWN* from Col-0 genomic DNA. The binary transformation vector for the C-terminal fusion to yellow fluorescent protein (YFP) coding sequence was constructed using pBI101, where the GUS coding region was replaced by the multiple cloning site of pUC19 and subsequently used to clone the YFP coding sequence (Wang et al., 2006).

3.3. Standard Molecular Protocols

Standard molecular biology procedures (e.g. restriction enzyme digestions, ligation reactions, plasmid DNA-preparations and gel electrophoresis) were performed as described in (Sambrook and Russell, 2001) or according to the enzymes and kit suppliers' instruction. Cloning PCRs were performed with Expand® Long Template PCR System (Roche), the Gateway cloning system was obtained from Invitrogen. Standard PCR was performed using DreamTaq™ buffer, 0.2 mM dNTP mix, 0.1 µM primers and 0.25 u DreamTaq™ DNA polymerase (Fermentas, cat. EP0702).

DNA was extracted based on the method described by (Edwards et al., 1991). Frozen plant leaf material (about 50 mg) was ground in a 1.5 ml reaction tube in a Silamat S5 mixer (Ivoclar Vivadent) using 1.25 - 1.65 mm diameter glass beads (Roth, Art.-Nr. A555.1). 500 µl of Edwards buffer (200 mM Tris [pH 7.5], 250 mM NaCl, 25 mM EDTA and 0.5 % (v/v) SDS) were added and samples were centrifuged for 1 min at 12'000 x g. 400 µl supernatant were transferred to a new reaction tube and mixed with 400 µl isopropanol.

After centrifugation for 5 min at 12'000 x g, the DNA pellet was washed once with 70 % (v/v) ethanol, air-dried and dissolved in 100 µl water.

3.4. Transcript Level Analysis

1 – 5-day-old seeds or unfertilized ovules were harvested into RNA*later* (Sigma, R0901) and ground frozen in a Silamat S5 for 3 times 7 sec. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, cat. 74903). RNA was treated with RNase-Free DNase (Qiagen, cat. 79254). RNA content was estimated on a NanoDrop1000 (Thermo Fisher Scientific) and equal amounts of RNA were used for cDNA synthesis with the First Strand cDNA Synthesis Kit (Fermentas, #K1611). Fast SYBR Green Master Mix (Applied Biosystems) and gene specific primers were used for RT-PRC performed on 7500 Fast Real-Time PCR device (Applied Biosystems). *PROTEIN PHOSPHATASE 2A* (*PP2A*) were used as reference genes. For sequences of primers see Table 3-3. Real-Time PCR was performed using three replicates and results were analyzed as described (Simon, 2003).

Table 3-3. Primers for transcript analysis

Primer	Sequence
PP2A qPCR primers	TAACGTGGCCAAAATGATGC
	GTTCTCCACAACCGCTTGGT
PHE1 qPCR primers	CGCATGTGCGGTCATCC
	TCCAACACCGAAAACTCCAT

3.5. Microarray Analysis

Wild type (L*er*) and double mutant *emf2-5/+ vrn2-1/-* plants were emasculated by manual removal of the anthers around 1-2 days prior anthesis. Two days after emasculation half of the wild type plants were hand pollinated. Unfertilized ovules and young seeds were dissected from the siliques 4 DAE and 2 DAP, respectively, collected to RNA*later* (Sigma, R0901) and ground frozen in a Silamat S5 for 3 times 7 sec. RNA was extracted using QIAGEN RNeasy KIT followed by on-column DNase treatment. Experiment was repeated 5 times and 3 out of 5 biological replicates were used for labeling and microarray hybridization.

Labeling and hybridization to AGRONOMICS1 arrays (Affymetrix, Santa Clara, CA) has been described previously (Rehrauer et al., 2010) and performed by Functional Genomic Center Zurich staff.

Signal values were derived from Affymetrix*.cel files using RMA (Rehrauer et al., 2010). All data processing was performed by Dr. Lars Hennig using the statistic package R (version 2.6.2) that is freely available at http://www.r-project.org/ (Ihaka and Gentleman, 1996). Quality control was done using the affyQCReport package in R. In addition, coefficients of variation (cv) were calculated between replicates as a quantitative measure of data quality and consistency between replicates as described previously (Köhler et al., 2003b). Differentially expressed genes were identified using the limma package in R (Smyth, 2004). Multiple-testing correction was done using the q-value method (Storey and Tibshirani, 2003). Probesets were called significantly differentially expressed when $q < 0.05$. To enrich for biologically relevant changes, only probesets with a minimal SLR of 0.6 were selected. Analysis of GO categories was performed using AtCOECiS (http://bioinformatics.psb.ugent.be/ATCOECIS/).

3.6. Histological Analysis

Samples were fixed in 4% formaldehyde, 50% ethanol, and 10% acetic acid overnight at 4°C. The fixed samples were dehydrated in a series of ethanol dilutions (50%, 70%, and 100%) for 1 hour each, followed by an overnight incubation in 100 % ethanol. After additional wash with 100 % ethanol, samples were embedded in Technovit 7100 (Kulzer) according to the manufacturer's instructions. Four-micrometer sections were prepared using a Leica RM2145 microtome.

For Schiff's staining, sections were oxidized for 30 min with 1 % periodic acid, washed and stained for another 30 min with Schiff's reagent (Sigma-Aldrich, cat. 3952016). The washed and dried sections were embedded in Entellan New mounting media (Electron Microscopy Sciences, cat. 14800) and allowed to cure for 30 min before microscopic analysis was performed.

For Toluidine blue staining, sections were incubated in 0.05% solution of Toluidine blue (Merck), washed with water, dried and mounted in Entellan New.

Seeds expressing β-GLUCURONIDASE (GUS) constructs were fixed for 1 h in acetone, washed twice in 50 mM sodium phosphate (pH 7.0) and immersed in GUS

solution (50 mM sodium phosphate [pH 7.0], 10 mM EDTA, 1 mM potassium ferrocyanide, 1 mM potassium ferricyanide, 0.1 % (v/v) Triton X-100 and 1 mg/ml X-Gluc (in Dimethylformamide)) at 37 °C for 24 h. The seeds were dehydrated and embedded as described above. Sections were directly mounted in Entellan New mounting media and used for microscopy.

3.7. Microscopy

For Nomarski microscopy, samples were fixed in ethanol/acetic acid in a ratio of 9:1, washed with 70% ethanol and mounted in chloral hydrate solution (glycerol/chloral hydrate/water in a ratio of 1:8:3). For vanillin staining, ovules were manually dissected from ovaries and mounted on slides in 1% (w/v) vanillin (4-hydroxy-3methoxybenzaldehyde; Sigma) in 6 N HCl solution. Slides were analyzed after 20 min of incubation. Samples were analyzed with a Leica DM2500 microscope using differential interference contrast optics. Images were recorded with a Leica DFC 300 FX digital camera. Samples from GFP lines were mounted in water and analyzed with epifluorescence optics.

3.8. Feulgen Staining and Confocal Microscopy

Fixation, staining and embedding was performed as described in (Braselton et al., 1996). Confocal imaging was performed using a Leica SP1-2, excitation wavelengths were set to 488 nm and detection to 535 nm and longer.

4. RESULTS

4.1. EMF2 and VRN2 are not Acting Redundantly with FIS2 in the Female Gametophyte

To determine the penetrance of autonomous seed development in different mutants lacking components of the FIS PcG complex I counted number of autonomous seeds in emasculated siliques at 7 days after emasculation (DAE) (Figure 4-1 A). The presence of four or more nuclei in the central cell was considered as criterion for autonomous endosperm and seed development. Alternatively, strongly enlarged autonomous seeds containing less than 4 endosperm nuclei were as well counted. Whereas *fis2* and *mea* mutants generated only very few autonomous seeds (about 3-8 %; penetrance 6-16%) in both L*er* and Col *Arabidopsis* accessions, almost every ovule inheriting a *fie* or *msi1* mutation initiated autonomous seed development (up to 42% or 45% autonomous seeds, penetrance 84% or 90%, respectively) with a stronger effect observed in Col background. Therefore, *fis* mutants have a strikingly different penetrance in autonomous seed formation, which is contrasted by a completely penetrant phenotype of each mutant after fertilization. Lack of any FIS complex subunit gives rise to 50% seed abortion (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998; Köhler et al., 2003a; Guitton et al., 2004), implicating no functional redundancy of *FIS* genes after fertilization. This either suggests the presence of genes acting redundantly with *MEA* and *FIS2* before fertilization or an increased capacity of *fie* and *msi1* to undergo autonomous seed development due to unknown defects caused by loss of FIE and MSI1 function. To test whether differences in the penetrance of autonomous seed development are caused by functional redundancy of FIS2 with VRN2 or EMF2 in the female gametophyte, I analyzed penetrance of autonomous seed development in double heterozygous mutants *fis2*/+ *vrn2*/+ and *fis2*/+ *emf2*/+ (Figure 4-1 B). In this genetic combination it is expected that 25% of female gametophytes lack FIS2 and VRN2 or FIS2 and EMF2 function, respectively. However, penetrance of autonomous seed development was not increased neither in *fis2*/+ *vrn2*/+ nor *fis2*/+ *emf2*/+ double mutants. I generated a triple *fis2*/+ *vrn2*/+ *emf2*/+ mutant containing 12.5% ovules lacking all three homologous genes. In this triple mutant 15% of the ovules initiated autonomous seed development (Figure 4-1 B), whereas autonomous seed development was never observed in *vrn2*/+ *emf2*/+ and *vrn2*/- *emf2*/+ double mutants (not shown). One explanation for this finding is that all three

43

genes act redundantly in the female gametophyte. I investigated this hypothesis further by testing whether expression of *EMF2* and *VRN2* under control of the *FIS2* promoter could rescue the *fis2* mutant. *FIS2* promoter is active in the central cell of the female gametophyte and during endosperm development before and after fertilization, respectively. I generated transgenic lines in the *fis2-1* and *fis2-5* mutant backgrounds expressing the *VRN2* and *EMF2* genes under control of the *FIS2* promoter.
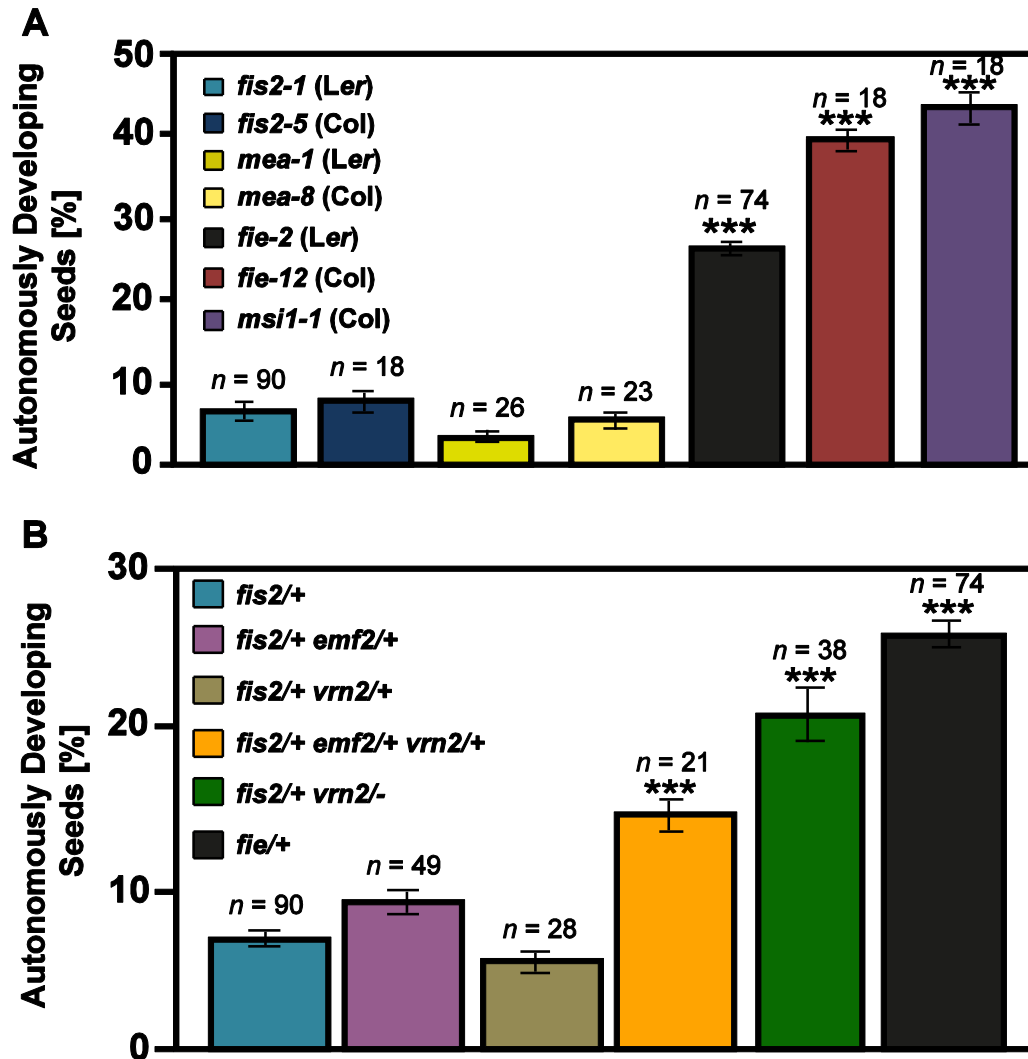

Figure 4-1. Penetrance of autonomous seed development in PcG mutants and mutant combinations. **(A)** Percentage of autonomously developing seeds in different *fis* mutants determined at 7 DAE. **(B)** Percentage of autonomously developing seed in the *fis2-1/+* single mutant and in double and triple mutant combinations of *fis2-1/+* with *emf2-5* and *vrn2-1*. The *fie-2* mutant served as a positive control. Asterisks indicate significant deviations from *fis2/+* determined by ANOVA testing (p<0.001). Numbers on top of each bar indicate analyzed siliques. Each silique contained on average 60 ovules or seeds.

However, among 19 *FIS2pro:VRN2* and 12 *FIS2pro:EMF2* lines, none of them complemented the *fis2* seed abortion phenotype, whereas expression of *FIS2* under control of the same promoter conferred complete complementation (out of 15 lines analyzed, all lines showed complementation revealed by reduction of seed abortion rates to 25% in lines hemizygous for the construct). These findings do not support the hypothesis that VRN2 and EMF2 act redundantly with FIS2 in the female gametophyte and rather suggest that double heterozygosity for *EMF2* and *VRN2* causes an effect on sporophytic tissues promoting the capacity of *fis2* mutants to form autonomous seeds.

Together, I conclude that FIS2 does not act redundantly with EMF2 and VRN2 in the female gametophyte, but that reduced dosage of EMF2 and VRN2 in sporophytic tissues promotes autonomous seed development of the *fis2* mutant. This suggests that PcG proteins act in maternal integument tissues surrounding the female gametophyte to restrict the development of autonomous seeds.

4.2. VRN2 and SWN Act in Sporophytic Tissues to Suppress Autonomous Seed Formation

I tested the hypothesis that decreased dosage of PcG proteins in sporophytic tissues enhances autonomous seed formation of *fis2* by analyzing autonomous seed development in a *fis2/+ vrn2/-* double mutant. In contrast to a double heterozygous *fis2/+ vrn2/+* mutant that did not increase penetrance of autonomous seed formation (Figure 4-1 B), a strongly increased penetrance was observed in the *fis2/+ vrn2/-* double mutant (Figure 4-1 B), supporting the idea that reduction of PcG function in sporophytic tissues promotes autonomous seed development. The expression of *VRN2* and *SWN* in sporophytic tissues of the integuments was confirmed by studying expression of the *VRN2pro:VRN2-GUS* and *SWNpro:SWN-YFP* reporter lines (Figure 4-2).
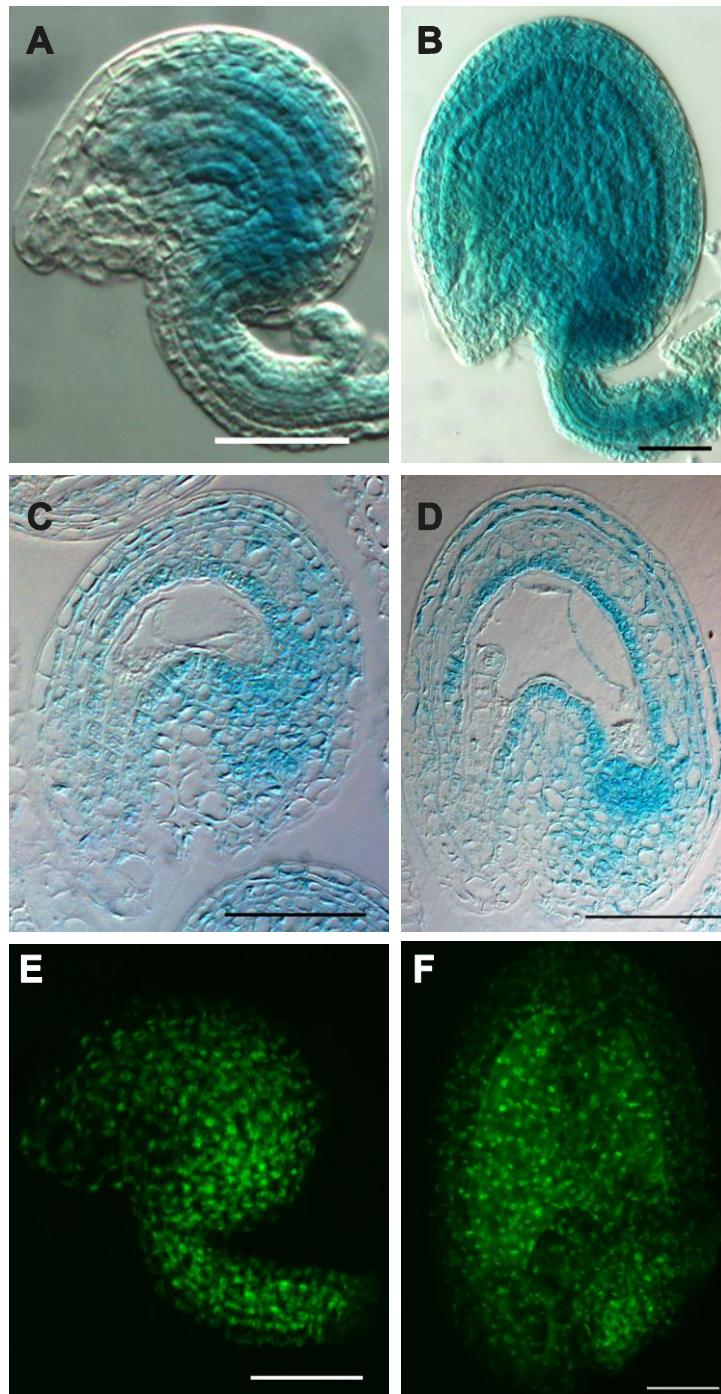
Figure 4-2. Expression of *VRN2* and *SWN* in the ovule integuments and the seed coat of young seeds. Microscopy images of cleared **(A and B)** and sectioned **(C and D)** *VRN2pro:VRN2-GUS* stained ovules **(A, C)** and young seeds **(B, D)**. *SWNpro:SWN-YFP* expression in the integuments **(E)** and the seed coat **(F)**. Scale bar = 50 μm.

Previous studies proposed that low penetrance of autonomous seed development in the *mea* mutant is caused by functional redundancy of MEA and the MEA homolog SWN

in the central cell (Wang et al., 2006). *SWN* is only weakly expressed in the central cell of the female gametophyte but strongly expressed in sporophytic tissues (Wang et al., 2006; Spillane et al., 2007) and (Figure 4-2 E), where it was previously shown to act redundantly with CLF (Chanvivattana et al., 2004).
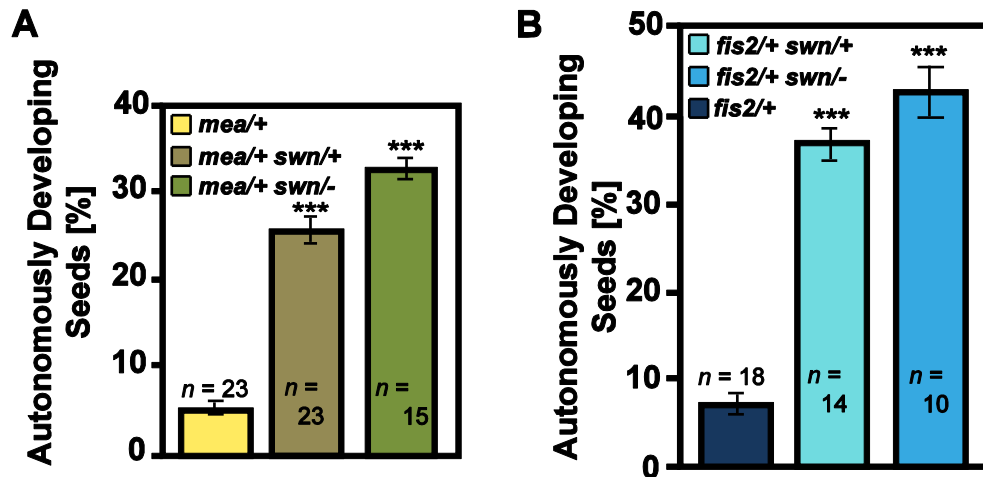

Figure 4-3. Penetrance of autonomous seed development in *mea* and *fis2* combined with *swn*. **(A)** Percentage of autonomously developing seeds in the *mea-8/+* single mutant and double mutant of *mea-8/+* with *swn-3*. **(B)** Percentage of autonomously developing seeds in the *fis2-5/+* single mutant and double mutant of *fis2-5/+* with *swn-3*. Asterisks indicate significant deviations from *mea/+* (A) or *fis2/+* (B) determined by ANOVA testing (p<0.001). Numbers on top of each bar indicate analyzed siliques. Each silique contained on average 60 ovules or seeds.

In agreement with previous findings (Wang et al., 2006), the *mea/+ swn/+* double heterozygous mutant had strongly increased penetrance of autonomous seed development in comparison to the *mea* single mutant (Figure 4-3 A). If MEA and SWN act redundantly in the female gametophyte to suppress autonomous seed development, it is expected that the *mea/+ swn/-* double mutant would have approximately twice as many autonomously developing seeds compared to the *mea/+ swn/+* double heterozygous mutant. However, only a minor increase of autonomously developing seeds was observed upon complete loss of SWN function, implicating that SWN, like VRN2 and EMF2, has a role in sporophytic tissues to suppress autonomous seed development.

To further explore this idea I generated and analyzed *fis2/+ swn/+* and *fis2/+ swn/-* mutants that are affected in non-homologous PcG genes *FIS2* and *SWN*. These double mutant combinations caused a dramatic increase in autonomous seed formation to 37 %

and 43 % compared to 7 % in the *fis2* single mutant (Figure 4-3 B). Consequently, *fis2/+ swn/-* did not produce twice more autonomous seeds as *fis2/+ swn/+* but only minor increase in penetrance was observed. These additional results strongly support a sporophytic role of SWN to suppress autonomous seed development.

4.3. FIS Genes Have a Non-Redundant Role after Fertilization

PHERES1 (*PHE1*) is a direct target gene of the FIS PcG complex and exclusively expressed in the endosperm (Köhler et al., 2003b; Weinhofer et al., 2010). Loss of FIS2 function causes increased expression of *PHE1* in the endosperm (Figure 4-4). Consistent with the idea that FIS2 does not act redundantly with VRN2 and EMF2, expression levels of *PHE1* were similarly increased in seeds of single *fis2* mutant as in double and triple mutants of *fis2* with *vrn2* and *emf2*. Similarly, there was no relevant difference in the expression of *PHE1* as well as of a *PHE1pro:PHE1-EGFP* reporter line in the *fie* mutant compared to the *fis2* mutant (Figure 4-4), supporting the idea that *FIS2* homologs do not act redundantly in controlling expression of FIS target genes. Together with the fact that *fis* mutants have a completely penetrant seed abortion phenotype, I conclude that *FIS* genes play a non-redundant role in the endosperm after fertilization and propose that differences among *fis* mutants prior to fertilization are caused by a haploinsufficient requirement of FIE and MSI1 in suppressing seed coat development.
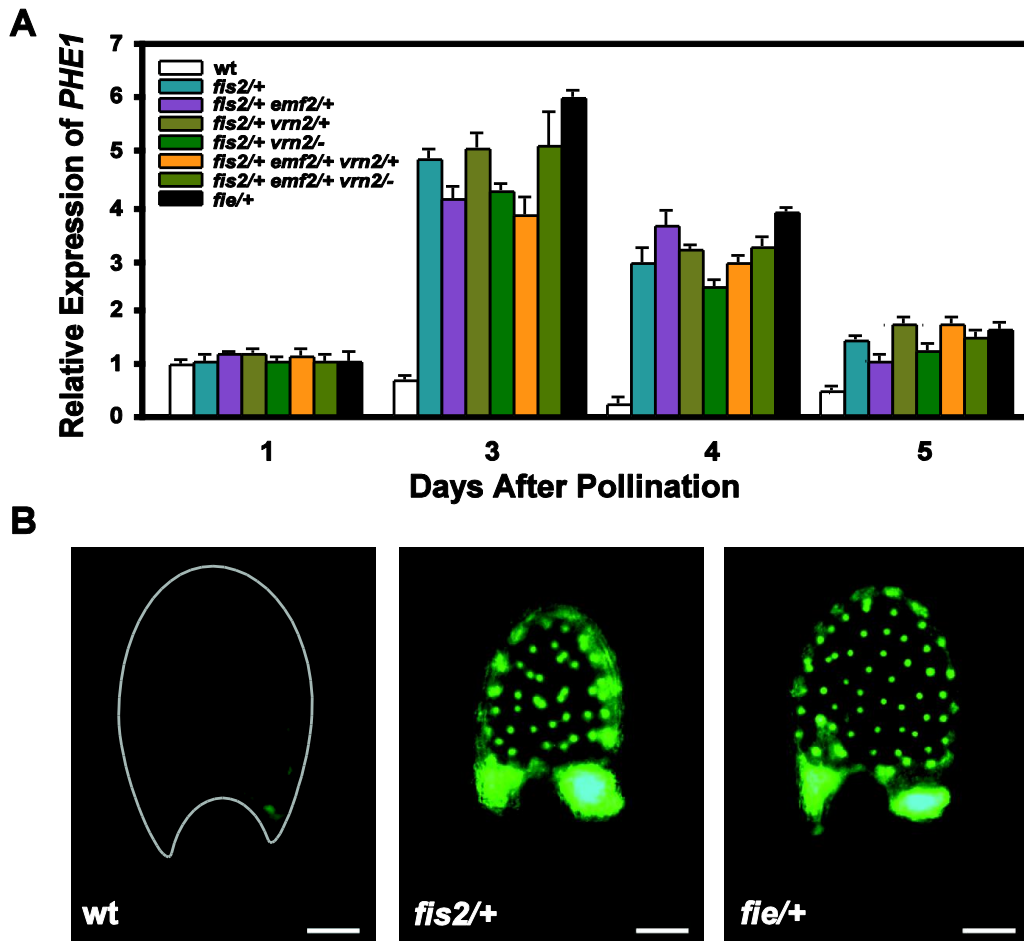
Figure 4-4. Expression analysis of the FIS target gene *PHE1* in seeds of *fis* mutants and mutant combinations with *vrn2* and *emf2*. **(A)** Quantitative RT-PCR analysis of *PHE1* expression in wild type, *fis2-1* and *fie-2* single mutants, and *fis2-1* in combination *vrn2-1* and *emf2-5* mutants. Error bars indicate SEM. **(B)** Expression of the *PHE1pro:PHE1-EGFP* reporter gene in wild-type (wt), *fis2-1* and *fie-2* seeds at 3 DAP. Gray line depicts border between seed coat and endosperm in wild-type. Bar = 50 μm.

4.4. *fis* Mutants Form Two Classes of Autonomously Developing Seeds

All four *fis* mutants generate autonomous seeds. However, I noticed that aside from the different penetrance of this phenotype, there were two distinct classes of autonomously developing seeds. While most of the autonomously developing seeds in *fis2* and *mea* remained small like unfertilized wild-type ovules (Figure 4-5 D and J), the majority of autonomously formed seeds in *fie* and *msi1* were much bigger (Figure 4-5 A, C, H and J). The size of the autonomous seeds was not directly correlated with endosperm growth, since I found small seeds with up to 64 nuclei and big seeds with as few as one, two or four endosperm nuclei (Figure 4-5 D, F and G). Furthermore, I found a fraction of big seeds without any visible endosperm nuclei (Figure 4-5 E).
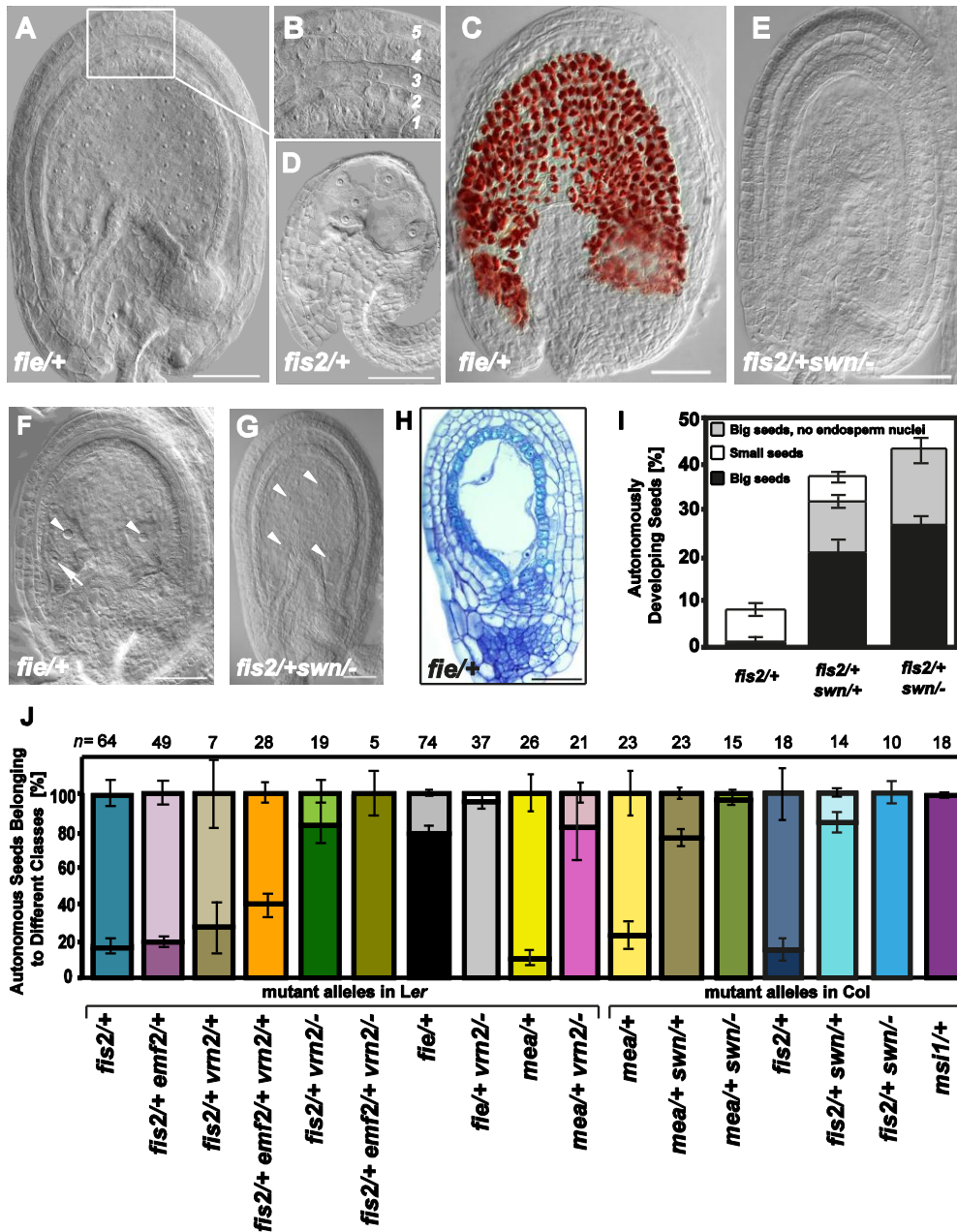
Figure 4-5. *fis* mutants form two classes of autonomously developing seeds. **(A)** Big autonomous *fie-2* seed with developed seed coat and endosperm. Inset window enlarged in **(B)** shows five seed coat layers. **(C)** Proanthocyanidin accumulation in *fie-2* seed. **(D)** A small autonomous *fis2-1* seed with developing endosperm but degenerated integuments. **(E)** Big autonomous seed of *fis2-5 swn-3* with developing seed coat but without visible endosperm nuclei. **(F)** Autonomous *fie-2* seed with developed seed coat and only two endosperm nuclei (arrow heads). Arrow indicates egg cell nucleus. **(G)** Autonomous *fis2-5 swn-3* seeds with developed seed coat and only four endosperm nuclei (arrow heads) **(H)** Section of toluidine blue stained autonomous *fie-2* seed. (A-H) Bar = 50 μm. **(I)** Percentage of autonomously developed seeds including seeds without endosperm nuclei visible in *fis2-5* and *fis2-5 swn-3* double mutant. **(J)** Percentage of each class of autonomous seeds (small – hatched and big non hatched) in respective genotypes (*fis2-1, fie-2, mea-1* in combination with *emf2-5* and *vrn2-1* for mutants in L*er* background and *mea-8, fis2-5* and *msi1-1* in combination with *swn-3* in Col background) determined at 7DAE. Numbers on top of each bar indicate analyzed siliques.

Lack of the endosperm nuclei in some of these seeds might result from a damage of the sample occurred during its preparation process. However, these seeds were rarely observed in the *fie* mutant, whereas much more abundant in double mutant of *fis2 swn* (Figure 4-5 E and I).

The second striking difference concerns the development of the integuments. Similar to unfertilized wild-type ovules, integuments of *fis2* and *mea* autonomous seeds did not progress in their development and degenerated about 5 DAE although they contained developing endosperm (Figure 4-5 D). In stark contrast, integuments of big autonomous seeds in *fie* and *msi1* mutants differentiated into a seed coat with five clearly distinguishable cell layers (Figure 4-5 B and H) that accumulated proanthocyanidins in the endothelium layer (Figure 4-5 C) as a hallmark for seed coat development (Haughn and Chaudhury, 2005).

These findings implicate that compromised sporophytic PcG function in *fie* and *msi1* mutants promotes autonomous seed coat development, suggesting that sporophytically acting PcG complexes prevent development of the seed coat in the absence of fertilization. In agreement with this hypothesis the majority of autonomously developing seeds in *fis2/+ vrn2/-* , *mea/+ vrn2/-* and *fis2/+ swn/-* double mutants were big and had developed seed coats (Figure 4-5 J). Partial loss of VRN2 in autonomous seeds of *fis2/+ vrn2/+* double mutants did not affect *fis2* phenotype. However, lowered dosage of SWN in *mea/+ swn/+* and *fis2/+ swn/+* dramatically influenced ratio of two classes of autonomous seeds (Figure 4-5 J). Triple heterozygous plants *fis2/+ vrn2/+ emf2/+* differed from *fis2* mostly in the number of autonomous seeds with developed seed coat, whereas the number of small and degenerated autonomous seeds was comparable between both genotypes (around 7%, Figures 4-1 B and 4-5 J). Together, I conclude that differences in developmental potential of autonomous seeds are controlled by sporophytically acting PcG complexes.

4.5. FIE is Haploinsufficient and Acts in Sporophytic Tissues to Suppress Seed Coat Development

The strong penetrance of autonomous seed formation of the heterozygous *fie* mutant suggests that FIE is a dosage-sensitive suppressor of integument development. I tested this hypothesis by analyzing whether I could increase the penetrance of

autonomous seed formation of the *fie/+* mutant by additionally depleting VRN2 function. Therefore, I generated and analyzed double mutants of *fie* with *vrn2.* Whereas about 20% of autonomously developing seeds in the *fie/+* mutant had early degenerating integuments, almost all developing autonomous seeds in *fie/+ vrn2/-* had developed seed coats (Figure 4-5 J). Thus, I conclude that incomplete penetrance of the *fie* mutant phenotype is caused by remaining PcG function in sporophytic tissues. To further test the hypothesis that FIE acts in sporophytic tissues to suppress autonomous seed development, I expressed the *FIE* gene under control of the 35S promoter of the *Cauliflower mosaic virus* (35S). The *35S* promoter is active in sporophytic tissues but is generally considered not to be active during the gametophytic stage (Acosta-García and Vielle-Calzada, 2004).
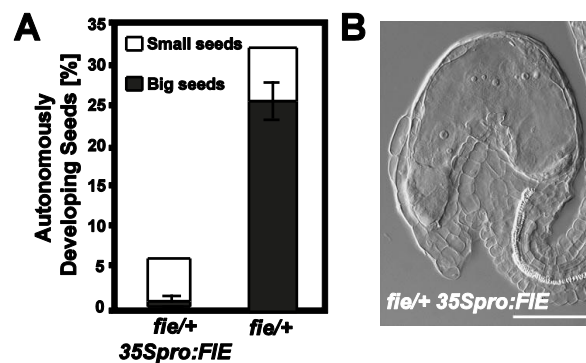

Figure 4-6. Integument-specific complementation of the *fie-2* mutation abolishes seed coat development in autonomous seeds. **(A)** Percentage of autonomously developed seeds in *fie-2* and *fie-2* mutant expressing a wild-type copy of the *FIE* gene under control of the *35S promoter*. **(B)** Small autonomous seed from *35Spro:FIE*; *fie/+* line.

Consistently, expression of *35Spro:FIE* in the *fie/+* mutant did not complement the *fie* seed abortion phenotype. Among 13 *35Spro:FIE*; *fie/+* lines analyzed, I did not identify a line with reduced numbers of aborted sexual seeds. However, I observed a complete suppression of the formation of big autonomous seeds with developed seed coats (Figure 4-6), strongly supporting the hypothesis that FIE is haploinsufficient and acts in integument tissues suppressing seed coat development.

4.6. Loss of VRN2 Initiates Endothelium Development in the Absence of Endosperm Development

Development of wild-type ovules is accompanied by cell divisions in all layers of the integuments. This process decelerates in mature ovules at the time of anthesis. Ovules at this stage await fertilization and if fertilization does not occur, they quickly degenerate at about 5 DAE (Skinner et al., 2004) and (Figure 4-7 A and G), corresponding to about 3-4 days post anthesis. In contrast, if fertilization occurs, development of seeds is accompanied by the fast growth of endosperm and seed coat (Figure 4-7 I and J). The initial phases of seed coat growth are mainly characterized by cell expansion and elongation rather than cell divisions (Garcia et al., 2005) and (Figure 4-7 J). I observed that *vrn2/-* ovules did not degenerate at the same time as wild-type ovules. Clearings and sections of unfertilized ovules revealed that even at 6 DAE *vrn2* ovules were not degenerated and five layers of integuments were clearly distinguishable (Figure 4-7 B, F, H and C).

I investigated whether unfertilized *vrn2/-* ovules initiate seed coat development by analyzing the formation of proanthocyanidins using vanillin staining. Whereas vanillin staining in unfertilized wild-type ovules was rarely observed, more than half of all tested *vrn2/-* ovules showed a strong staining in the micropylar region of the ovule (Figure 4-7 D, K and L), which was, however, not accompanied by a size increase of *vrn2/-* ovules (Figure 4-7 E, F and M). Together, loss of VRN2 function initiates seed coat differentiation in unfertilized ovules, supporting the view that PcG proteins suppress seed coat development in the absence of fertilization. I tested whether seed coat development would also be initiated in heterozygous *vrn2/+, emf2/+* and double heterozygous *vrn2/+ emf2/+* mutants. Consistent with the finding that neither heterozygous *vrn2/+* nor *emf2/+* could enhance the *fis2* phenotype, I did not observe initiation of autonomous seed coat development in both mutants (Figure 4-7 C). Similar to homozygous *vrn2* ovules, about 22% of the ovules of double heterozygous *vrn2/+ emf2/+* mutants remained intact and did not collapse 6 DAE (Figure 4-7 C), supporting the view that the observed increased penetrance of autonomous seed formation in *fis2/+ vrn2/+ emf2/+* triple mutants (Figure 4-1 B) is caused by depletion of PcG function in the integuments, promoting seed coat formation.
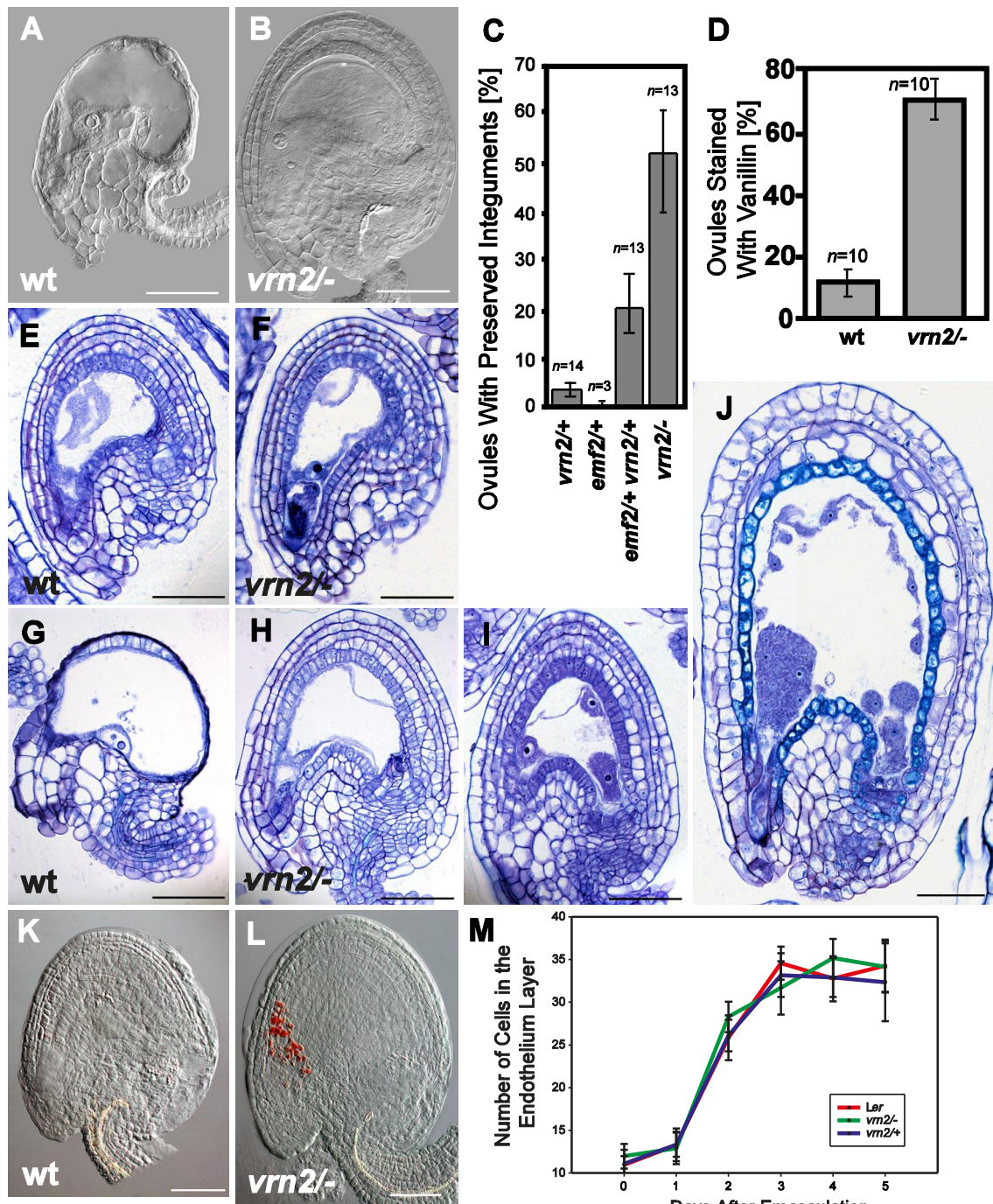
Figure 4-7. Loss of VRN2 initiates endothelium development in the absence of endosperm development. Microscopy images of cleared **(A and B)**, sectioned **(E-J)**, and vanillin stained **(K, L)** ovules and autonomous seeds at 5 DAE **(E, F, K, L)**, 6 DAE **(A, B, G, H)** and seeds at 0.5 DAP **(I)** and 2 DAP **(J)**. **(A, E, G, K, I and J)** wild type (wt), **(B, F, H, L)** *vrn2-1/-*. Scale bar = 50 µm. **(C)** Percentage of ovules with non-degenerated five integument layers at 6 DAE, error bars indicate SEM. **(D)** Percentage of wild-type and *vrn2-1/-* ovules stained with vanillin at 5 DAE, error bars indicate SEM. **(M)** Number of cells in the visual section (DIC microscopy) of endothelium layer of wild-type, *vrn2-1/+* and *vrn2/-* ovules from 0-5 DAE. Cells were counted in triplicates.

4.7. A Signal for Seed Coat Development is Generated by the Sexual Endosperm

Although data from several studies support a role of the endosperm in regulating seed coat growth (Luo et al., 1999; Weijers et al., 2003; Garcia et al., 2005; Ingouff et al., 2006), formal evidence for the sexual endosperm being sufficient for generating a signal initiating seed coat development is missing. To define the origin of the signal initiating seed coat development I investigated seeds of the *kokopelli* (*kpl)* mutant. The *kpl* mutation affects male gametogenesis, leading to the formation of single sperm cell male gametophytes and random single fertilization events (Ron et al., 2010). Consequently, seeds derived after fertilization with *kpl* pollen frequently contain only embryos or only endosperm, resulting from fertilization of egg or central cell, respectively (Figure 4-8 A and B). Seed coat development of both seed classes was strikingly different, whereas seeds containing only endosperm had normally developed seed coats (Figure 4-8 A and D), seeds containing only an embryo did not initiate seed coat development and integuments appeared similar as in unfertilized wild-type ovules (Figure 4-8 B and C). In this seed class embryo development arrested at the globular stage and integuments degenerated at the same time as in unfertilized wild-type ovules (Figure 4-8 E and F). Given that the autonomously formed endosperm in *fis2* and *mea* mutants is rarely sufficient to initiate seed coat formation, I conclude that a signal initiating seed coat development in *Arabidopsis* is derived from the sexual endosperm.
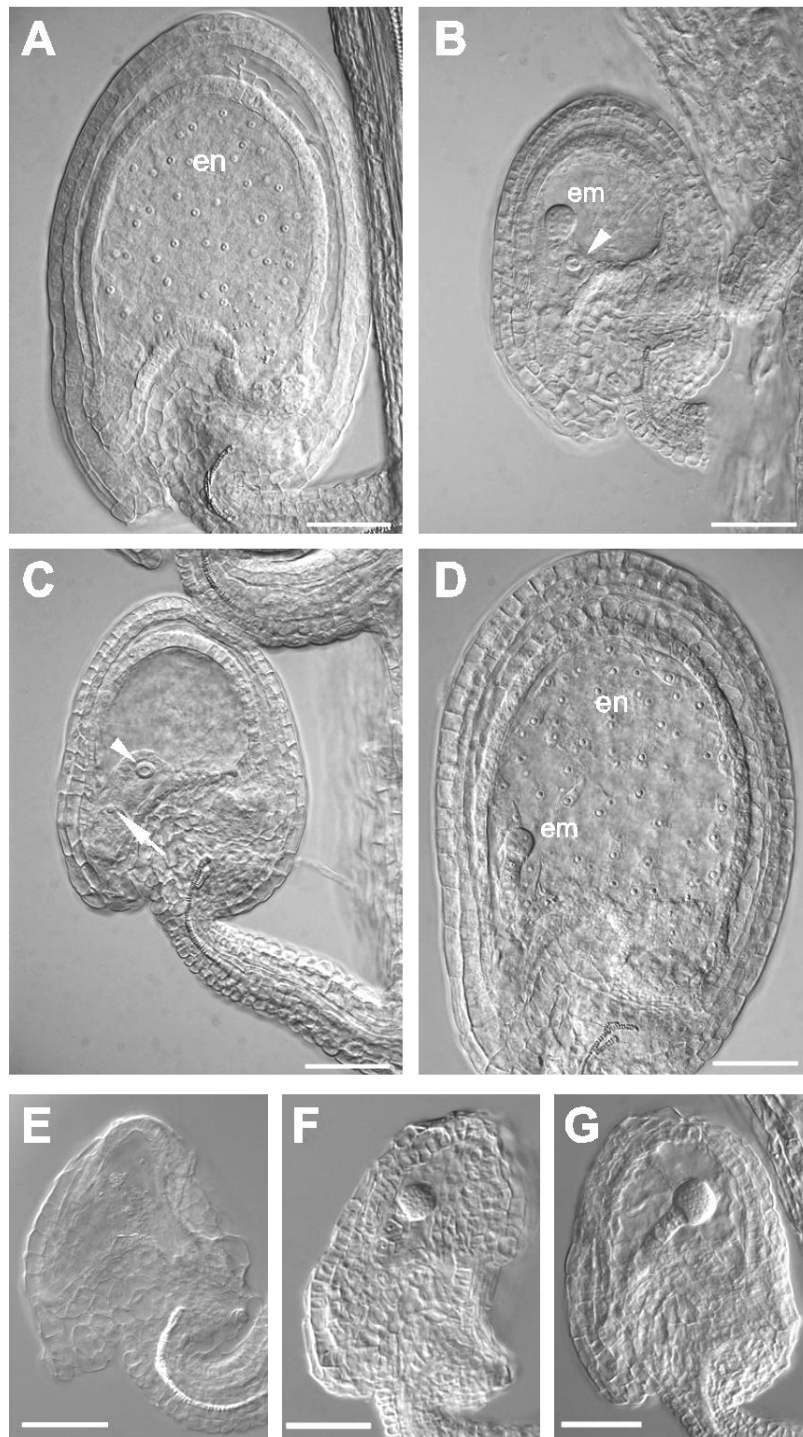
Figure 4-8. The sexual endosperm generates a signal for seed coat development. **(A)** Seed developing after fertilization with *kpl-2* pollen at 3 DAP. Seed contains only endosperm and no embryo. **(B)** Seed developing after fertilization with *kpl-2* pollen at 3 DAP. Seed contains only an embryo and no endosperm. **(C)** Unfertilized ovule shortly before degeneration. **(D)** Seed developing after fertilization with wild-type pollen at 3 DAP. Degenerating seed coat of wild type ovule at 6 DAE **(E)** and seed fertilized with *kpl-2* pollen at 6 DAP **(F and G)**. Seed contains only an embryo and no endosperm. An arrow indicates unfertilized egg cell nucleus. An arrowhead indicates unfertilized central cell nucleus. Embryo (em), endosperm (en), scale bars = 50 μm.

4.8. A Signal for Seed Coat Development Depends on AGL62

Disruption of the signaling pathway from the endosperm to the seed coat is expected to result in the formation of early arresting seeds containing embryo and endosperm, but without a developed seed coat. Lack of the type I MADS-box transcription factor AGL62 causes precocious endosperm cellularization after about 3-4 nuclei divisions and early embryo arrest (Kang et al., 2008). Although *agl62* seeds initiate embryo and endosperm development, I found that this was generally not accompanied by seed coat development (Figure 4-9 A and B).
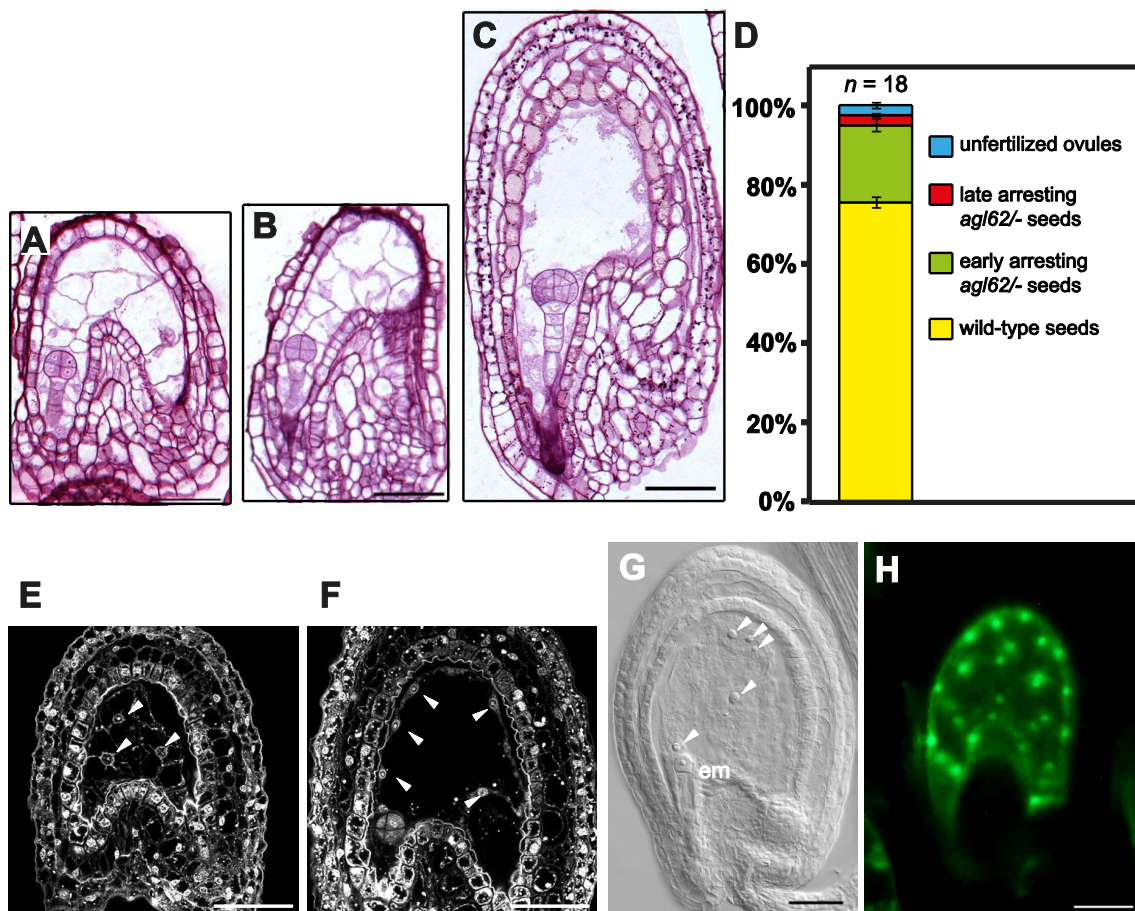

Figure 4-9. A signal for seed coat development depends on AGL62. **(A, B)** Sections of early arresting *agl62-2/-* seeds at 3 DAP. **(C)** Section of late arresting *agl62/-* seed at about 5 DAP. **(D)** Percentage of different classes of seeds in the siliques of self-fertilized *agl62/+* plants. **(E, F)** Confocal pictures of Feulgen stained *agl62/-* seeds, early and late arresting, respectively. Arrow heads indicate endosperm nuclei. Endosperm cell walls are visible in (E) but not (F). **(G)** Cleared *ttn2* seed at 2 DAP. **(H)** *AGL62pro:AGL62-GFP* expression in autonomous *fie* seeds at 4DAE. Scale bars = 50 µm.

Testing cellularization process in *agl62/-* seeds, 3 DAP with a use of sections and Feulgen staining technique, I noted a minor fraction of *agl62/-* seeds arresting development later with about 50 endosperm cells and a developed seed coat (Figure 4-9 C, D and E, F), suggesting incomplete penetrance of the *agl62-2* mutant phenotype. The majority of *agl62/-* seeds did not initiate seed coat formation and had collapsed integuments after about 3DAP (Figure 4-9 A and B), similar to the phenotype observed in autonomously developing *fis2* seeds (Figure 4-5 D) and *kpl* seeds containing only embryo (Figure 4-8 B). It is unlikely that failure of seed coat initiation is a consequence of endosperm proliferation failure, as seed coat development is clearly initiated in the *titan 2* (*ttn2*) mutant that has severe endosperm proliferation defects and arrests development containing a comparable number of endosperm nuclei as the *agl62* mutant (Figure 4-9 G and (Liu and Meinke, 1998)). This strongly implicates that AGL62 is required in the endosperm to initiate seed coat formation.

I analyzed *AGL62* expression in wild-type seeds after fertilization and autonomously developing *fie* seeds using a translational fusion of the *AGL62* gene under control of the native promoter fused to GFP. This construct completely complemented the *agl62* mutant phenotype, indicating that the *AGL62pro:AGL62-GFP* expression reflects the native *AGL62* expression pattern. In wild-type as in autonomously developing *fie* seeds *AGL62* remained exclusively expressed in endosperm nuclei (Figure 4-9 H), indicating that AGL62 is rather required to form the mobile signal than being the mobile signal itself.

4.9. PHE1 is not Necessary for AGL62 Activity in the Endosperm

According to (Kang et al., 2008) and my own *AGL62pro:AGL62-GFP* expression studies AGL62 gets activated in the endosperm only after fertilization. In agreement with Kang and colleagues I did not detect any GFP signal in unfertilized central cell of the wild-type, neither in additionally tested *fis2* mutant ovules. Together with the fact that *AGL62* is not imprinted but expressed bi-allelically (Kang et al., 2008) it suggests an imprinted nature of its activator. *PHE1* is endosperm specific, imprinted gene, expressed predominately from the paternal allele only after fertilization. PHE1 has been shown to physically interact with AGL62 in yeast two-hybrid studies (deFolter et al., 2005). To test whether PHE1 is an activator necessary for AGL62 functional activity I examined *phe1/-* seeds and seeds derived from *agl62/+* pollinated with *phe1/-* mutant pollen.
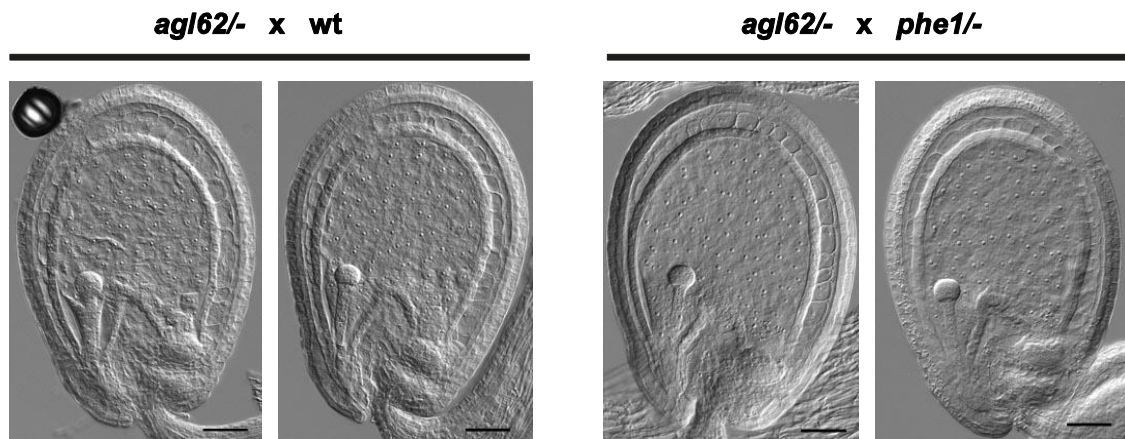
Figure 4-10. PHE1 is not essential for AGL62 activity. Cleared seeds coming from the cross of *agl62-2/-* with wild-type (wt) plant **(A)** and *phe1-1/-* **(B)** inspected at 3 DAP.

Reducing the dosage of AGL62 by using *agl62/+* plant as a mother could possibly sensitize the system and allow detection of a synergistic effect in the double mutant combination. However, I did not identify any seed with *agl62/-* phenotype neither in *phe1/-* mutant nor in the several siliques of *agl62/+* pollinated with *phe1/-* (Figure 4-10), suggesting that PHE1 is not necessary for AGL62 activity. However, functional redundancy of PHE1 with PHE2 might obscure the interpretation of this phenotype.

4.10. AGL62 is a Negative Regulator of Fertilization Independent Seed Development

The results of this study implicate that AGL62 controls not only timing of endosperm cellularization but also affects formation of the signal for seed coat development. Therefore, I investigated how the *agl62* mutation affects these processes in autonomously developing seeds of the *fie* mutant. I anticipated this double mutant to show early symptoms of cellularization or a completed cellularization process in autonomously developing endosperm, accompanied by an undeveloped and degenerated seed coat. Surprisingly, I found strongly reduced penetrance of the FIS phenotype in *fie/+ agl62/+* mutant (Figure 4-11). The number of autonomous seeds was reduced by half, indicating that every *fie* ovule inheriting an *agl62* mutation was arrested at the very beginning of autonomous seed development even before the first division of the central cell.
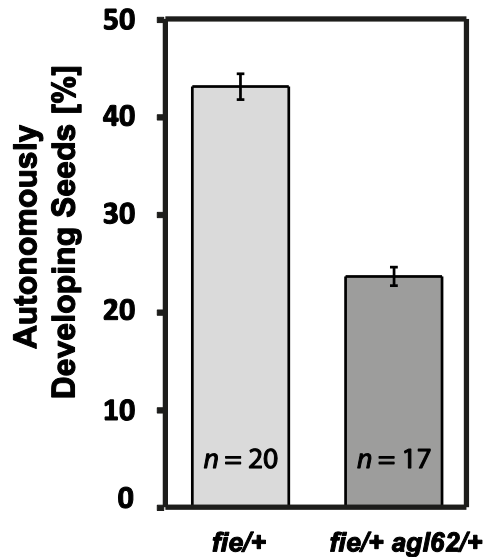
Figure 4-11. Loss of AGL62 suppresses *fie* autonomous endosperm development. Percentage of autonomously developing seeds in *fie-12/+* and *fie-12/+ agl62-2/+* double mutant.

These data strongly suggest that AGL62 is required for autonomous endosperm development. Importantly, it reveals that *AGL62* gene is expressed and active in the central cell of the female gametophyte before fertilization, although it could not be detected with use of the *AGL62pro:AGL62-GFP* expressing line in our group neither by Kang and colleagues (Kang et al., 2008).

4.11. Transcriptional Profiling of Early Developing Seeds and Mature Ovules

To investigate the molecular bases of seed coat development, I performed transcriptional profiling of wild-type and *emf2-5/+ vrn2/-* mutant ovules and early developing wild-type seeds. Wild-type seeds were collected at 2 DAP, corresponding to the first moment when elongation of seed coat cells was clearly visible. Mutant ovules showing symptoms of seed coat development were collected together with wild-type ovules at 4 DAE, before wild-type integuments showed any signs of degeneration. I compared the transcriptional profiles of developing seeds and wild-type ovules as well as the profiles of wild-type and mutant ovules. I anticipated that genes being upregulated after fertilization in wild-type seeds might as well being upregulated in PcG mutant ovules initiating autonomous seed coat development. There are two possible hypotheses explaining the autonomous onset of seed coat development in PcG mutants, (i) loss of PcG

function causes the formation of the seed coat initiation signal, and (ii), loss of PcG function causes the initiation of the processes downstream of the signal, without initiating signal formation. I found 551 genes being similarly upregulated (SLR>0.6, p<0.05) in both comparisons, of which 128 genes were not significantly expressed before fertilization (Appendix, Table 7-1). These genes were significantly enriched for genes regulating cell wall loosening and modification, as well as protein regulating cell growth and cell size, consistent with the observation that loss of PcG function initiates seed coat development (Table 4-1).

Table 4-1. GO categories of genes upregulated in wild-type young seeds and *emf2/+ vrn2/-* ovules and not expressed before fertilization.

GO-term	p-value	Number of genes	Definition
Functions			
GO:0016891	6.33E-04	3	endoribonuclease activity, producing 5'-phosphomonoesters
GO:0016893	6.87E-04	3	endonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters
GO:0019887	7.43E-04	3	protein kinase regulator activity
GO:0019207	8.02E-04	3	kinase regulator activity
GO:0004521	1.66E-03	3	endoribonuclease activity
GO:0004519	4.33E-03	3	endonuclease activity
GO:0004540	6.06E-03	3	ribonuclease activity
Processes			
GO:0009828	5.34E-04	3	cell wall loosening (sensu Magnoliophyta)
GO:0009827	6.87E-04	3	cell wall modification (sensu Magnoliophyta)
GO:0042545	1.35E-03	4	cell wall modification
GO:0009826	1.56E-03	4	unidimensional cell growth
GO:0016049	4.58E-03	4	cell growth
GO:0008361	5.18E-03	4	regulation of cell size
GO:0009831	6.04E-03	2	cell wall modification during multidimensional cell growth (sensu Magnoliophyta)
GO:0042547	7.03E-03	2	cell wall modification during multidimensional cell growth
GO:0051726	8.68E-03	3	regulation of cell cycle
GO:0040007	9.62E-03	4	growth
Compartments			
GO:0005618	9.40E-04	6	cell wall
GO:0030312	1.02E-03	6	external encapsulating structure

I tested how many of those genes are marked by H3K27me3, using publicly available datasets generated from seedlings and adult plants (Zhang et al., 2007; Oh et al., 2008). Of the 128 deregulated genes 111 genes are present on the ATH1 microarray; of those 35 genes are marked by H3K27me3 (P>0.05). Thus, there is no significant enrichment for PcG

target genes among the set of deregulated genes, implicating that many of the deregulated genes are downstream target genes of PcG target genes. Genes being marked by H3K27me3 are enriched for transcription factors as well as genes responsive to hormones (Table 4-2), among them are two auxin responsive genes and two genes responsive to jasmonic acid and ethylene. To explore whether indeed hormones form the signal will be part of future investigation.

Table 4-2. GO categories of genes enriched for H3K27me3, upregulated in wild-type young seeds and *emf2/+ vrn2/-* ovules and not expressed before fertilization.

GO Term	p value	Number of genes	Definition
GO:0003700	1.56E-03	8	Transcription factor activity
GO:0030528	2.92E-03	8	Transcription regulator activity
GO:0003677	8.13E-03	8	DNA binding
GO:0044275	8.38E-03	2	cellular carbohydrate catabolic process
GO:0016052	8.38E-03	2	Carbohydrate catabolism
GO:0009725	9.35E-03	4	Response to hormone stimuli
GO:0009723	1.37E-02	2	Response to ethylene stimuli

5. DISCUSSION

5.1. Seed Coat Initiation is regulated by Sporophytically Acting PcG proteins

The FIS PcG complex prevents the formation of autonomous development by repressing divisions of the central cell in the absence of fertilization. In this study I could show that initiation of seed development is controlled by PcG proteins on another level that was not identified previously. While specific impairment of FIS function in *fis2* and *mea* mutants elicits autonomous replication of the central cell, it fails to efficiently initiate seed coat development. In contrast, concomitant loss of FIS function as well as compromised function of sporophytic PcG complexes in the integuments initiates formation of autonomous seeds containing endosperm as well as seed coat. The percentage of autonomously developing seeds has been determined for different FIS mutant alleles: *fie* (L*er*) 47% (Ohad et al., 1996), *mea-3* (L*er*) 7.5%, (Col) 10.3% (Kiyosue et al., 1999), *msi1-1* (Col) 41.2% (Köhler et al., 2003a), *msi1-3* 40.9% (Ngo et al., 2007), *mea-3* (backcrossed to Col) 3.09-11.82 % (Wang et al., 2006), *mea-1* (L*er*) 12.8%, *mea-8* (Col) 28.7% (Ngo et al., 2007). Data obtained in this study reflect the same trend as observed in previously published experiments - high penetrance of the FIS phenotype in mutants affected in the gametophytically and sporophytically active genes *FIE* and *MSI1*, and much weaker effect in mutants affected in the specifically gametophytically active genes *FIS2* and *MEA.* In my study I obtained generally lower values for the autonomous endosperm penetrance in the *mea* mutation than previously published. One possible explanation for this difference is the use of a different method of counting autonomous seeds. Here, the requirement for autonomous seed to be counted was the presence of at least four endosperm nuclei. In wild-type as well as in *fis* mutant ovules two polar nuclei could be observed in the central cell before fertilization. In order not to account central cell polar nuclei as a pair of endosperm nuclei originating from the first division of the central cell nucleus, I decided to only count autonomous seeds in which the central cell nucleus had undergone at least two mitotic divisions. This might have potentially lowered the values obtained for *fis2* and *mea* that more frequently than *fie* and *msi1* show low number of endosperm nuclei in autonomous seeds. Another factor influencing the penetrance of the FIS phenotype is the accession background. Previous studies as well as my data reveal that alleles in the Col background have a higher potential for fertilization independent seed

formation than those obtained in L*er,* e.g. *fie-2* in L*er* 26% and *fie-12* in Col 38%. Hence, for all experiments only mutants within the same accession background were used.

5.2. VEFS Domain proteins FIS2, EMF2 and VRN2 do not Act Redundantly in the Female Gametophyte neither in the Endosperm

Several *Arabidopsis* PcG genes acting in sporophytic tissues act redundantly. It has been suggested that the low penetrance of *fis2* and *mea* mutants to form autonomous endosperm is caused by redundant action of FIS2 homologues EMF2, VRN2 and the MEA homolog SWN (Kiyosue et al., 1999; Chanvivattana et al., 2004; Wang et al., 2006). Here I show that FIS2 does not act redundantly with EMF2 and VRN2 in the female gametophyte and neither in the developing endosperm. If EMF2 and VRN2 could replace the function of FIS2 in the central cell of the female gametophyte, the phenotype of the double mutants *fis2/+ emf2/+* and *fis2/+ vrn2/+* should be enhanced and increased penetrance of autonomous seed formation would be expected. Furthermore, if homologues could replace FIS2 in the FIS complex during endosperm development, complete de-repression of the FIS target gene *PHE1* would only be possible in the double or triple mutant and *fis2* single mutants should not have a completely penetrant seed abortion phenotype. However, none of these predictions hold true. Moreover, a strong enhancing effect of the *swn* mutation on *mea* and *fis2* autonomous seed penetrance supports the view that loss of SWN causes a sporophytic effect rather than a gametophytic effect by redundant action. Whereas the homologous proteins MEA and SWN could act redundantly, redundant action of SWN and FIS2 is rather unlikely as FIS2 is a different type of protein with different functional domains. The effect of the homozygous, sporophytic affect of the *vrn2/-* mutation on *fis2* and *mea* phenotype adds additional support to the hypothesis that sporophytically acting PcG proteins repress development of autonomous seeds. Of particular importance is the phenotypic change of autonomous seeds observed in *mea vrn2/-*, a combination mutants in two non-homologues genes similar to the *fis2 swn* mutant. My analysis of heterozygous *fie* as well as double heterozygous *emf2* and *vrn2* mutants revealed that sporophytic PcG complexes are haploinsufficient to suppress autonomous seed coat formation, revealing a dosage-sensitive function of plant PcG proteins similar to the fine-tuned dosage-sensitive requirement of PcG proteins in animals (Mollaaghababa et al., 2001). Haploinsufficiency of the double heterozygous *emf2/+*

64

vrn2/+ mutant depleted in two redundantly acting genes, explains enhancement of the FIS phenotype in the triple heterozygous mutant *fis2/+ emf2/+ vrn2/+*. The strongest support for non-redundant action of FIS2, EMF2 and VRN2 was derived by expression of *EMF2* and *VRN2* genes driven by the *FIS2* promoter in the *fis2* mutant. EMF2 and VRN2 did not rescue *fis2* seed abortion after fertilization, strongly supporting the view that EMF2 and VRN2 are not components of the FIS complex even in the absence of the FIS2 protein.

5.3. Autonomous Seeds of *fie* and *msi1* Develop a Seed Coat Due to their Role in Sporophytic Tissues

Fertilization independent seeds are characterized by autonomously developing endosperm, however, based on the development of the seed coat they can be grouped into two subclasses. While most of the autonomous seeds in *fie* and *msi1* mutants initiate development of the seed coat, this is not the case for most of autonomous seeds of *fis2* and *mea*. The seed coat of autonomously developing seeds does not differ from wild-type seed coat in its early developmental stages. Proanthocyanidin biosynthesis – a hallmark of seed coat development initiation is detectable early in all cells of the endothelium. Two subsequent layers of the inner integument undergo programed cell death like in wild-type fertilized seeds. Also the outer integument differentiates and the columella structures are detectable in the cells of the outer seed coat layer. These morphological features of the seed coat of autonomous seeds are indistinguishable from normal zygotic seeds (Chaudhury et al., 1997). Although *fis2* and *mea* mutants produce autonomous seeds, I observed that only a minor fraction initiates seed coat development. Seven days after emasculation when siliques were dissected and investigated for autonomous seed development, the integuments of *fis2* and *mea* autonomous seeds were degenerated. All five layers surrounding the developing endosperm collapsed, often forming a thick layer of cell walls that were protecting the developing seed from complete collapse. The process of integument degeneration in *fis2* and *mea* resembles the degeneration of the integuments in unfertilized ovules. In the absence of fertilization the mature wild-type ovules persisted unaffected until the 4th-5th day after emasculation, corresponding to about 2-3 days post anthesis. After that time cell layers of both integuments quickly collapsed. Degeneration of unfertilized ovules correlates in time with degeneration of the papilla cells of the stigma. These two processes together limit the time when pollination and fertilization can occur

while increased longevity of these organs is a desirable agronomic trait (Carbonell-Bejerano et al., 2010).

5.4. The Requirement of the Seed Coat Initiation Signal is Partially Bypassed in PcG Mutants

The results from this work strongly suggest that upon fertilization a mobile signal is formed in the endosperm that migrates to the integuments and relieves PcG repression at specific target loci. The importance of the endosperm for seed coat development has been implicated before (Garcia et al., 2003; Weijers et al., 2003; Garcia et al., 2005; Luo et al., 2005; Ingouff et al., 2006), however, this study makes a significant advance revealing that the sexual endosperm is required for seed coat formation, whereas asexually formed endosperm rarely suffices to initiate its formation. This implicates that signal formation is initiated after fertilization and that formation of the signal is not or not sufficiently initiated in the asexual endosperm of *fis* mutants. To test if the seed coat develops independently of endosperm formation in PcG deficient mutants I investigated the sporophytic role of VRN2 and FIE. While unfertilized wild-type ovules collapsed quickly after anthesis, homozygous *vrn2* mutant ovules persisted longer and initiated proanthocyanidin synthesis, indicating onset of seed coat development. Consistently, expression of a *35Spro:FIE* construct in the integuments suppressed development of the seed coat but not of the endosperm in the autonomously developing seeds. Together, I demonstrated that seed coat formation is initiated in mutants of sporophytically acting PcG genes, implicating that upon loss of sporophytic PcG function formation of a signal is initiated in the integuments independently of a sexual endosperm. Alternatively, processes downstream of the signal are initiated upon loss of PcG function. Importantly though, loss of sporophytic PcG function only initiates proanthocyanidin formation as one of the first detectable processes of seed coat differentiation, whereas complete differentiation of seed coat layers was not observed. Interestingly, preliminary data of the analysis of a *fis2 swn/-* double mutant suggests that seed coat development can proceed beyond initial proanthocyanidin synthesis in the absence of a developing endosperm. Observation of autonomous seeds with developed seed coat but without visible endosperm nuclei implicate that indeed seed coat development can occur without a developing endosperm. However, detailed further studies are required to substantiate these findings.

5.5. Central Role of AGL62 in the Process of Signal Formation

Analysis of the literature describing mutants affected in integument growth and differentiation revealed a number of mutants with seeds containing early arrested globular embryos, not dividing central cells and early degenerating integuments. Published data are consistent with the hypothesis that seed coat development is induced by the growing endosperm and not the embryo (Figure 5-1). Finally, I discovered the type I MADS-box transcription factor AGL62 as a candidate for being a central component required for the formation of the mobile signal. Seeds lacking AGL62 fail to initiate seed coat formation, similar to autonomously developing *fis2* and *mea* seeds as well as seeds containing only a developing embryo but no endosperm. It is unlikely that this is a consequence of *agl62* endosperm proliferation failure and early endosperm cellularization, as the *ttn2* mutant that is defective in nuclear proliferation still develops a normal seed coat. Conversely, defects in seed coat growth negatively impact on endosperm growth (Garcia et al., 2003; Garcia et al., 2005), implicating that impaired seed coat development is cause rather than consequence of early *agl62* endosperm arrest. Furthermore, I could show that AGL62 is unlikely to be the signal itself, as expression of AGL62-GFP after fertilization remained confined to the endosperm and was not detected in the seed coat. AGL62 has been shown to interact with type I MADS-box proteins including the specifically paternally expressed PHERES1 (de Folter et al., 2005), suggesting that an AGL62 containing protein complex is required to activate downstream target genes to form the seed coat initiation signal. When tested, the *phe1* mutation did not show a phenotype similar to *agl62*, likely because *PHE1* acts redundantly with its close homologue *PHE2*. The effect of the *agl62* mutation on seed development turned out to be even stronger when analyzed in autonomous seeds. In combination with *fie,* mutant ovules of *fie agl62* genotype did not even initiate divisions of the central cell. This indicates that *AGL62* is expressed in the central cell before fertilization, although AGL62-GFP was only detected in the endosperm after fertilization. Furthermore, it implicates the presence of additional factors in the fertilized zygote that enable initiation of endosperm development after fertilization in *agl62* mutant seeds. AGL62-GFP reporter line analysis revealed a GFP signal in antipodal cells in the developing female gametophyte before fertilization ((Kang et al., 2008) and my observation). It is possible that before fertilization the AGL62 protein is expressed in the central cell but lacks binding partners that would facilitate its transportation to the central

cell nucleus. At the same time AGL62 could be transported to antipodal cells by interacting partners. Why *AGL62pro:AGL62-GFP* is not detected in the central cell nucleus before fertilization where it is indispensable for the process of autonomous endosperm formation remains to be elucidated. Nuclear localization of another, central cell specific, type I MADS box protein AGL61 depends on the presence of the AGL80 protein, which has been shown to interact also with AGL62 in yeast two-hybrid experiments (Kang et al., 2008; Steffen et al., 2008). Together, it indicates that the described AGL proteins are transported to the nucleus in complexes, in a tightly regulated way and only at the time when they are needed to facilitate central cell nucleus and endosperm development.
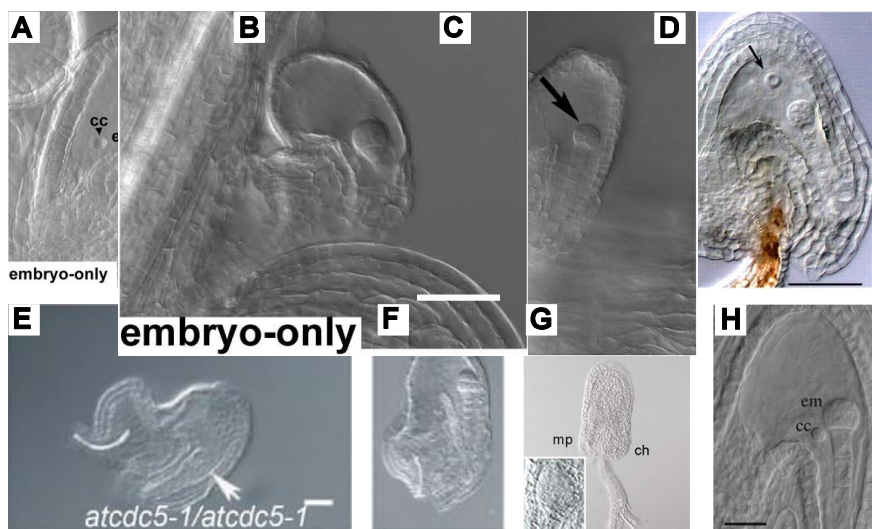

Figure 5-1. Seeds without an endosperm do not develop a seed coat. **(A and B)** Single fertilization event in ovules fertilized with pollen expressing Diphtheria Toxin A (Frank and Johnson, 2009) **(C)** Embryo development in wild type-plant fertilized with *fbl17-1* mutant (Gusti et al., 2009) **(D)** *glc* seed without endosperm and with embryo arrested at octan stage (Ngo et al., 2007) **(E)** *atcdc5-1* (-/-) seed at 72 hours after pollination (Lin et al., 2007), **(F)** Embryo at globular stage in *fbl17* mutant seed (Kim et al., 2008) **(G)** Collapsed seed containing heart stage embryo and diphtheria toxin-mediated ablation of the endosperm (Weijers et al., 2003) **(H)** Single fertilization event in *msi1* mutant. Seed with an embryo but without endosperm, 3 DAP. At the side of the embryo the nucleus of central cell that remains unfertilized is still observed (Chen et al., 2008)

5.6. A Signal from the Sexual Endosperm Initiates Seed Coat Cell Elongation

Growth of plant tissues is generally characterized by two phases: initial proliferation and subsequent cell elongation. Shoot lateral organs are initiated at the shoot apical meristem from a group of long-living stem cells and their immediate daughters at the growing tip of the plant. Before floral transition, the meristem produces leaves, while after that change the shoot meristem gives rise to floral meristems, which then initiate floral organs. After initiation, the organs grow by cell proliferation which encompasses coupled accumulation of cytoplasmic mass, increase in cell size, and subsequent cell division giving rise to two daughter cells. Eventually, proliferation ceases and further growth occurs by cell expansion until the final size and shape is reached. The cell cycle control that underlies the proliferative phase of organ growth is much better studied and understood than the process of cell elongation. The list of transcription factors being involved in the control of the cell divisions is much longer than the very limited number of known factors controlling initiation of the cell expansion (Breuninger and Lenhard, 2010). The process of cell expansion is driven by loosening of cell walls and uptake of water into the vacuole. Down-regulation of the cell wall associated expansin proteins that are responsible for breaking noncovalent bonds between cell wall polysaccharides, leads to reduced cell wall extensibility and in turn reduced cell and organ size (Cho and Cosgrove, 2000). Cell growth is also accompanied by endoreduplication, increasing the ploidy level of cells. There is a positive correlation between cell size and DNA content, thus mutations in genes regulating endoreduplication like *FIZZY-RELATED* (*FZR*) indirectly cause a reduction of cell size (Larson-Rabin et al., 2009). DELLA proteins are another group of factors inhibiting cell elongation (Schwechheimer and Willige, 2009). They specifically block the DNA-binding domain of PHYTOCHROM INTERACTING FACTORs (PIFs), a subfamily of bHLH transcription factors that are involved in the regulation of light-regulated genes (de Lucas et al., 2008). Control of the well-studied process of hypocotyl elongation in the dark involves the action of these genes. DELLA protein degradation is induced by the phytohormone gibberellic acid (GA), causing activation of PIFs and resulting in expansion of the hypocotyl cells in the absence of light. In the light PIFs are rapidly degraded, causing termination of cell expansion. DELLA proteins are also degraded by auxin (Fu and Harberd, 2003) and stabilized by ethylene (Achard et al., 2003), indicating an integrative role of these proteins in the responses to hormones and environmental cues. Whether any of

69

these regulatory processes is a primary target of the signal originating from the sexual endosperm remains to be elucidated.

Signaling from the sexual endosperm to the surrounding integuments induces their growth in all cell layers. Previous studies reported that signal-induced initial growth of the seed coat combines mitotic cell divisions as well as cell expansion (Western et al., 2000) and (Garcia et al., 2005). In my study I show that in the absence of fertilization the maturing wild-type ovule undergoes rapid mitotic divisions and that this process declines after the time of anthesis (2-3 DAE), but does not terminate immediately. Garcia and colleagues observed ceasing of mitotic divisions at 3 DAP, what in their study corresponds to 3.5 DAE. There is no obvious change in endothelium cell numbers of mature ovules and developing seeds (pictures from the figure 4-4 E, H and J), suggesting that mitotic divisions during early seed coat development are independent from any signal sent by the sexual endosperm. Conversely, induction of cell expansion and differentiation fully relies on the perception of the signal. The signal is unlikely to involve increased mechanical pressure through endosperm growth, as I observed enlarged autonomous seeds containing only two or four endosperm nuclei and small seeds without developing seed coat containing up to 64 nuclei.

5.7. Are Plant Hormones Involved in Endosperm - Seed Coat Signaling?

One of the first detectable changes in the developing seed coat is accumulation of proanthocyanidins in the endothelium layer. PA biosynthesis starts in the cells surrounding the micropylar endosperm and proceeds in a wave-like manner throughout the endothelium until it reaches the opposite end surrounding the chalazal endosperm (Debeaujon et al., 2003). The PA accumulation pattern suggests initial signaling from the fertilized central cell or early endosperm to the neighboring integument cells. Does spreading of PA accumulation in the seed coat reflects continuous signalling from the developing endosperm or does it result from signal amplification within the cells of the seed coat independently from the endosperm? To distinguish between both possibilities will require knowing the nature of the signal. Transmission electron microscopy of osmium tetroxide stained seeds revealed the presence of an electron-dense, lipid layer on the endothelium surface facing the embryo sac (Beeckman et al., 2000). This cell wall impregnation has been considered as an original cuticle of the inner integument and it was

not found on the other integument layers at any time. It raises the question whether signals from the endosperm can pass through this barrier and stimulate growth of the integuments. On the other hand it is believed that although there are no symplastic connections between the inner integument and the endosperm, assimilates, like sucrose are delivered from the phloem through the outer and subsequently inner integuments to the endosperm and embryo using active transporters (Stadler et al., 2005). The absence of plasmodesmata connections between seed coat and endosperm would rather rule out the possibility of small RNA signaling between endosperm and seed coat since their intercellular movement relies on plasmodesmata connections (Kehr and Buhtz, 2008).

Transcriptional profiling of unfertilized ovules and early wild-type seeds as well as ovules of the *vrn2 emf2* mutant with first symptoms of endothelium differentiation gave first hints about the possible nature of the signal. Initial data mining revealed that upregulated genes involved in cell wall loosening and modification and cell growth were not regulated by GA or IAA treatment, implicating that GA and IAA are unlikely to be part of the signal (Hruz et al., 2008). Most of the genes commonly deregulated in the wild-type seeds and mutant ovules were highly upregulated after treatment with Jasmonic Acid (JA), suggesting a possible link of the signal with JA signalling.

6. REFERENCES

Achard, P., Vriezen, W.H., Van Der Straeten, D., and Harberd, N.P. (2003). Ethylene Regulates Arabidopsis Development via the Modulation of DELLA Protein Growth Repressor Function. The Plant Cell **15,** 2816-2825.

Acosta-García, G., and Vielle-Calzada, J.-P. (2004). A Classical Arabinogalactan Protein Is Essential for the Initiation of Female Gametogenesis in Arabidopsis. The Plant Cell **16,** 2614-2628.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. Science **301,** 653-657.

Baker, S.C., Robinson-Beers, K., Villanueva, J.M., Gaiser, J.C., and Gasser, C.S. (1997). Interactions Among Genes Regulating Ovule Development in *Arabidopsis thaliana*. Genetics **145,** 1109-1124.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. The Plant Journal **39,** 366-380.

Beeckman, T., De Rycke, R., Viane, R., and Inzé, D. (2000). Histological Study of Seed Coat Development in *Arabidopsis thaliana*. Journal of Plant Research **113,** 139-148.

Boisnard-Lorig, C., Colon-Carmona, A., Bauch, M., Hodge, S., Doerner, P., Bancharel, E., Dumas, C., Haseloff, J., and Berger, F. (2001). Dynamic Analyses of the Expression of the HISTONE::YFP Fusion Protein in Arabidopsis Show That Syncytial Endosperm Is Divided in Mitotic Domains. The Plant Cell **13,** 495-509.

Braselton, J.P., Wilkinson, M.J., and Clulow, S.A. (1996). Feulgen Staining of Intact Plant Tissues for Confocal Microscopy. Biotechnic & Histochemistry **71,** 84-87.

Breuninger, H., and Lenhard, M. (2010). Chapter Seven - Control of Tissue and Organ Growth in Plants. In Current Topics in Developmental Biology, C.P.T. Marja, ed (Academic Press), pp. 185-220.

Brown, R.C., Lemmon, B.E., and Nguyen, H. (2003). Events during the first four rounds of mitosis establish three developmental domains in the syncytial endosperm of *Arabidopsis thaliana*. Protoplasma **222,** 167-174.

Brown, R.C., Lemmon, B.E., Nguyen, H., and Olsen, O.-A. (1999). Development of endosperm in *Arabidopsis thaliana*. Sexual Plant Reproduction **12,** 32-42.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Current Opinion in Genetics & Development **14,** 155-164.

Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing. Molecular and Cellular Biology **28,** 1862-1872.

Carbonell-Bejerano, P., Urbez, C., Carbonell, J., Granell, A., and Perez-Amador, M.A. (2010). A Fertilization-Independent Developmental Program Triggers Partial Fruit Development and Senescence Processes in Pistils of Arabidopsis. Plant Physiology **154,** 163-172.

Chandler, J., Wilson, A., and Dean, C. (1996). Arabidopsis mutants showing an altered response to vernalization. The Plant Journal **10,** 637-644.

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z.R., and Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development **131,** 5263-5276.

Chapman, L.A., and Goring, D.R. (2010). Pollen–pistil interactions regulating successful fertilization in the *Brassicaceae*. Journal of Experimental Botany **61,** 1987-1999.

Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **94,** 4223-4228.

Chen, L.-J., Diao, Z.-Y., Specht, C., and Sung, Z.R. (2009). Molecular Evolution of VEF-Domain-Containing PcG Genes in Plants. Molecular Plant **2,** 738-754.

Chen, L., Cheng, J.C., Castle, L., and Sung, Z.R. (1997). EMF Genes Regulate Arabidopsis Inflorescence Development. The Plant Cell **9,** 2011-2024.

Chen, Z., Tan, J.L.H., 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.

Cho, H.-T., and Cosgrove, D.J. (2000). Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **97,** 9783-9788.

Choi, J., Hyun, Y., Kang, M.-J., In Yun, H., Yun, J.-Y., Lister, C., Dean, C., Amasino, R.M., Noh, B., Noh, Y.-S., and Choi, Y. (2009). Resetting and regulation of FLOWERING LOCUS C expression during Arabidopsis reproductive development. The Plant Journal **57,** 918-931.

Costa, L.M., Gutièrrez-Marcos, J.F., and Dickinson, H.G. (2004). More than a yolk: the short life and complex times of the plant endosperm. Trends in Plant Science **9,** 507-514.

de Folter, S., Immink, R.G.H., Kieffer, M., Parenicová, L., Henz, S.R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M.M., Davies, B., and Angenent, G.C. (2005). Comprehensive Interaction Map of the Arabidopsis MADS Box Transcription Factors. The Plant Cell **17,** 1424-1433.

de Lucas, M., Daviere, J.-M., Rodriguez-Falcon, M., Pontin, M., Iglesias-Pedraz, J.M., Lorrain, S., Fankhauser, C., Blazquez, M.A., Titarenko, E., and Prat, S. (2008). A molecular framework for light and gibberellin control of cell elongation. Nature **451,** 480-484.

De Lucia, F., Crevillen, P., Jones, A.M.E., Greb, T., and Dean, C. (2008). A PHD-Polycomb Repressive Complex 2 triggers the epigenetic silencing of FLC during vernalization. Proceedings of the National Academy of Sciences **105,** 16831-16836.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M., and Lepiniec, L. (2003). Proanthocyanidin-Accumulating Cells in Arabidopsis Testa: Regulation of Differentiation and Role in Seed Development. The Plant Cell **15,** 2514-2531.

Dixon, R.A., Xie, D.-Y., and Sharma, S.B. (2005). Proanthocyanidins – a final frontier in flavonoid research? New Phytologist **165,** 9-28.

Drews, G.N., and Yadegari, R. (2002). DEVELOPMENT AND FUNCTION OF THE ANGIOSPERM FEMALE GAMETOPHYTE. Annual Review of Genetics **36,** 99-124.

Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research **19(6),** 1349-1349.

Elliott, R.C., Betzner, A.S., Huttner, E., Oakes, M.P., Tucker, W., Gerentes, D., Perez, P., and Smyth, D.R. (1996). *AINTEGUMENTA*, an *APETALA2-like* Gene of Arabidopsis with Pleiotropic Roles in Ovule Development and Floral Organ Growth. The Plant Cell **8,** 155-168.

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Scheid, O.M., Hennig, L., and Köhler, C. (2009). Imprinting of the Polycomb Group Gene *MEDEA* Serves as a Ploidy Sensor in Arabidopsis. PLoS Genetics **5,** e1000663.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science **306,** 1574-1577.

Frank, A.C., and Johnson, M.A. (2009). Expressing the Diphtheria Toxin A Subunit from the HAP2(GCS1) Promoter Blocks Sperm Maturation and Produces Single Sperm-Like Cells Capable of Fertilization. Plant Physiology **151,** 1390-1400.

Fu, X., and Harberd, N.P. (2003). Auxin promotes Arabidopsis root growth by modulating gibberellin response. Nature **421,** 740-743.

Garcia, D., Fitz Gerald, J.N., and Berger, F. (2005). Maternal Control of Integument Cell Elongation and Zygotic Control of Endosperm Growth Are Coordinated to Determine Seed Size in Arabidopsis. The Plant Cell **17,** 52-60.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G., and Berger, F. (2003). Arabidopsis *haiku* Mutants Reveal New Controls of Seed Size by Endosperm. Plant Physiology **131,** 1661-1670.

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., and Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature **386,** 44-51.

Grossniklaus, U., Vielle-Calzada, J.-P., Hoeppner, M.A., and Gagliano, W.B. (1998). Maternal Control of Embryogenesis by MEDEA, a Polycomb Group Gene in Arabidopsis. Science **280,** 446-450.

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

Guitton, A.-E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.-E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. Development **131,** 2971-2981.

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.

Haig, D., and Westoby, M. (1989). Parent-Specific Gene Expression and the Triploid Endosperm. The American Naturalist **134,** 147-155.

Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, Masahiro M., Sasaki, N., Nakano, A., Berger, F., and Higashiyama, T. (2011). Live-Cell Imaging Reveals the Dynamics of Two Sperm Cells during Double Fertilization in *Arabidopsis thaliana*. Current Biology **21,** 497-502.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends in Plant Science **10,** 472-477.

Higashiyama, T., and Hamamura, Y. (2008). Gametophytic pollen tube guidance. Sexual Plant Reproduction **21,** 17-26.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S.-y., Kuroiwa, H., and Kuroiwa, T. (2001). Pollen Tube Attraction by the Synergid Cell. Science **293,** 1480-1483.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes. Advances in Bioinformatics **2008**.

Hulskamp, M., Schneitz, K., and Pruitt, R.E. (1995). Genetic Evidence for a Long-Range Activity That Directs Pollen Tube Guidance in Arabidopsis. The Plant Cell **7,** 57-64.

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics **5,** 299-314

Ingouff, M., Jullien, P.E., and Berger, F. (2006). The Female Gametophyte and the Endosperm Control Cell Proliferation and Differentiation of the Seed Coat in Arabidopsis. The Plant Cell **18,** 3491-3501.

Jiang, D., Wang, Y., Wang, Y., and He, Y. (2008). Repression of *FLOWERING LOCUS C* and *FLOWERING LOCUS T* by the *Arabidopsis* Polycomb Repressive Complex 2 Components. PLoS ONE **3,** e3404.

Jofuku, K.D., Boer, B., Montagu, M.V., and Okamuro, J.K. (1994). Control of Arabidopsis Flower and Seed Development by the Homeotic Gene *APETALA2*. The Plant Cell **6,** 1211-1225.

Johnson, C.S., Kolevski, B., and Smyth, D.R. (2002). *TRANSPARENT TESTA GLABRA2*, a Trichome and Seed Coat Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. The Plant Cell **14,** 1359-1375.

Johnson, M., and Lord, E. (2006). Extracellular Guidance Cues and Intracellular Signaling Pathways that Direct Pollen Tube Growth, R. Malhó, ed (Springer Berlin / Heidelberg), pp. 223-242.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature **316,** 153-155.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb Complexes and the Propagation of the Methylation Mark at the Drosophila Ubx Gene. Journal of Biological Chemistry **281,** 29064-29075.

Kang, I.-H., Steffen, J.G., Portereiko, M.F., Lloyd, A., and Drews, G.N. (2008). The AGL62 MADS Domain Protein Regulates Cellularization during Endosperm Development in Arabidopsis. The Plant Cell **20,** 635-647.

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

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 SCFFBL17 degradation of cell cycle inhibitors. Nature **455,** 1134-1137.

Kinoshita-Tsujimura, K., and Kakimoto, T. (2011). Cytokinin receptors in sporophytes are essential for male and female functions in *Arabidopsis thaliana*. Plant Signaling and Behavior **6,** 66-71.

Kinoshita, T., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2001). Polycomb repression of flowering during early plant development. Proceedings of the National Academy of Sciences **98,** 14156-14161.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999). Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in Arabidopsis. Proceedings of the National Academy of Sciences **96,** 4186-4191.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes & Development **20,** 1110-1122.

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

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003a). Arabidopsis MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. EMBO Journal **22,** 4804-4814.

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

Larson-Rabin, Z., Li, Z., Masson, P.H., and Day, C.D. (2009). *FZR2/CCS52A1* Expression Is a Determinant of Endoreduplication and Cell Expansion in Arabidopsis. Plant Physiology **149,** 874-884.

Lau, S., Ehrismann, J.S., Schlereth, A., Takada, S., Mayer, U., and Jürgens, G. (2010). Cell–cell communication in Arabidopsis early embryogenesis. European Journal of Cell Biology **89,** 225-230.

Laux, T., Würschum, T., and Breuninger, H. (2004). Genetic Regulation of Embryonic Pattern Formation. The Plant Cell **16,** S190-S202.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., and Caboche, M. (2006). Genetics and Biochemistry of Seed Flavonoids. Annual Review of Plant Biology **57,** 405-430.

Leroy, O., Hennig, L., Breuninger, H., Laux, T., and Köhler, C. (2007). Polycomb group proteins function in the female gametophyte to determine seed development in plants. Development **134,** 3639-3648.

Levine, S.S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R.E. (2002). The Core of the Polycomb Repressive Complex Is Compositionally and Functionally Conserved in Flies and Humans. Molecular and Cellular Biology **22,** 6070-6078.

Li, S.F., Milliken, O.N., Pham, H., Seyit, R., Napoli, R., Preston, J., Koltunow, A.M., and Parish, R.W. (2009). The Arabidopsis MYB5 Transcription Factor Regulates Mucilage Synthesis, Seed Coat Development, and Trichome Morphogenesis. The Plant Cell **21,** 72-89.

Lin, Z., Yin, K., Zhu, D., Chen, Z., Gu, H., and Qu, L.-J. (2007). AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of Arabidopsis shoot apical meristem. Cell Research **17,** 815-828.

Linkies, A., Graeber, K., Knight, C., and Leubner-Metzger, G. (2010). The evolution of seeds. New Phytologist **186,** 817-831.

Liu, C.-m., and Meinke, D.W. (1998). The *titan* mutants of Arabidopsis are disrupted in mitosis and cell cycle control during seed development. The Plant Journal **16,** 21-31.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J., and Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proceedings of the National Academy of Sciences **97,** 10637-10642.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J., and Chaudhury, A. (2005). *MINISEED3 (MINI3)*, a WRKY family gene, and *HAIKU2 (IKU2)*, a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **102,** 17531-17536.

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

Mollaaghababa, R., Sipos, L., Tiong, S.Y.K., Papoulas, O., Armstrong, J.A., Tamkun, J.W., and Bender, W. (2001). Mutations in Drosophila heat shock cognate 4 are enhancers of Polycomb. Proceedings of the National Academy of Sciences **98,** 3958-3963.

Müller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Current Opinion in Genetics & Development **16,** 476-484.

Nakaune, S., Yamada, K., Kondo, M., Kato, T., Tabata, S., Nishimura, M., and Hara-Nishimura, I. (2005). A Vacuolar Processing Enzyme, δVPE, Is Involved in Seed Coat Formation at the Early Stage of Seed Development. The Plant Cell **17,** 876-887.

Nawaschin, S.G. (1898). Resultate einer revision der befruchtungsvorgaenge bei Lilium martagon und Fritillaria tenella. Bul. Acad. Imp. Sci. St Petersburg **9,** 377-382.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Muller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO Journal **26,** 4078-4088.

Nesi, N., Debeaujon, I., Jond, C., Stewart, A.J., Jenkins, G.I., Caboche, M., and Lepiniec, L. (2002). The *TRANSPARENT TESTA16* Locus Encodes the ARABIDOPSIS BSISTER MADS

Domain Protein and Is Required for Proper Development and Pigmentation of the Seed Coat. The Plant Cell **14,** 2463-2479.

Ngo, Q.A., Moore, J.M., Baskar, R., Grossniklaus, U., and Sundaresan, V. (2007). Arabidopsis GLAUCE promotes fertilization-independent endosperm development and expression of paternally inherited alleles. Development **134,** 4107-4117.

Nguyen, H., Brown, R.C., and Lemmon, B.E. (2000). The specialized chalazal endosperm in *Arabidopsis thaliana* and *Lepidium virginicum*; (*Brassicaceae*). Protoplasma **212,** 99-110.

Nowack, M.K., Ungru, A., Bjerkan, K., Grini, P.E., and Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. Biochemical Society Transactions **38,** 604–612.

Oh, S., Park, S., and van Nocker, S. (2008). Genic and Global Functions for Paf1C in Chromatin Modification and Gene Expression in Arabidopsis. PLoS Genetics **4,** e1000077.

Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. Proceedings of the National Academy of Sciences **93,** 5319-5324.

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.

Olsen, O.-A. (2004). Nuclear Endosperm Development in Cereals and *Arabidopsis thaliana*. The Plant Cell **16,** S214-S227.

Otegui, M., and Staehelin, L.A. (2000). Syncytial-Type Cell Plates: A Novel Kind of Cell Plate Involved in Endosperm Cellularization of Arabidopsis. The Plant Cell **12,** 933-947.

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.

Papini, A., Mosti, S., Milocani, E., Tani, G., Di Falco, P., and Brighigna, L. (2011). Megasporogenesis and programmed cell death in *Tillandsia (Bromeliaceae)*. Protoplasma **248,** 651-662.

Peris, C.I.L., Rademacher, E.H., and Weijers, D. (2010). Chapter One - Green Beginnings – Pattern Formation in the Early Plant Embryo. In Current Topics in Developmental Biology, C.P.T. Marja, ed (Academic Press), pp. 1-27.

Pourcel, L., Routaboul, J.-M., Kerhoas, L., Caboche, M., Lepiniec, L., and Debeaujon, I. (2005). *TRANSPARENT TESTA10* Encodes a Laccase-Like Enzyme Involved in Oxidative Polymerization of Flavonoids in Arabidopsis Seed Coat. The Plant Cell **17,** 2966-2980.

Rehrauer, H., Aquino, C., Gruissem, W., Henz, S.R., Hilson, P., Laubinger, S., Naouar, N., Patrignani, A., Rombauts, S., Shu, H., Van de Peer, Y., Vuylsteke, M., Weigel, D., Zeller, G., and Hennig, L. (2010). AGRONOMICS1: A New Resource for Arabidopsis Transcriptome Profiling. Plant Physiology **152,** 487-499.

Robinson-Beers, K., Pruitt, R.E., and Gasser, C.S. (1992). Ovule Development in Wild-Type Arabidopsis and Two Female-Sterile Mutants. The Plant Cell **4,** 1237-1249.

Rodkiewicz, B. (1970). Callose in cell walls during megasporogenesis in angiosperms. Planta **93,** 39-47.

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

Sambrook, J., and Russell, D. (2001). Molecular cloning: A laboratory manual. (Cold Spring Harbor Laboratory Press).

Schubert, D., Clarenz, O., and Goodrich, J. (2005). Epigenetic control of plant development by Polycomb-group proteins. Current Opinion in Plant Biology **8,** 553-561.

Schuettengruber, B., and Cavalli, G. (2009). Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development **136,** 3531-3542.

Schwechheimer, C., and Willige, B.C. (2009). Shedding light on gibberellic acid signalling. Current Opinion in Plant Biology **12,** 57-62.

Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G. (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. Development **125,** 3329-3341.

Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response. Genes & Development **20,** 3079-3083.

Shirley M, T. (1999). The Sins of the Fathers and Mothers: Genomic Imprinting in Mammalian Development. Cell **96,** 185-193.

Shirzadi, R., Andersen, E.D., Bjerkan, K.N., Gloeckle, B.M., Heese, M., Ungru, A., Winge, P., Koncz, C., Aalen, R.B., Schnittger, A., and Grini, P.E. (2011). Genome-Wide Transcript Profiling of Endosperm without Paternal Contribution Identifies Parent-of-Origin–Dependent Regulation of *AGAMOUS-LIKE36*. PLoS Genetics **7,** e1001303.

Skinner, D.J., Hill, T.A., and Gasser, C.S. (2004). Regulation of Ovule Development. The Plant Cell **16,** S32-S45.

Smyth, D.R., Bowman, J.L., and Meyerowitz, E.M. (1990). Early Flower Development in Arabidopsis. The Plant Cell **2,** 755-767.

Smyth, G. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology **3,** Article 3.

Spillane, C., Schmid, K.J., Laoueille-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D.R., Wolfe, K.H., and Grossniklaus, U. (2007). Positive darwinian selection at the imprinted MEDEA locus in plants. Nature **448,** 349-352.

Stadler, R., Lauterbach, C., and Sauer, N. (2005). Cell-to-Cell Movement of Green Fluorescent Protein Reveals Post-Phloem Transport in the Outer Integument and Identifies Symplastic Domains in Arabidopsis Seeds and Embryos. Plant Physiology **139,** 701-712.

Steffen, J.G., Kang, I.-H., Portereiko, M.F., Lloyd, A., and Drews, G.N. (2008). AGL61 Interacts with AGL80 and Is Required for Central Cell Development in Arabidopsis. Plant Physiology **148,** 259-268.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America **100,** 9440-9445.

Sundaresan, V. (2005). Control of seed size in plants. Proceedings of the National Academy of Sciences of the United States of America **102,** 17887-17888.

Tie, F., Furuyama, T., Prasad-Sinha, J., Jane, E., and Harte, P.J. (2001). The Drosophila Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. Development **128,** 275-286.

Twell, D. (2006). A blossoming romance: gamete interactions in flowering plants. Nature Cell Biology **8,** 14-16.

Vielle-Calzada, J.-P., Baskar, R., and Grossniklaus, U. (2000). Delayed activation of the paternal genome during seed development. Nature **404,** 91-94.

Wang, D., Tyson, M.D., Jackson, S.S., and Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in Arabidopsis. Proceedings of the National Academy of Sciences **103,** 13244-13249.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. Nature **431,** 873-878.

Weijers, D., van Hamburg, J.-P., van Rijn, E., Hooykaas, P.J.J., and Offringa, R. (2003). Diphtheria Toxin-Mediated Cell Ablation Reveals Interregional Communication during Arabidopsis Seed Development. Plant Physiology **133,** 1882-1892.

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., and Köhler, C. (2010). H3K27me3 Profiling of the Endosperm Implies Exclusion of Polycomb Group Protein Targeting by DNA Methylation. PLoS Genetics **6,** e1001152.

Western, T.L., Skinner, D.J., and Haughn, G.W. (2000). Differentiation of Mucilage Secretory Cells of the Arabidopsis Seed Coat. Plant Physiology **122,** 345-356.

Western, T.L., Young, D.S., Dean, G.H., Tan, W.L., Samuels, A.L., and Haughn, G.W. (2004). *MUCILAGE-MODIFIED4* Encodes a Putative Pectin Biosynthetic Enzyme Developmentally Regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the Arabidopsis Seed Coat. Plant Physiology **134,** 296-306.

Williams, J.H., and Friedman, W.E. (2004). The four-celled female gametophyte of Illicium (*Illiciaceae; Austrobaileyales*): implications for understanding the origin and early evolution of monocots, eumagnoliids,and eudicots. American Journal of Botany **91,** 332-351.

Winkel-Shirley, B. (2001). Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant Physiology **126,** 485-493.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm. PLoS Genetics **7,** e1002126.

Wood, C.C., Robertson, M., Tanner, G., Peacock, W.J., Dennis, E.S., and Helliwell, C.A. (2006). The *Arabidopsis thaliana* vernalization response requires a polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. Proceedings of the National Academy of Sciences **103,** 14631-14636.

Yang, C.-H., Chen, L.-J., and Sung, Z.R. (1995). Genetic Regulation of Shoot Development in Arabidopsis: Role of the *EMF* Genes. Developmental Biology **169,** 421-435.

Yang, W.-C., Ye, D., Xu, J., and Sundaresan, V. (1999). The *SPOROCYTELESS* gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes & Development **13,** 2108-2117.

Yoshida, N., Yanai, Y., Chen, L., Kato, Y., Hiratsuka, J., Miwa, T., Sung, Z.R., and Takahashi, S. (2001). EMBRYONIC FLOWER2, a Novel Polycomb Group Protein Homolog, Mediates Shoot Development and Flowering in Arabidopsis. The Plant Cell **13,** 2471-2481.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in *Arabidopsis*. PLoS Biology **5,** e129.

7. Appendix

Table 7-1. List of the genes being commonly upregulated in the wild-type seeds 2 DAP and *emf2-5/+ vrn2/-* ovules 4 DAE which were not significantly expressed before fertilization.

AT5G14700	cinnamoyl-CoA reductase-related
AT3G28455	"CLE25 (CLAVATA3/ESR-RELATED 25); receptor binding"
AT2G10940	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein
AT4G29360	glycosyl hydrolase family 17 protein
AT3G56220	transcription regulator
AT5G09980	PROPEP4 (Elicitor peptide 4 precursor)
AT5G67150	transferase family protein
AT2G45680	TCP family transcription factor, putative
AT1G14840	"ATMAP70-4 (microtubule-associated proteins 70-4); microtubule binding"
AT2G26440	pectinesterase family protein
AT2G33735	DNAJ heat shock N-terminal domain-containing protein
AT4G37650	"SHR (SHORT ROOT); transcription factor"
AT3G12270	methyltransferase
AT2G20560	DNAJ heat shock family protein
AT5G10920	argininosuccinate lyase, putative / arginosuccinase, putative
AT1G75030	ATLP-3 (*Arabidopsis* thaumatin-like protein 3)
AT2G47930	AGP26/ATAGP26 (ARABINOGALACTAN PROTEINS 26)
AT2G44640	NA
AT2G16780	MSI2 (NUCLEOSOME/CHROMATIN ASSEMBLY FACTOR GROUP C 2)
AT2G29570	"PCNA2 (PROLIFERATING CELL NUCLEAR ANTIGEN 2); DNA binding / DNA polymerase processivity factor"
AT3G02640	NA
AT3G25940	transcription factor S-II (TFIIS) domain-containing protein
AT5G49560	NA
AT1G78370	"ATGSTU20 (*Arabidopsis thaliana* Glutathione S-transferase (class tau) 20); glutathione transferase"
AT5G04530	beta-ketoacyl-CoA synthase family protein
AT2G02080	"ATIDD4 (ARABIDOPSIS THALIANA INDETERMINATE(ID)-DOMAIN 4); transcription factor"
AT4G28310	NA
AT5G05620	"ATGCP2/TUBG2 (GAMMA-TUBULIN); structural molecule"
AT1G09812	NA
AT3G56710	"SIB1 (SIGMA FACTOR BINDING PROTEIN 1); binding"
AT1G72240	NA
AT4G18970	GDSL-motif lipase/hydrolase family protein
AT3G12920	protein binding / zinc ion binding
AT3G52630	NA
AT1G08845	structural constituent of ribosome
AT5G17660	methyltransferase
AT3G05600	epoxide hydrolase, putative
AT2G03090	ATEXPA15 (ARABIDOPSIS THALIANA EXPANSIN A15)
AT1G47670	amino acid transporter family protein

AT1G17430	hydrolase, alpha/beta fold family protein
AT2G33510	protein binding
AT5G57440	"GS1 (GLYCEROL-3-PHOSPHATASE 2); hydrolase"
AT3G52750	"FTSZ2-2 (FtsZ2-2); structural molecule"
AT3G13740	URF 4-related
AT5G50150	NA
AT1G66430	pfkB-type carbohydrate kinase family protein
AT1G64185	lactoylglutathione lyase family protein / glyoxalase I family protein
AT5G40080	60S ribosomal protein-related
AT2G37510	RNA-binding protein, putative
AT1G53520	chalcone-flavanone isomerase-related
AT1G14720	"XTR2 (XYLOGLUCAN ENDOTRANSGLYCOSYLASE RELATED 2); hydrolase, acting on glycosyl bonds"
AT2G45340	leucine-rich repeat transmembrane protein kinase, putative
AT1G07370	"PCNA1 (PROLIFERATING CELLULAR NUCLEAR ANTIGEN); DNA binding / DNA polymerase processivity factor"
AT2G48030	endonuclease/exonuclease/phosphatase family protein
AT3G13674	NA
AT1G26770	ATEXPA10 (ARABIDOPSIS THALIANA EXPANSIN A10)
AT5G56530	NA
AT1G69530	ATEXPA1 (ARABIDOPSIS THALIANA EXPANSIN A1)
AT2G44840	"ATERF13/EREBP (ETHYLENE-RESPONSIVE ELEMENT BINDING FACTOR 13); DNA binding / transcription factor"
AT4G26790	GDSL-motif lipase/hydrolase family protein
AT1G09440	protein kinase family protein
AT4G24780	pectate lyase family protein
AT1G08560	"SYP111 (syntaxin 111); SNAP receptor"
AT4G16980	arabinogalactan-protein family
AT3G18850	"LPAT5; acyltransferase"
AT4G32980	"ATH1 (ARABIDOPSIS THALIANA HOMEOBOX GENE 1); transcription factor"
AT2G27775	NA
AT2G31725	NA
AT4G01790	ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein / ribonuclease P-related
AT4G23820	glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein
AT2G01520	MLP328 (MLP-LIKE PROTEIN 328)
AT1G02370	pentatricopeptide (PPR) repeat-containing protein
AT3G51240	"F3H (TRANSPARENT TESTA 6); naringenin 3-dioxygenase"
AT3G10150	"ATPAP16/PAP16 (purple acid phosphatase 16); acid phosphatase/ protein serine/threonine phosphatase"
AT3G48490	NA
AT3G13175	NA
AT1G14440	"ATHB31 (ARABIDOPSIS THALIANA HOMEOBOX PROTEIN 31); transcription factor"
AT2G04530	CPZ
AT2G43550	trypsin inhibitor, putative
AT3G22886	"MIR167A; miRNA"
AT2G26760	"CYCB1;4; cyclin-dependent protein kinase regulator"
AT4G22212	NA
AT1G22250	NA
AT1G23750	DNA-binding protein-related

AT5G62840	phosphoglycerate/bisphosphoglycerate mutase family protein
AT4G34800	auxin-responsive family protein
AT4G10470	NA
AT5G03545	unknown protein
AT1G70210	"CYCD1;1 (CYCLIN D1;1); cyclin-dependent protein kinase regulator"
AT2G36145	NA
AT5G66800	NA
AT5G67180	AP2 domain-containing transcription factor, putative
AT5G06150	"CYC1BAT (CYCLIN B 1;2); cyclin-dependent protein kinase regulator"
AT5G47240	"ATNUDT8 (*Arabidopsis thaliana* Nudix hydrolase homolog 8); hydrolase"
AT1G52040	MBP1 (MYROSINASE-BINDING PROTEIN 1)
AT5G54490	"PBP1 (PINOID-BINDING PROTEIN 1); calcium ion binding"
AT5G02890	transferase family protein
AT1G26210	NA
AT1G62480	vacuolar calcium-binding protein-related
AT1G15260	NA
AT4G16140	proline-rich family protein
AT4G13840	transferase family protein
AT5G25830	zinc finger (GATA type) family protein
AT5G41460	fringe-related protein
AT2G22170	lipid-associated family protein
AT1G22330	RNA binding
AT4G15210	"ATBETA-AMY (BETA-AMYLASE); beta-amylase"
AT1G55205	NA
AT1G79110	protein binding / zinc ion binding
AT5G08330	TCP family transcription factor, putative
AT1G31320	LBD4 (LOB DOMAIN-CONTAINING PROTEIN 4)
AT3G58850	"PAR2 (PHY RAPIDLY REGULATED 2); transcription regulator"
AT5G55620	NA
AT4G29905	NA
AT5G62627	NA
AT5G44420	PDF1.2 (Low-molecular-weight cysteine-rich 77)
AT4G30410	transcription factor
AT1G04520	33 kDa secretory protein-related
AT1G64780	"ATAMT1;2 (AMMONIUM TRANSPORTER 1;2); ammonium transmembrane transporter"
AT2G32487	unknown protein
AT3G28220	meprin and TRAF homology domain-containing protein / MATH domain-containing protein
AT1G09470	NA
AT1G25220	"ASB1 (ANTHRANILATE SYNTHASE BETA SUBUNIT 1); anthranilate synthase" (auxin)
AT5G46690	"BHLH071 (BETA HLH PROTEIN 71); DNA binding / transcription factor"
AT4G21870	26.5 kDa class P-related heat shock protein (HSP26.5-P)
AT4G30650	hydrophobic protein, putative / low temperature and salt responsive protein, putative
AT2G41090	calmodulin-like calcium-binding protein, 22 kDa (CaBP-22)
AT4G34760	auxin-responsive family protein

These two articles were published together with other colleagues during the course of my studies.

"H3K27me3 Profiling of the Endosperm Implies Exclusion of Polycomb Group Protein Targeting by DNA Methylation" (2010) Isabelle Weinhofer, Elisabeth Hehenberger, Pawel Roszak, Lars Hennig, Claudia Köhler. PLoS Genet 6, e1001152.

"High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm" (2011) Philip Wolff, Isabelle Weinhofer, Jonathan Seguin, Pawel Roszak, Christian Beisel, Mark T. A. Donoghue, Charles Spillane, Magnus Nordborg, Marc Rehmsmeier, Claudia Köhler. PLoS Genet 7, e1002126.

PLoS GENETICS

H3K27me3 Profiling of the Endosperm Implies Exclusion of Polycomb Group Protein Targeting by DNA Methylation

Isabelle Weinhofer[1], Elisabeth Hehenberger[1], Pawel Roszak[1], Lars Hennig[1,2], Claudia Köhler[1,2]*

1 Department of Biology and Zurich-Basel Plant Science Center, Swiss Federal Institute of Technology, Zurich, Switzerland, 2 Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

Polycomb group (PcG) proteins act as evolutionary conserved epigenetic mediators of cell identity because they repress transcriptional programs that are not required at particular developmental stages. Each tissue is likely to have a specific epigenetic profile, which acts as a blueprint for its developmental fate. A hallmark for Polycomb Repressive Complex 2 (PRC2) activity is trimethylated lysine 27 on histone H3 (H3K27me3). In plants, there are distinct PRC2 complexes for vegetative and reproductive development, and it was unknown so far whether these complexes have target gene specificity. The FERTILIZATION INDEPENDENT SEED (FIS) PRC2 complex is specifically expressed in the endosperm and is required for its development; loss of FIS function causes endosperm hyperproliferation and seed abortion. The endosperm nourishes the embryo, similar to the physiological function of the placenta in mammals. We established the endosperm H3K27me3 profile and identified specific target genes of the FIS complex with functional roles in endosperm cellularization and chromatin architecture, implicating that distinct PRC2 complexes have a subset of specific target genes. Importantly, our study revealed that selected transposable elements and protein coding genes are specifically targeted by the FIS PcG complex in the endosperm, whereas these elements and genes are densely marked by DNA methylation in vegetative tissues, suggesting that DNA methylation prevents targeting by PcG proteins in vegetative tissues.

Introduction

Polycomb group (PcG) proteins are evolutionary conserved master regulators of cell identity and balance the decision between cell proliferation and cell differentiation [1]. PcG proteins act in multimeric complexes that repress transcription of target genes; the best characterized complexes are the evolutionary conserved Polycomb Repressive Complex 2 (PRC2) that catalyzes the trimethylation of histone H3 on lysine 27 (H3K27me3), and PRC1, which binds to this mark and catalyzes ubiquitination of histone H2A at lysine 119 [1]. Plants contain multiple genes encoding homologs of PRC2 subunits that have different roles during vegetative and reproductive plant development [2]. Whereas the EMBRYONIC FLOWER (EMF) and VERNALIZATION (VRN) complexes control vegetative plant development, reproductive development in Arabidopsis crucially depends on the presence of the FERTILIZATION INDEPENDENT SEED (FIS) PcG complex that is comprised of the subunits MEDEA (MEA), FERTILIZATION INDEPENDENT SEED2 (FIS2), FERTILIZATION INDEPENDENT ENDOSPERM (FIE) and MSI1 [2]. The FIS PcG complex is required to suppress autonomous endosperm development; loss of FIS function initiates the fertilization-independent formation of seed-like structures containing diploid endosperm [3]. In most angiosperms the endosperm is

a triploid zygotic tissue that develops after fusion of the homodiploid central cell with a haploid sperm cell. The endosperm regulates nutrient transfer to the developing embryo and regular endosperm development is essential for embryo development [4]. Loss of FIS function also dramatically impacts on endosperm development after fertilization, causing endosperm overproliferation and cellularization failure, eventually leading to seed abortion [5]. Thus far, only few direct target genes of the FIS PcG complex are known, among them the MADS-box transcription factor PHERES1 (PHE1) [6], FUSCA3 [7] and MEA itself [8–10]. All three genes are also targets of vegetatively active PcG complexes [7,11], suggesting that different PcG complexes share at least a subset of target genes [7].

Similar to extraembryonic tissues in mammals [12], the endosperm has reduced levels of DNA methylation compared to the embryo or vegetative tissues [13,14]. Hypomethylation is established by transcriptional repression of the maintenance DNA-methyltransferase MET1 during female gametogenesis [15], together with active DNA demethylation by the DNA glycosylase DEMETER (DME) [13,16]. Whereas the global DNA methylation levels differ only slightly between embryo and endosperm (~6% for CG methylation), methylation differences at transposable elements and repeat sequences are significantly more pronounced [13,14]. The functional significance of this genome-wide demethylation of the endosperm is

Author Summary

Cell identity is established by the evolutionary conserved Polycomb group (PcG) proteins that repress transcriptional programs which are not required at particular developmental stages. The plant FERTILIZATION INDEPENDENT SEED (FIS) PcG complex is specifically expressed in the endosperm where it is essential for normal development. The endosperm nourishes the embryo, similar to the physiological function of the placenta in mammals. In this study, we established the cell type–specific epigenome profile of PcG activity in the endosperm. The endosperm has reduced levels of DNA methylation, and based on our data we propose that PcG proteins are specifically targeted to hypomethylated sequences in the endosperm. Among these endosperm-specific PcG targets are genes with functional roles in endosperm cellularization and chromatin architecture, implicating a fundamental role of PcG proteins in regulating endosperm development. Importantly, we identified transposable elements and genes among the specific PcG targets in the endosperm that are densely marked by DNA methylation in vegetative tissues, suggesting an antagonistic placement of DNA methylation and H3K27me3 at defined sequences.

not yet understood. However, it has been proposed that DNA demethylation might cause transposon activation and generation of small interfering RNAs (siRNA) that might move to egg cell or embryo where siRNA-mediated DNA methylation would lead to increased methylation of parasitic genomic sequences [13]. This notion is supported by the observation of accumulating 24nt siRNAs in the female gametophyte and in the endosperm [17]. However, functional loss of RNA polymerase IV, the enzyme responsible for the biogenesis of siRNAs, does not cause reactivation of most transposons [18], suggesting the presence of redundant pathways to silence transposable elements.

In this study, we profiled the H3K27me3 pattern in the endosperm and identified many target genes that were known previously to be targeted by vegetatively active PcG complexes, supporting the idea that different PcG complexes share a common set of target genes. However, we also identified endosperm-specific H3K27me3 target genes that have functional roles in endosperm cellularization and chromatin architecture, suggesting that the FIS PcG complex has endosperm-specific functions and that PcG targeting in plants has tissue specific roles. Finally and most importantly, we discovered that the FIS PcG complex in the endosperm targets transposable elements (TEs) that are protected by DNA methylation in vegetative tissues, implicating that DNA methylation and H3K27me3 are alternative repressive marks that may compensate for each other in the repression of a subset of TEs.

Results

Isolation of Endosperm Nuclei by Fluorescent Activated Cell Sorting

We established a transgenic line expressing PHE1 fused to the enhanced green fluorescent protein (EGFP) under control of the native promoter and 3′ regulatory elements. Strong EGFP fluorescence was exclusively detected in endosperm nuclei from 1 day after pollination (DAP) until 4 DAP, whereas only a weak signal was detectable in the chalazal endosperm at 5 DAP (Figure 1A). EGFP-labeled nuclei from 1–4 DAP-old seeds were isolated with the use of a fluorescence-activated cell sorter. High-throughput techniques allowed the harvesting, nuclei isolation, and sorting of

approximately 100 000 nuclei in about 4 hours. Within this time period, endosperm nuclei did apparently not undergo substantial changes in their transcriptional identity, as judged by a relatively low expression of embryo and seed coat marker genes in relation to the *PHE1* gene (Figure 1B). Expression of seed coat and embryo marker genes followed a similar trend in microdissected endosperm samples (Figure 1C). To identify endosperm-specific PcG target genes we performed chromatin immunoprecipitation (ChIP) of chromatin from sorted endosperm nuclei using H3K27me3 specific antibodies followed by hybridization to high resolution whole-genome tiling microarrays (Chip-on-chip). As a control, we performed ChIP with unspecific IgG antibodies. Genomic regions marked by H3K27me3 ("H3K27me3 regions") were identified as continuous runs of probes with a MAT-score of at least 3.5 (see Materials and Methods). We identified 2282 regions that were significantly enriched for H3K27me3, covering ~1.9 Mb and representing ~1.6% of the sequenced genome. This corresponds to about one fourth the number of H3K27me3 regions identified in seedling tissues [11,19], indicating that there are substantially fewer H3K27me3 targets in the endosperm than in vegetative tissues. Similar to the H3K27me3 distribution in Arabidopsis seedlings [11], most H3K27me3 regions in the endosperm were located on euchromatic chromosome arms and only 17 of the 2282 regions (0.7%) were from centromeric or pericentromeric heterochromatin (Figure 2A). The distribution of H3K27me3 in endosperm over genes had a pronounced maximum in the transcribed region, similar to the distribution of H3K27me3 in vegetative tissues (Figure 2B, [11]). Notably, there was a small but distinct drop of H3K27me3 at the transcriptional start and shortly after the transcriptional stop, possibly caused by localized nucleosome depletion. This interpretation would be in agreement with previous observations made in yeast and human cells, revealing nucleosome depletion at the transcriptional start and around polyadenylation sites [20–22]. The length of H3K27me3 regions in the endosperm was comparable to the length of H3K27me3 regions in vegetative tissues [11], with a median region size of about 750 bps (Figure 2C). *MEA*, *PHE1*, *MEIDOS (MEO)* and *FUSCA3 (FUS3)* as well as other genes that were previously identified as sporophytic H3K27me3 targets were among the endosperm H3K27me3 targets (Figure 2D and Figure 3A), indicating that our procedure successfully identified H3K27me3 targets in the endosperm.

Transposable Elements Are Specifically Targeted by H3K27me3 in the Endosperm

We identified 1773 genes to be associated with H3K27me3; of those, 1533 genes (~86.5%) overlapped with H3K27me3 marked loci identified in seedling tissues ("shared H3K27me3 targets") [11,19], whereas 240 loci (~13.5%) were specifically enriched only in the endosperm ("endosperm-specific H3K27me3 targets") (Figure 3A and Table S1). Most H3K27me3 targets in both sample sets are protein-coding genes of known or unknown functions, similar to the H3K27me3 targets in seedling tissues [11,19] (Figure 3B). The overall distribution of H3K27me3 marked pseudogenes and TEs in the endosperm and seedling tissues was similar; TEs and transposable element genes (TEGs; correspond to genes encoded within a transposable element) were clearly underrepresented among H3K27me3 targets compared to the genome average (Figure 3B). However, the frequency of TEs and TEGs was much higher among the endosperm-specific H3K27me3 targets than among the shared H3K27me3 targets, indicating that a subset of TEs and TEGs are specifically marked by H3K27me3 in the endosperm (Figure 3B). While 16% of all TEs and 46% of all TEGs probed by the microarray are located in centromeric and pericentromeric heterochromatin, only 5% of the TEs with
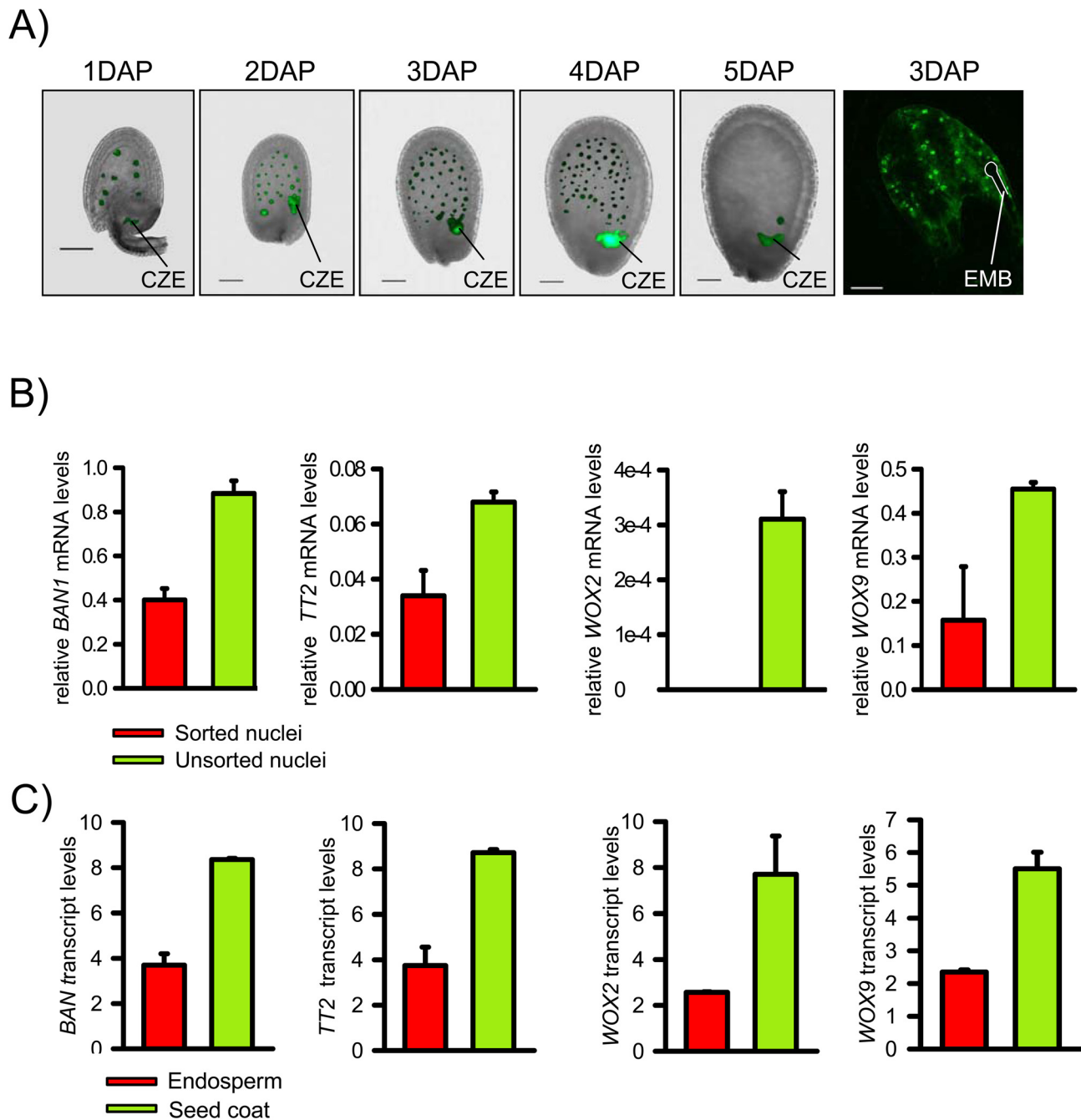
Figure 1. Isolation of EGFP Positive Endosperm Nuclei and Validation of the Technique. A) *PHE1::PHE1-EGFP* is specifically expressed in endosperm nuclei from 1 DAP to 4 DAP. First five images are fluorescence images overlaid with bright-field images. The chalazal endosperm (CZE) is indicated. Last image of the row shows a confocal image. The position of the embryo (EMB) is indicated by a white line. Scale bars, 50 μM. B) Quantitative RT-PCR expression analysis of seed coat marker genes *BANYULS* (*BAN*) and *TRANSPARENT TESTA2* (*TT2*), and embryo marker genes *WUSCHEL-RELATED HOMEOBOX* (*WOX*) *2* and *WOX9* in sorted endosperm nuclei and total nuclei isolated from *PHE1::PHE1-EGFP* expressing 1–4 DAP-old seeds. Values are shown relative to *PHE1* expression. Error bars, s.e.m. C) Transcript levels of seed coat marker genes *BAN* and *TT2* and embryo marker genes *WOX2* and *WOX9* in peripheral endosperm and seed coats of seeds from 1 DAP to 3 DAP, corresponding to seeds containing preglobular to globular stage embryos. Values are based on ATH1 microarray signals after RMA normalization. Error bars, s.e.m.
doi:10.1371/journal.pgen.1001152.g001

H3K27me3 and 16% of the TEGs with H3K27me3 were from these heterochromatic regions. Frequencies of almost all super families of TEs were similar among H3K27me3-marked endosperm-specific TEs and among all TEs detectable by the microarray (Figure S1). Among the shared H3K27me3 targets LTR/COPIA ($p<5E-4$), LINE/L1 ($p<0.05$), and RathE1 elements ($p<0.05$) were significantly enriched, indicating non-random targeting of TEs by PcG proteins. We verified the specificity of our analysis by qPCR validation of endosperm-specific and shared H3K27me3 targets using independently prepared ChIP samples. We randomly selected 10 endosperm-specific TEGs, 9 endosperm-specific genes and 8 shared target genes and could confirm all loci in an independent
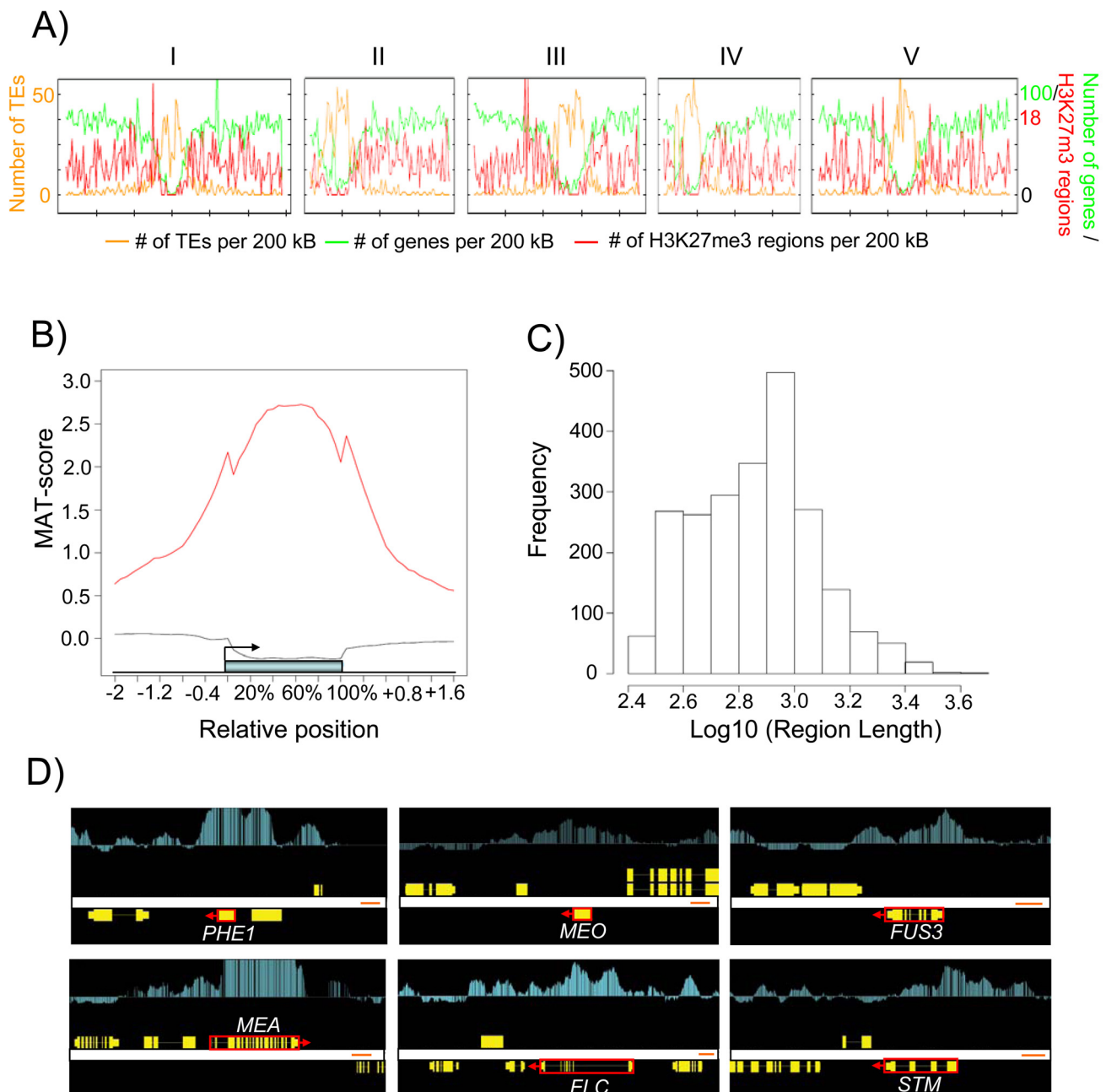
Figure 2. Genome-Wide Identification of H3K27me3 Regions in the Endosperm. A) Chromosomal distribution of H3K27me3 regions. The H3K27me3 regions per 200 kb and genes per 200 kb (y-axis, right-side scale) and number of transposons (y-axis, left-side scale). Numbers on top indicate chromosome number. B) Average H3K27me3 profiles (red line) over H3K27me3 targets. The black line represents the H3K27me3 profile over genes not marked by H3K27me3. The blue bar represents the annotated gene body from transcription start (left) to transcription end (right). Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. C) Length distribution of H3K27me3 regions. D) Comparison of ChIP-chip results with Arabidopsis genes (red boxes, where arrows indicate direction of transcription) that were previously shown to be H3K27me3 targets [6,7,11,57]. Genes are shown as yellow boxes (exons) and lines (introns), and H3K27me3 is shown as vertical light blue bars [MAT score ranging from −1 (bottom) to 6 (top)].
doi:10.1371/journal.pgen.1001152.g002

ChIP experiment (Figure S2), indicating that our procedure was specific with a low false discovery rate.

Functional Roles and Expression of H3K27me3 Target Genes in the Endosperm

Shared H3K27me3 targets in the endosperm were highly enriched for genes involved in transcriptional regulation, with MADS-box transcription factors being a prominently enriched subclass of transcription factors (p = 3.01E-05; Table S2). However, many other GO categories were enriched among shared H3K27m3 target genes, including regulation of metabolism, flower development, cell wall organization, secondary metabolism and others (Table S3). This indicates that the FIS PcG complex acts to repress a large set of genes that are not required during early endosperm development. Among endosperm-specific H3K27me3 targets, there were many genes with potential roles
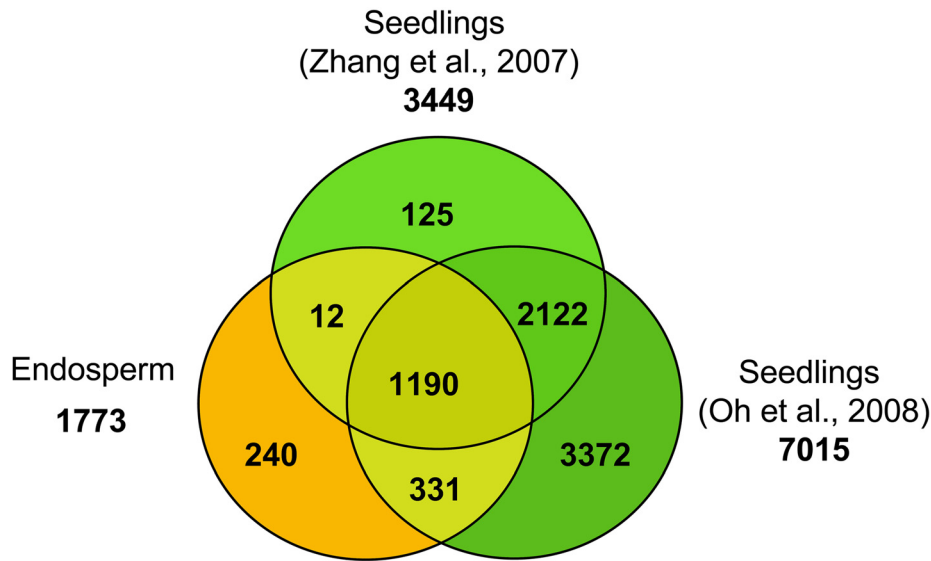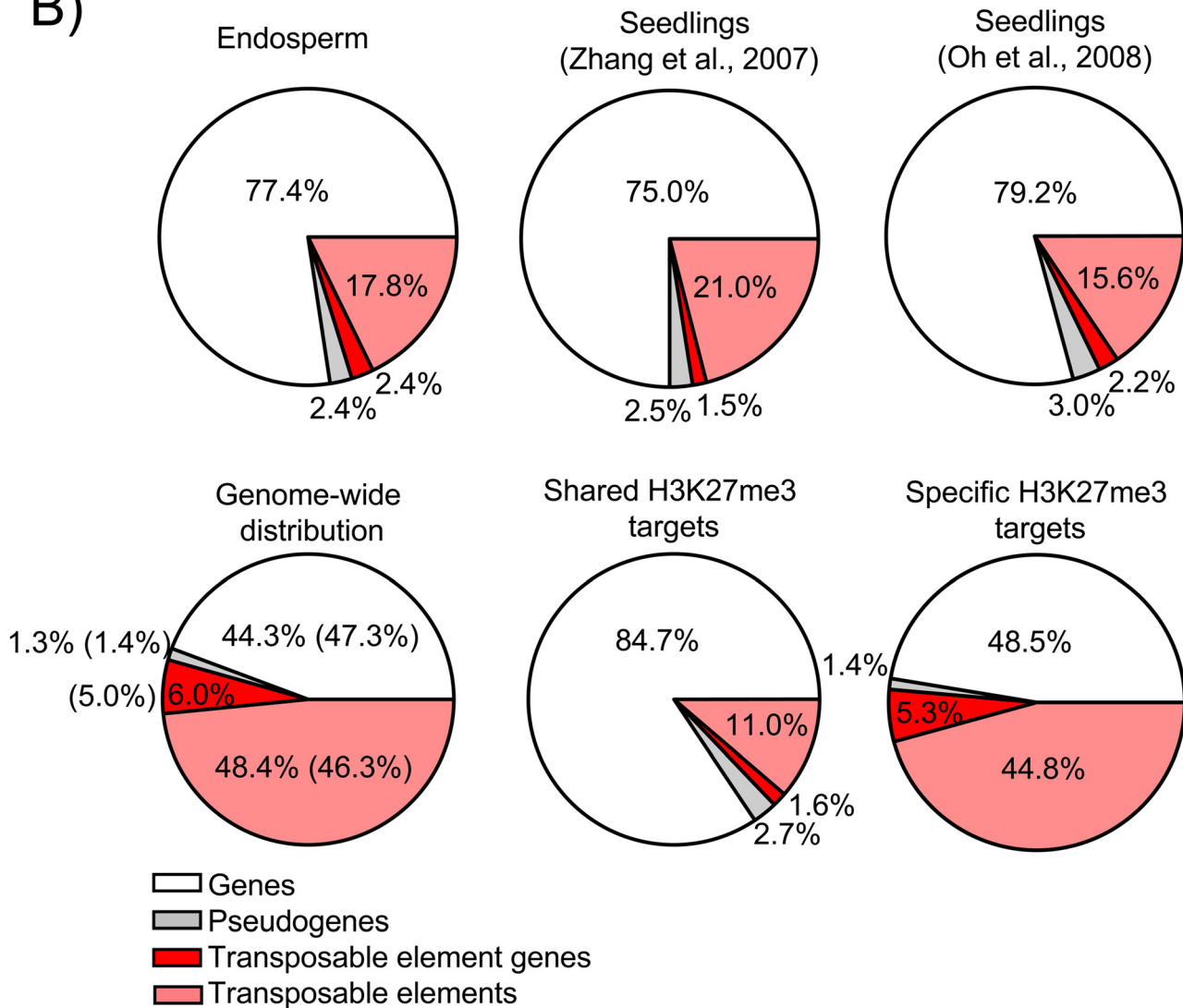
Figure 3. Characteristics of H3K27me3 Target Genes in the Endosperm. A) Venn diagram showing overlap of H3K27me3 target genes in seedlings [11,19] and endosperm. B) Distribution of different types of H3K27me3 targets (genes, pseudogenes, transposable elements, transposable

element genes) in endosperm and seedling tissues. Upper panel shows distribution in the endosperm (this study) and seedlings [11,19], lower panel shows genome-wide distribution of genes, pseudogenes, transposable elements and transposable element genes in comparison to the distribution of shared and specific endosperm H3K27me3 targets. Number in parenthesis reflect number of detectable targets.
doi:10.1371/journal.pgen.1001152.g003

in vesicle-mediated transport and cytoskeleton organization (Table S4), suggesting a specific function of the FIS PcG complex in endosperm cellularization. Furthermore, many genes with functional roles in chromatin organization, such as the PcG protein encoding genes *EMF2, VRN2, MSI1*, the DNA glycosylase *ROS1* as well as DNA helicases were among specific H3K27me3 target genes (Table S4), implicating a role of the FIS PcG complex in establishing specific chromatin architectures in the endosperm.

Next, we analyzed the relation between H3K27me3 modification and gene expression. Gene expression data were derived from the peripheral endosperm of seeds containing globular stage embryos, corresponding to the main fraction of the sorted endosperm nuclei used in our ChIP-chip experiment. Consistent with the function of H3K27me3 in transcriptional silencing, the majority of shared endosperm H3K27me3 target genes were expressed at low levels (Figure 4A). In contrast, a fraction of the endosperm-specific H3K27me3 targets was moderately expressed (Figure 4A). Endosperm-specific H3K27me3 target genes had lower average H3K27me3 scores compared to shared targets independent of their expression level (Figure 4B), suggesting that there is different efficiency of PcG protein targeting or PRC2 activity for endosperm-specific versus shared endosperm H3K27me3 targets.

Using publicly available datasets we tested the tissue-specific expression of endosperm-specific H3K27me3 target genes by cluster analysis. Consistent with the idea that the FIS PcG complex is required for repression of target genes in the endosperm, genes present in clusters I, II and V (45%, n = 75) were specifically repressed in the endosperm (Figure 4C). However, about half of all endosperm-specific H3K27me3 targets were expressed in the endosperm (clusters III and IV, 55%, n = 91; Figure 4C), in agreement with the higher average expression levels of endosperm-specific H3K27me3 target genes compared to non-H3K27me3 target genes (Figure 4A). We consider three not mutually exclusive explanations for this observation: (i) H3K27me3 is not necessarily connected with gene silencing in the endosperm. (ii) For a subset of genes only one of the alleles is marked by H3K27me3. In this case expression of the non-marked allele would be detected, whereas the H3K27me3 allele remains silenced, as it was shown before for *PHE1* and *MEA* [8,9,23,24]. However, imprinted genes predicted by Gehring and colleagues [14] were not among genes present in clusters III and IV. (iii) PcG target genes are differentially regulated in the different domains of the endosperm, i.e. the micropylar, peripheral and chalazal domains).

DNA Methylated Loci Become Targets of H3K27me3 in the Endosperm

TEs were strongly overrepresented among the endosperm-specific H3K27me3 targets compared to the shared H3K27me3 targets (Figure 3B). Hence, we hypothesized that the global DNA demethylation in the endosperm [13,14] caused H3K27me3 to accumulate in regions that are DNA methylated in vegetative tissues and, therefore, H3K27me3-poor. This hypothesis predicts that TEs marked by H3K27me3 in the endosperm have reduced endosperm DNA methylation levels compared to all TEs. Indeed, median endosperm CG and CHG DNA methylation levels were lower at H3K27me3 marked TEs than at other TEs (Figure 5A). CHH methylation levels were generally low and did not differ between H3K27me3 marked TEs and all TEs (data not shown). TEs that carried H3K27me3 in endosperm and vegetative tissues were almost

devoid of CG DNA methylation in endosperm and vegetative tissues. In contrast, TEs that carried H3K27me3 only in the endosperm had high DNA methylation levels in vegetative tissues while DNA methylation levels in the endosperm were markedly below the average over all TEs. Similarly, shared TEGs were almost devoid of DNA methylation in vegetative tissues and in the endosperm. Endosperm DNA methylation levels of specific H3K27me3 TEGs were comparable to the average DNA methylation levels in the endosperm of all TEGs present in the genome (Figure 5B), indicating that reduced DNA methylation levels in the endosperm might allow targeting of PcG proteins to defined sequences independent of residual DNA methylation. CHG methylation followed a similar trend as CG methylation (Figure 5B). In contrast, no substantial changes in CHH methylation levels were observed (data not shown). Protein coding genes were generally much less DNA methylated than TEs or TEGs. Similar to shared TEs and TEGs, shared H3K27me3 target genes were almost devoid of DNA methylation in vegetative tissues and the endosperm (Figure 5C). In marked contrast, endosperm-specific H3K27me3 target genes had significantly higher CG DNA methylation levels in vegetative tissues than the genome-wide average (Figure 5C), supporting the idea that CG DNA methylation prevents these genes being targeted by PcG proteins in vegetative tissues. CG DNA methylation level of endosperm-specific H3K27me3 genes was reduced in the endosperm compared to vegetative tissues, again suggesting that reduced DNA methylation levels in the endosperm enable targeting of PcG proteins to selected loci. Shared and specific protein coding H3K27me3 target genes were almost devoid of CHG and CHH methylation in vegetative tissues and the endosperm (Figure 5C and data not shown). Together, we conclude that DNA methylation and H3K27me3, which both can bring about transcriptional repression of target genes, usually exclude each other at target chromatin. In the endosperm, where DNA methylation is naturally reduced, some loci that were DNA methylated in other tissues become targeted by the FIS PcG complex and marked by H3K27me3. This hypothesis predicts that experimental reduction of DNA methylation levels in vegetative tissues will cause PcG proteins to be targeted to some loci that are usually DNA methylated. Indeed, in *met1* mutants H3K27me3 was found at some TEs that did not carry H3K27me3 in wild type [25], strongly supporting this idea.

Based on their expression in the endosperm, two main clusters of protein coding genes and TEGs that were DNA methylated in vegetative tissues and carried H3K27me3 in the endosperm were apparent (Figure 5D); the first cluster contained genes and TEGs that were weakly expressed in other tissues and became specifically repressed in the endosperm, whereas the second cluster contained genes and TEGs that were mainly repressed in other tissues and became specifically expressed in the endosperm, indicating that loss of DNA methylation fostered expression of several genes and transposons in the endosperm independent of their gain of H3K27me3.

Only Few H3K27me3 Target Genes Are Deregulated in *fis2* Mutants

We wondered whether loss of FIS activity would cause a global deregulation of H3K27me3 target genes. Therefore, we profiled the *fis2* transcriptome of seeds harvested at 3 DAP and 6 DAP and searched for deregulated genes that were marked by H3K27me3 in the endosperm. Loss of FIS function profiled at 3 DAP and 6 DAP
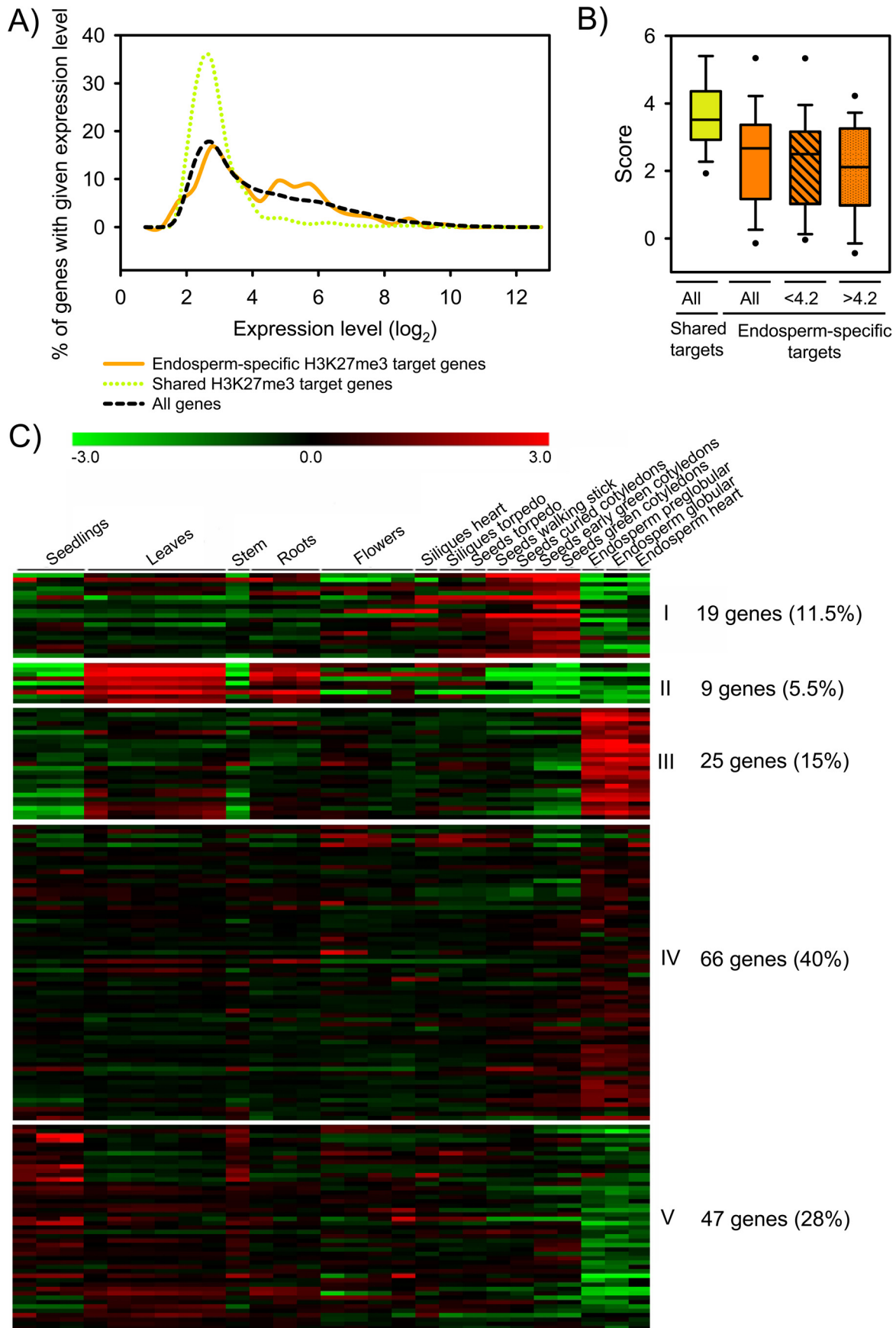
Figure 4. Expression of H3K27me3 Target Genes in the Endosperm. A) Expression level of shared (green) and endosperm-specific (orange) H3K27me3 target genes compared to all genes (black) in the peripheral endosperm of seeds containing globular stage embryos. B) Box plot of H3K27me3 MAT scores of shared (green) and endosperm-specific (orange) target genes. H3K27me3 MAT scores of endosperm-specific target genes with low expression levels ($\log_2 < 4.2$) and moderate expression levels in the endosperm ($\log_2 > 4.2$) are symbolized by striped and dotted fill patterns, respectively. C) Cluster analysis of endosperm-specific H3K27me3 target genes. H3K27me3 target genes are grouped into five mutually exclusive clusters based on their expression patterns. Each row represents a gene, and each column represents a tissue type. Tissue types are: seedlings, leaves, stems, roots, flowers, siliques containing seeds with embryos in the heart or torpedo stage, seeds with embryos in the torpedo, walking stick, curled cotyledon, early green, and green cotyledon stage and endosperm derived from seeds with embryos in the preglobular, globular and heart stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively.
doi:10.1371/journal.pgen.1001152.g004

resulted in different and largely non-overlapping gene expression profiles (Figure 6A). Although the overlap of H3K27me3 target genes and genes deregulated upon loss of FIS function was significant (p = 3.0E-05 and 5.7E-04 for 3 DAP and 6 DAP, respectively), expression of surprisingly few target genes (~1.5% and ~1.8% at 3 DAP and 6 DAP, respectively) was increased upon loss of FIS function (Figure 6A, Table S5). *EMF2* and *VRN2* expression was not increased in *fis2* seeds at 3 or 6 DAP, indicating that loss of FIS2 function is not compensated by increased expression of *FIS2* homologous genes. Genes deregulated at 3 DAP and 6 DAP fell into two largely distinct clusters. Whereas most of early deregulated genes were not expressed in the wild-type endosperm until heart stage, late deregulated genes were predominantly expressed during early wild-type endosperm development and became repressed around heart stage (Figure 6B), supporting the idea that the FIS PcG complex is required for the repression of a defined set of genes around endosperm cellularization [26,27]. Genes deregulated in *fis2* at 3 DAP and 6 DAP were prominently enriched for glycosyl hydrolases (Table S6), with a strong enrichment of Family 17 of plant glycoside hydrolases at 6 DAP. Family 17 members preferentially hydrolyse the major component of endosperm cell walls, callose, [28], suggesting that repression of cell wall degrading enzymes is a requirement for successful endosperm cellularization. Conversely, this implicates that increased expression of these genes in *fis* mutants might contribute to the failure of *fis* mutant endosperm to undergo endosperm cellularization [29].

Importantly, we did not detect increased expression of TEGs in *fis2* mutants, suggesting that loss of H3K27me3 might be compensated by other repressive mechanisms. If so, we wondered whether in seeds lacking both, FIS activity and CG DNA methylation, repression of TEGs would be relieved. Therefore, we generated *fis2/FIS2; met1/MET1* double mutants that contain 12.5% seeds homozygous for *met1* and devoid of FIS activity. We randomly selected eight endosperm-specific H3K27me3 TEGs (At4g16870, At5g37880, At3g32110, At2g13890, At5g35710, At1g35480, At3g28400, At2g16010) that were DNA methylated in vegetative tissues and had decreased DNA methylation levels in the endosperm (Figure S3). Among those, At4g16870, At5g37880 had increased expression levels in *fis2;met1* double mutants compared to *met1* and *fis2* single mutants (Figure 6C), whereas expression of At3g32110 equally increased in *met1* and *fis2; met1* double mutants. Expression of the other TEGs was not significantly changed compared to wild type (data not shown). Based on these data we conclude that DNA methylation and FIS-mediated H3K27me3 can act synergistically to repress a subset of TEGs in the endosperm, but that there are additional mechanisms to silence TEGs in the absence of both mechanisms.

Discussion

Identification of tissue-specific target genes and unraveling how PcG proteins regulate their target genes are important steps to understand how tissue specificity is established. In this study we

established the endosperm-specific H3K27me3 profile and the following main conclusions can be drawn based on our results: (1) The majority of PcG target genes are shared among the endosperm and vegetative tissues, indicating that the reproductively active FIS PcG complex and vegetatively active PcG complexes are recruited to a common set of genes. (2) Expression of only few PcG target genes is induced upon loss of FIS activity, suggesting the activation of alternative repressive mechanisms in the absence of PcG function and/or the lack of appropriate transcriptional activators in the endosperm. (3) Selected TEs, TEGs and protein coding genes are specifically targeted by the FIS PcG complex in the endosperm; these elements and genes are densely marked by DNA methylation in vegetative tissues, suggesting that DNA methylation prevents targeting by PcG proteins in vegetative tissues. (4) DNA demethylation in the endosperm may be required, but not sufficient for targeting of the FIS PcG complex. DNA demethylation in the endosperm is a global phenomenon [13,14], whereas only selected loci become specifically targeted by the FIS PcG complex, suggesting that additional factors are decisive for PcG recruitment.

Functional Roles of H3K27me3 Target Genes in the Endosperm

PcG proteins are largely viewed as general suppressors of genomic programmes that are not required in a specific tissue type or during a particular developmental stage of an organism [1]. This would predict that a large set of PcG target genes is shared in different tissues, as only a small set of genes is expressed in a tissue-specific fashion [30]. In line with this view, we found that the majority of PcG target genes identified in the endosperm are also targeted by PcG proteins in vegetative tissues [11,19], suggesting that different PcG complexes share a common set of target genes during different stages of plant development. However, we identified substantially fewer PcG target genes in the endosperm than previous studies found in seedlings consisting of a mixture of many diverse cell types [11,19] as well as in root hair and non-hair specific cell types [31].

The low number of identified H3K27me3 target genes in endosperm correlates well with reduced expression of the critical PRC2 components *MEA* and *FIS2* in the same tissue [8,27]. A reason for lower expression of PcG proteins and only few PcG protein target genes in endosperm at 1–4 DAP could be that at this time, when mitotic activity is high, the endosperm has not yet acquired its terminal differentiation status [32]. In contrast, the cells profiled in the other studies [11,19,31] were mostly fully differentiated. This is similar to the situation in mammals, where lineage-specific genes often become targeted by PcG proteins only upon cell-fate commitment [33], leading to cell-type specific PcG target profiles and gene expression patterns [34,35]. Furthermore, it should be noted that the endosperm has fundamentally different developmental origin and fate than vegetative tissues; it is derived after fertilization of the diploid central cell and will not contribute any cells to embryo and the developing new plant. Therefore, it is also possible that the reduced number of H3K27me3 target genes
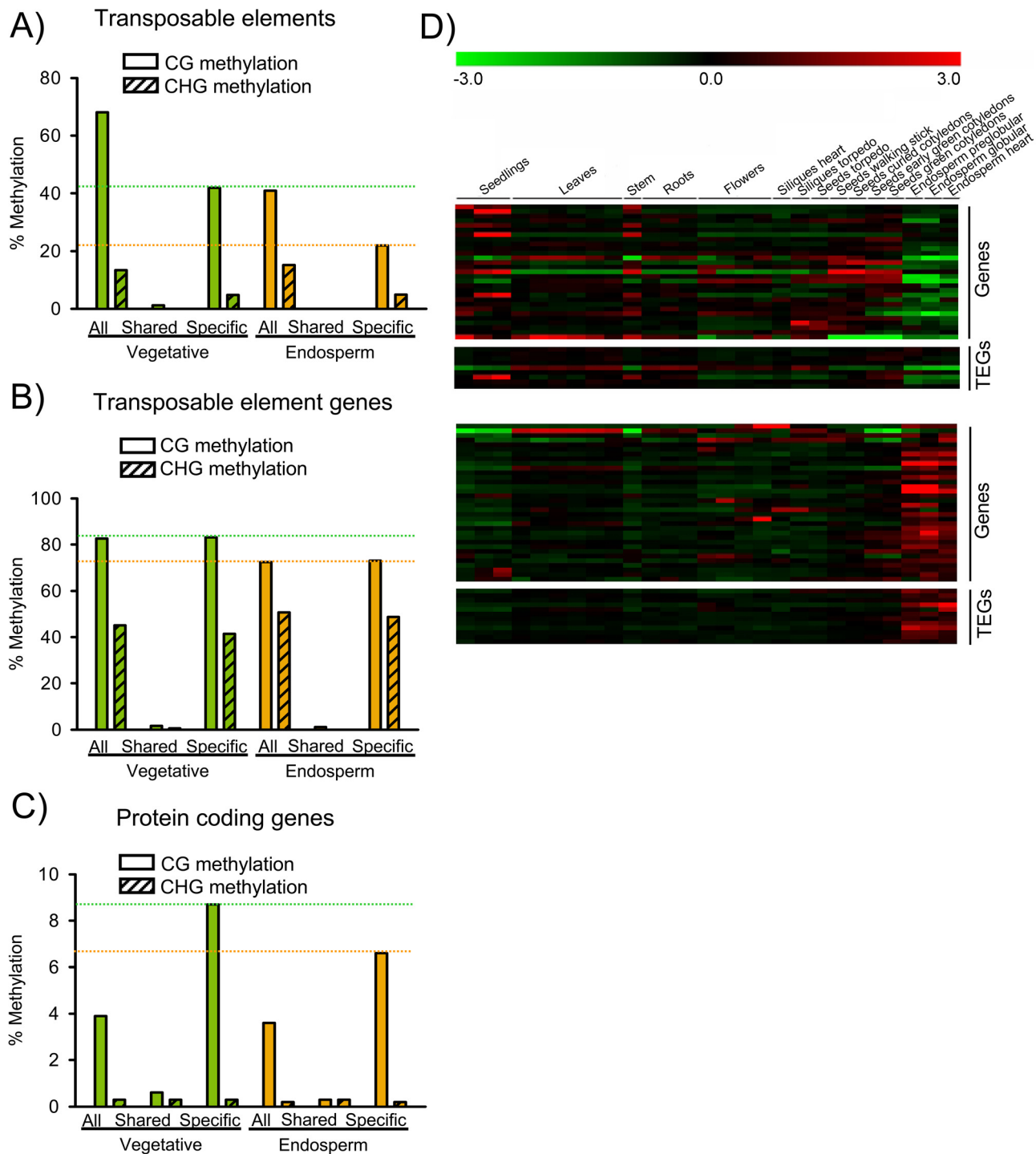
Figure 5. DNA Methylated Genes Become Targets for H3K27me3 in the Endosperm. A) Median DNA methylation levels of TEs in vegetative tissues (green) and endosperm (orange). CG and CHG methylation levels were analyzed for all TEs, TEs marked specifically by H3K27me3 in the endosperm ("Specific") and TEs marked by H3K27me3 in the endosperm and in seedlings [11,19] ("Shared"). Median DNA methylation levels of shared TEs are close to zero. Green and orange dotted horizontal lines mark CG methylation levels of specific H3K27me3 TEs in vegetative and endosperm tissues, respectively. B) Median DNA methylation levels of TEGs in vegetative tissues (green) and endosperm (orange). CG and CHG methylation levels were analyzed for all TEGs, TEGs marked specifically by H3K27me3 in the endosperm ("Specific") and TEGs marked by H3K27me3 in the endosperm and in seedlings [11,19] ("Shared"). Median DNA methylation levels of shared TEGs are close to zero. Green and orange dotted horizontal lines mark CG methylation levels of specific H3K27me3 TEGs in vegetative and endosperm tissues, respectively. C) Median DNA methylation levels of protein coding genes in vegetative tissues (green) and endosperm (orange). CG and CHG methylation levels were analyzed for all protein coding genes, protein coding genes marked specifically by H3K27me3 in the endosperm ("Specific") and protein coding genes marked by H3K27me3 in the endosperm and in seedlings [11,19] ("Shared"). Green and orange dotted horizontal lines mark CG methylation levels of specific H3K27me3 protein coding genes in vegetative and endosperm tissues, respectively. D) Cluster analysis of DNA methylated H3K27me3 target genes and transposons. Genes and transposons are grouped into two mutually exclusive clusters based on their expression patterns in different tissues. Each row represents a gene, and each column represents a tissue type. Tissue types are: seedlings, leaves, stems, roots, flowers, siliques containing

in the endosperm might reflect a less stringent requirement of PcG-mediated gene regulation in the endosperm than in vegetative tissues.

In the endosperm as well as in vegetative tissues, genes encoding for transcription factors were highly enriched among PcG target genes (this study and [11]), supporting the general idea that PcG proteins regulate cell identity by controlling expression of transcription factors [36]. Importantly however, H3K27me3 target genes were also prominently enriched for pectinesterases and glycosyl hydrolases - two enzyme classes that degrade major components of plant cell walls [28,37], indicating an important role of the FIS PcG complex in the regulation of endosperm cellularization. The observed deregulation of both enzyme classes in *fis2* mutant seeds might be the underlying cause of endosperm cellularization failure of *fis* mutants [29].

Only Few H3K27me3 Target Genes Are Deregulated upon Depletion of FIS Activity

Loss of FIS function caused deregulation of only few H3K27me3 genes, similar to observations made in mammalian and *Drosophila* cells, where only a small subset of PcG target genes were deregulated upon depletion of PcG proteins [33,38,39]. Stable repression of FIS target genes could be due to secondary epigenetic modifications that together with FIS-mediated H3K27me3 keep PcG target genes repressed and which are not alleviated in FIS-depleted cells. Alternatively, it is possible that secondary epigenetic modifications are only recruited to FIS target genes upon loss of FIS function. As a third and complementary explanation for the lack of expression of a large number of FIS target genes in FIS-depleted endosperm we propose that the promoters of many PcG target genes lack binding sites for endosperm-specific transcriptional activators required for substantially increased expression in this tissue. This last explanation would imply that those FIS target genes that are deregulated in the *fis2* mutant are even in wild type expressed in the endosperm. Indeed, deregulated FIS target genes were predominantly expressed during wild-type seed development (Figure 6B), supporting the hypothesis that cis-acting tissue-specific enhancers are required for full induction of FIS target genes upon loss of H3K27me3.

Transposable Elements Are Targeted by the FIS PcG Complex in the Endosperm

TEs and TEGs were most prominently enriched among endosperm-specific H3K27me3 targets. This is in contrast to the situation in vegetative tissues, where these elements are largely excluded from PcG target genes [11]. We propose that reduced levels of DNA methylation in the endosperm allow targeting of the FIS PcG complex to defined sequence elements that are protected by DNA methylation in vegetative tissues. This conclusion is supported by the following findings made in this study: (i) Shared H3K27me3 targets were completely devoid of DNA methylation, indicating that DNA methylation prevents targeting by PcG proteins. (ii) Endosperm-specific H3K27me3 protein coding genes had much higher CG DNA methylation levels in vegetative tissues compared to genome-wide average DNA methylation levels, supporting the view that DNA methylation prevents these genes being targeted by PcG proteins in vegetative tissues. (iii) In the endosperm, the average DNA methylation level of endosperm-

specific H3K27me3 targets was reduced compared to vegetative tissues. This trend was most pronounced for TEs, where DNA methylation level of endosperm-specific TEs were much lower compared to the genome-wide average DNA methylation of TEs in the endosperm. However, also TEGs and protein-coding genes had reduced DNA methylation levels in the endosperm compared to vegetative tissues, supporting the notion that reduced DNA methylation levels in the endosperm allow targeting of the FIS PcG complex to defined sequence elements. However, DNA demethylation is a global phenomenon [13,14], but only selected sequences were targeted by the FIS complex, suggesting that DNA demethylation is necessary, but not sufficient for targeting of the FIS complex. The conclusion that DNA methylation and H3K27me3 are usually exclusive epigenetic marks is strongly supported by previous studies on seedlings with experimentally altered DNA methylation. When DNA methylation was reduced, H3K27me3 localized to defined regions within heterochromatin [25], and when DNA methylation was increased H3K27me3 levels dropped [40]. Mutual antagonistic placement of DNA methylation and H3K27me3 was also identified at the imprinted *Rasgrf1* locus in mouse [41], suggesting an evolutionary conserved basis of the underlying mechanism. Together, we conclude that DNA methylation prevents targeting of PcG proteins to sequence elements that have the potential to recruit PcG proteins.

Materials and Methods

### Plant Material and Growth Conditions

A transgenic *Arabidopsis thaliana* (Landsberg *erecta* (L*er*)) line in which endosperm nuclei were specifically marked by EGFP was established by expressing a translational fusion of *PHE1* with EGFP under the transcriptional control of the *PHE1* promoter (*PHE1::PHE1-EGFP*) and 3 kb regulatory 3′ sequences. A transgenic Arabidopsis (Columbia, Col) line constitutively expressing YFP fused to histone H3.2 (*35S::H3.2-YFP*) served as a positive control. The *fis2-1* allele (L*er* accession) has been described previously [3]. The *met1-3* (Col accession) allele was described in [42]. For *met1;fis2* double mutant analysis the newly identified *fis2-5* allele (SALK_009910; Col accession) was used, containing a T-DNA insertion within the first exon. The *fis2-5* seed abortion ratio and mutant seed phenotypes were analyzed and found to be similar to the *fis2-1* allele (data not shown).

Seeds were surface sterilized (5% sodium hypochlorite, 0.1% Tween-20) and plated on MS medium (MS salts, 1% sucrose, pH 5.6, 0.8% bactoagar). Plants were grown in a growth cabinet under long day photoperiods (16 h light and 8 h dark) at 22°C. After 10 days, seedlings were transferred to soil and plants were grown in a growth chamber at 60% humidity and daily cycles of 16 h light at 22°C and 8 h darkness at 18°C. Inflorescences were harvested approximately 21 days after transfer to soil, shock-frozen in liquid nitrogen and stored at −80°C. For analysis of seedlings, seeds were stratified for 2 days at 4°C before incubation in a growth cabinet. After 10 days, whole seedling tissue was harvested, shock-frozen in liquid nitrogen and stored at −80°C before further usage.
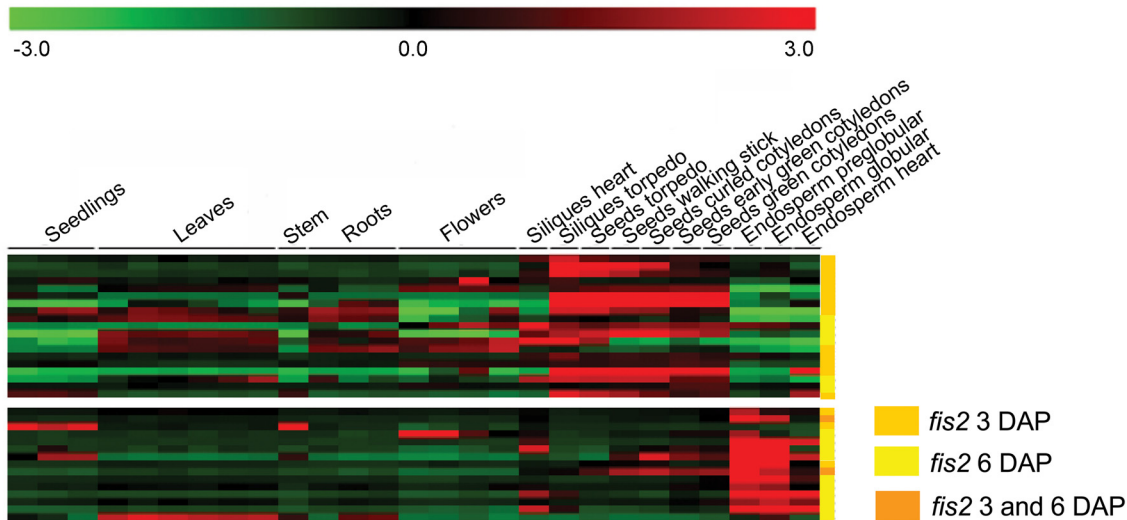
GFP Expression Analysis

Microscopy imaging was performed using a Leica DM 2500 microscope (Leica Microsystems, Wetzlar, Germany) with either
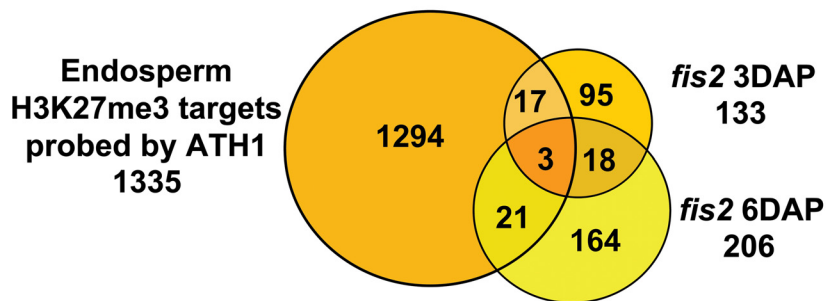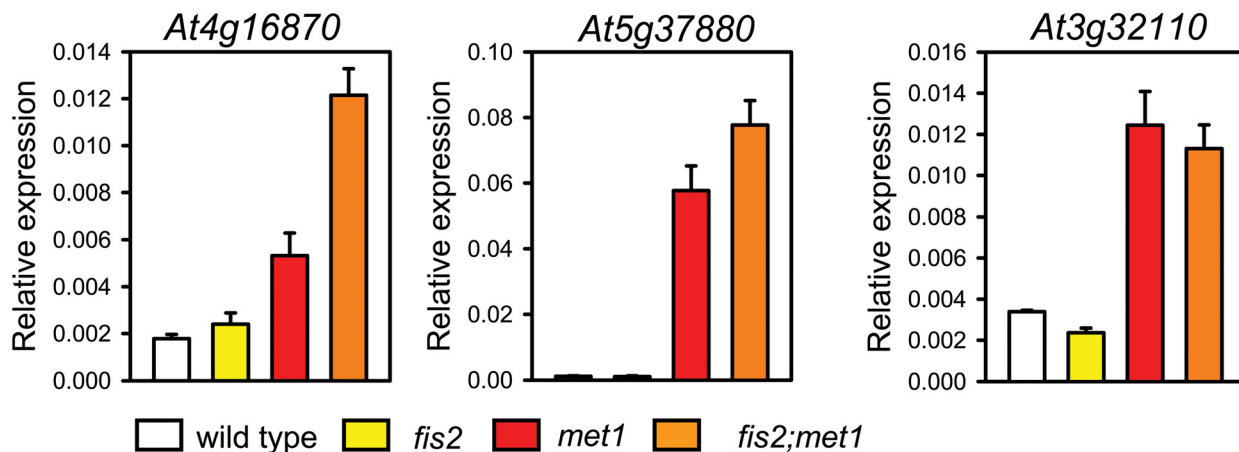
Figure 6. Only Few H3K27me3 Target Genes Are Deregulated in *fis2* Mutants. A) Venn diagram showing overlap of H3K27me3 target genes with genes deregulated in *fis2* seeds at 3 DAP and 6 DAP. Only genes present on the ATH1 microarray were included in the analysis. B) Cluster analysis of H3K27me3 target genes that are deregulated in *fis2* mutant. H3K27me3 target genes are grouped into two main clusters based on their expression patterns in different domains of the endosperm. Each row represents a gene, and each column represents a tissue type. Tissue types are: seedlings, leaves, stems, roots, flowers, siliques containing seeds with embryos in the heart or torpedo stage, seeds with embryos in the torpedo, walking stick, curled cotyledon, early green, and green cotyledon stage and endosperm derived from seeds with embryos in the preglobular, globular and heart stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively. Colors at the right side symbolize genes deregulated in *fis2* at 3 DAP (orange), 6DAP (yellow), or at both time points (yellow with orange stripes). C) Quantitative RT-PCR analysis of TEGs in wild-type, *fis2/FIS2*, *met1/MET1* and *fis2/FIS2;met1/MET1* seeds. Error bars, s.e.m.
doi:10.1371/journal.pgen.1001152.g006

bright-field or epifluorescence optics. Images were captured using a Leica DFC300 FX digital camera, exported using Leica Application Suite Version 2.4.0.R1, and processed using Photoshop 7.0 (Adobe Systems Incorporated, San Jose, USA). Confocal imaging was performed on a Leica SP1-2.

Isolation of GFP Positive Endosperm Nuclei

Nuclei were isolated from 3.5 g of inflorescences following the protocol described in [43]. Isolated nuclei were resuspended in 1× PBS, and proteins were crosslinked to DNA with 1% formaldehyde for 8 min. After adding glycine to 125 mM final concentration and incubation for 5 min, crosslinked nuclei were washed and resuspended in 1× PBS and stained by addition of Propidium Iodide (PI) or DAPI to a final concentration of 1 µg/ml or 0.5 µg/ml, respectively. Biparametric flow analysis of EGFP fluorescence versus nuclear DNA content was performed on a fluorescence activated cell sorter (FACS Aria II, Becton, Dickinson, Franklin Lakes, USA) equipped with a 70 µm flow tip and operated at a sheath pressure of 70 psi. Events were thresholded on forward scatter and samples were sorted at the event rate of 15000/sec. For EGFP and PI excitation a 488 nm laser and for DAPI excitation a 407 laser were used. The barrier filters were 610/20 nm for PI, 450/40 for DAPI and 530/30 for EGFP fluorescence.

The position of the nuclei gate was defined using 6 µm beads (Becton Dickinson), forwards (FSC-A) and sidewards scatter (SSC-A) and was verified by DAPI-staining (Figure S4A). The position of the sort region was established by first determining the baseline of green fluorescence using inflorescence nuclei from EGFP-negative Ler control plants (Figure S4B). The upper and left- and right-hand boundaries of the sort window were adjusted to include all nuclei derived from YFP-positive 35S::H3.2-YFP control plants (Figure S4B). Sorted GFP positive nuclei from PHE1::PHE1-EGFP plants were reanalyzed to verify sorting conditions (Figure S4C).

Transcript Level Analysis

For expression analysis from sorted nuclei, RNA was isolated by flow sorting nuclei directly into 450 µl of RLT lysis buffer (Qiagen, Hilden, Germany) and using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's recommendation. For other expression analyses, siliques were harvested at the indicated time points and RNA extraction and generation of cDNAs were performed using RNeasy Plant Mini Kit (Qiagen) according to the supplier's instructions. For quantitative RT-PCR, RNA was treated with DNaseI and reverse transcribed using the First strand cDNA synthesis kit (Fermentas, Ontario, Canada). Gene-specific primers and Fast-SYBR-mix (Applied Biosystems, Carlsbad, USA) were used on a 7500 Fast Real-Time PCR system (Applied Biosystems). Analysis was performed using three replicates and results were analyzed as described [44]. Briefly, mean expression values and standard errors for the reference gene as well as for the target genes were determined, taking into consideration the primer efficiency that was determined for each primer pair used. Relative expression values were determined by calculating the ratio of target gene expression and reference gene expression and error bars were derived by error propagation calculation. The primers used in this study are specified in Table S7.

Chromatin Immunoprecipitation

ChIP with 500 to 700 ng of chromatin derived from approximately 100'000 sorted nuclei was performed as described [45] using antibodies against H3K27me3 (Millipore, cat. 07-449) and rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, USA, cat. Sc-2027). ChIP-DNA was amplified using the WGA-4 single cell amplification kit (Sigma-Aldrich, St. Louis, USA). For amplifica-

tion of input DNA, 10 ng of chromatin was used. Amplified DNA was purified with the QIAquick PCR purification kit (Qiagen) and eluted with 50 µL of water. DNA concentration was measured using a NanoDrop 1000 (NanoDrop Technologies, Wilmington, USA).

Microarray Analysis

H3K27me3 profiling. Amplified ChIP DNA was fragmented and labelled with the GeneChip WT Terminal Labeling kit (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions. Fragmentation was confirmed using an RNA Nano 1000 kit on a 2100 Bioanalyzer lab-on-chip platform (Agilent, Waldbronn, Germany), revealing an average fragment size of 90 nucleotides. Labelled samples (Input, ChIP with anti-H3K27me3 and ChIP with unspecific IgG) from three independent experiments were hybridized to AGRONOMICS1 arrays (Affymetrix) as previously described [46].

Transcriptional profiling. The transcriptional profile of wild-type and fis2 seeds at 3 DAP was established using ATH1 microarrays (Affymetrix) following previously published procedures [27] with three biological replicates.

Validation of ChIP-chip results. Selected regions were validated using independently prepared chromatin samples immunoprecipitated with H3K27me3 and IgG antibodies. Amplified ChIP-DNA was analyzed by quantitative PCR using 2 µl of 1:30 diluted samples. Three replicates were performed for each sample and results were analyzed as described [44] and presented as percent of input. The primers used in this study are specified in Table S7.

Bioinformatic analysis. All analysis was performed in R 2.9.1 [47]. ChIP-chip data were normalized with MAT [48] implemented in the aroma.affymetrix package [49] with the window-size parameter set to 500. H3K27me3 enrichments were calculated against signals from both input and IgG samples and averaged. Enriched regions were defined as continuous runs of probes with a MAT-score of at least 3.5 and were selected using the package BAC [50] with minRun and maxGap parameters set to 300 and 200, respectively. A gene-specific MAT-score was defined as the 75% ile of all probe-specific MAT-scores for the probes located entirely within the transcribed region of a gene. Visualization of tiling array data was done using the Integrated Genome Browser at http://igb. bioviz.org/download.shtml [51]. Transcript profiling data were normalized with GCRMA [52]; differentially expressed genes were identified with the rankproduct algorithm [53]; false discovery rate = 0.1, fold change >0.6). Clustering analysis was performed using TM4 software [54], Enrichment of GO categories (obtained from TAIR) was tested based on the hypergeometric test and multiple-testing correction according to [55] with a critical p-value of 1.0E-03. Comparisons with whole genome data were based on the sequences probed by the AGRONOMICS1 microarray.

The transcriptional profile of wild-type and fis2 seeds at 6 DAP has been previously published [27]. Reference transcript profiles during development were taken from [30]. DNA methylation profiles were taken from [13,56]. Data for transcript profiles from endosperm were taken from experiments carried out in the laboratories of Bob Goldberg (UCLA), John Harada (UC Davis), Brandon Le (UCLA), Anhthu Bui (UCLA), and Julie Pelletier (UC Davis) and are available under http://estdb.biology.ucla.edu/genechip/project. Microarray raw data generated in this study are available at ArrayExpress, accession numbers E-TABM-1007 and E-TABM-1008.

Supporting Information

Figure S1 Specific Transposon Superfamilies Are Enriched or Depleted among H3K27me3 Targets. Frequency of transposon

superfamilies among endosperm-specific and shared H3K27me3 targets as well as among H3K27me3 targets in seedlings [19] in comparison to the genome-wide transposon frequency that was calculated based on sequences probed by the microarray.
Found at: doi:10.1371/journal.pgen.1001152.s001 (0.01 MB PDF)

Figure S2 Confirmation of Randomly Selected H3K27me3 Target Genes. A) Confirmation of endosperm-specific TEGs. B) Confirmation of endosperm-specific protein coding genes. C) Confirmation of shared H3K27me3 protein coding genes. ChIP was performed using nuclei isolated from endosperm (red bars) or seedlings (green bars) with H3K27me3 specific antibodies and randomly selected target genes were tested by qPCR. Enrichment levels are indicated as % input. Error bars correspond to standard deviation.
Found at: doi:10.1371/journal.pgen.1001152.s002 (0.02 MB PDF)

Figure S3 CG methylation and H3K27me3 Profiles at Selected TEGs. CG methylation profiles of TEGs in vegetative tissues and the endosperm [13,56] were plotted together with the endosperm H3K27me3 profiles obtained in this study.
Found at: doi:10.1371/journal.pgen.1001152.s003 (0.49 MB PDF)

Figure S4 Establishing GFP Sorting Conditions. A) Biparametric flow sort analysis of nuclei isolated from wild-type inflorescences (upper panel), from 35S::H3.2-YFP inflorescences (middle panel) and from PHE1::PHE1-EGFP inflorescences (lower panel). P3 represents the region employed for sorting GFP-negative nuclei. P4 represents the region containing GFP-positive nuclei. B) The presence of nuclei and purity of the defined nuclei gate was verified by analyzing GFP positive nuclei isolated from PHE1:: PHE1-EGFP plants by flow cytometry before (blue line) and after DAPI staining (red line). After addition of DAPI, the whole population of particles present in the defined nuclei gate is shifted to higher DAPI fluorescence, indicating high purity of isolated nuclei. C) The purity of isolated GFP positive nuclei from PHE1::PHE1-EGFP plants was verified by re-analysis of the sorted sample. The sorted sample (green line) was clearly enriched for GFP positive nuclei compared to the unsorted sample (blue line). Bars indicate GFP positive signals. The calculated purity of nuclei was 92%. The presence of two peaks is likely contributed to endoreduplication and correspondingly increased GFP signal intensity.
Found at: doi:10.1371/journal.pgen.1001152.s004 (0.04 MB PDF)

Table S1 Endosperm-specific H3K27me3 targets.
Found at: doi:10.1371/journal.pgen.1001152.s005 (0.15 MB XLS)

Table S2 MADS-box transcription factors among shared H3K27me3 target genes.
Found at: doi:10.1371/journal.pgen.1001152.s006 (0.01 MB PDF)

Table S3 GO analysis of shared endosperm H3K27m3 target genes.
Found at: doi:10.1371/journal.pgen.1001152.s007 (0.01 MB PDF)

Table S4 Endosperm-specific H3K27me3 target genes with specific roles in cellularization and chromatin architecture.
Found at: doi:10.1371/journal.pgen.1001152.s008 (0.01 MB PDF)

Table S5 H3K27me3 target genes deregulated in fis2 seeds at 3 DAP and 6 DAP.
Found at: doi:10.1371/journal.pgen.1001152.s009 (0.01 MB PDF)

Table S6 GO analysis of genes deregulated in *fis2* at 3 DAP and 6 DAP.
Found at: doi:10.1371/journal.pgen.1001152.s010 (0.01 MB PDF)

Table S7 Primers used in this study.
Found at: doi:10.1371/journal.pgen.1001152.s011 (0.01 MB PDF)

Author Contributions

Conceived and designed the experiments: IW EH PR LH CK. Performed the experiments: IW EH PR. Analyzed the data: IW EH PR LH CK. Contributed reagents/materials/analysis tools: CK. Wrote the paper: IW LH CK.

References

1. Schuettengruber B, Cavalli G (2009) Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development 136: 3531–3542.
2. Hennig L, Derkacheva M (2009) Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet 25: 414–423.
3. Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, et al. (1997) Fertilization-independent seed development in Arabidopsis thaliana. Proc Natl Acad Sci USA 94: 4223–4228.
4. Costa LM, Gutierrez-Marcos JF, Dickinson HG (2004) More than a yolk: the short life and complex times of the plant endosperm. Trends Plant Sci 9: 507–514.
5. Köhler C, Makarevich G (2006) Epigenetic mechanisms governing seed development in plants. EMBO Rep 7: 1223–1227.
6. Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, et al. (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes Dev 17: 1540–1553.
7. Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, et al. (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO Rep 7: 947–952.
8. Baroux C, Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of Polycomb group genes: *MEDEA* autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev 20: 1081–1086.
9. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, et al. (2006) DEMETER DNA glycosylase establishes *MEDEA* Polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495–506.
10. Jullien PE, Katz A, Oliva M, Ohad N, Berger F (2006) Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in Arabidopsis. Curr Biol 16: 486–492.
11. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, et al. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol 5: e129. doi:10.1371/journal.pbio.0050129.
12. Santos F, Hendrich B, Reik W, Dean W (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol 241: 172–182.
13. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451–1454.
14. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447–1451.
15. Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, et al. (2008) Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol 6: e194. doi:10.1371/journal.pbio.0060194.
16. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell 110: 33–42.
17. Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, et al. (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. Nature 460: 283–286.
18. Mosher RA, Melnyk CW (2010) siRNAs and DNA methylation: seedy epigenetics. Trends Plant Sci 15: 204–210.

19. Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. PLoS Genet 4: e1000077. doi:10.1371/journal.pgen.1000077.

20. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, et al. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309: 626–630.

21. Ozsolak F, Song JS, Liu XS, Fisher DE (2007) High-throughput mapping of the chromatin structure of human promoters. Nat Biotechnol 25: 244–248.

22. Spies N, Nielsen CB, Padgett RA, Burge CB (2009) Biased chromatin signatures around polyadenylation sites and exons. Mol Cell 36: 245–254.

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

24. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, et al. (2006) DEMETER DNA glycosylase establishes *MEDEA* Polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495–506.

25. Mathieu O, Probst AV, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J 24: 2783–2791.

26. Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. Plant Cell 20: 635–647.

27. Erilova A, Brownfield L, Exner V, Rosa M, Twell D, et al. (2009) Imprinting of the Polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. PLoS Genet 5: e1000663. doi:10.1371/journal.pgen.1000663.

28. Minic Z, Jouanin L (2006) Plant glycoside hydrolases involved in cell wall polysaccharide degradation. Plant Physiol Biochem 44: 435–449.

29. Sorensen MB, Chaudhury AM, Robert H, Bancharel E, Berger F (2001) Polycomb group genes control pattern formation in plant seed. Curr Biol 11: 277–281.

30. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, et al. (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501–506.

31. Deal RB, Henikoff S (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev Cell 18: 1030–1040.

32. Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, et al. (2001) Dynamic analyses of the expression of the histone::YFP fusion protein in Arabidopsis show that syncytial endosperm is divided in mitotic domains. Plant Cell 13: 495–509.

33. Mohn F, Schübeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet 25: 129–136.

34. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 20: 1123–1136.

35. Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, et al. (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res 16: 890–900.

36. Ringrose L, Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development 134: 223–232.

37. Pelloux J, Rusterucci C, Mellerowicz EJ (2007) New insights into pectin methylesterase structure and function. Trends Plant Sci 12: 267–277.

38. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, et al. (2004) Silencing of human Polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev 18: 1592–1605.

39. Schwartz YB, Kahn TG, Stenberg P, Ohno K, Bourgon R, et al. (2010) Alternative epigenetic chromatin states of Polycomb target genes. PLoS Genet 6: e1000805. doi:10.1371/journal.pgen.1000805.

40. Naumann K, Fischer A, Hofmann I, Krauss V, Phalke S, et al. (2005) Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. EMBO J 24: 1418–1429.

41. Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, et al. (2008) Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. PLoS Genet 4: e1000145. doi:10.1371/journal.pgen.1000145.

42. Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet 34: 65–69.

43. Weigel D, Glazebrook J (2002) Arabidopsis-a laboratory manual. New York: Cold Spring Harbor Laboratory Press.

44. Simon P (2003) Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics 19: 1439–1440.

45. Acevedo LG, Iniguez AL, Holster HL, Zhang X, Green R, et al. (2007) Genome-scale ChIP-chip analysis using 10,000 human cells. Biotechniques 43: 791–797.

46. Rehrauer H, Aquino C, Gruissem W, Henz S, Hilson P, et al. (2010) AGRONOMICS1 - A new resource for Arabidopsis transcriptome profiling. Plant Physiol 152: 487–499.

47. R Core Development Team (2009) A language and environment for statistical computing. Vienna: R foundation for statistical computing.

48. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, et al. (2006) Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci USA 103: 12457–12462.

49. Bengtsson H, Simpson K, Bullard J, Hansen K (2008) Aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Berkeley: Department of Statistics.

50. Gottardo R, Li W, Johnson WE, Liu XS (2008) A flexible and powerful bayesian hierarchical model for ChIP-Chip experiments. Biometrics 64: 468–478.

51. Nicol JW, Helt GA, Blanchard SG, Jr., Raja A, Loraine AE (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25: 2730–2731.

52. Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F (2003) A model based background adjustment for oligonucleotide expression arrays. Baltimore: John Hopkins University, Department of Biostatistics.

53. Breitling R, Armengaud P, Amtmann A, Herzyk P (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett 573: 83–92.

54. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.

55. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B 57: 289–300.

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

57. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, et al. (2006) Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. EMBO J 25: 4638–4649.

High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Arabidopsis Endosperm

Philip Wolff[1,2,9], Isabelle Weinhofer[1,9], Jonathan Seguin[1,9], Pawel Roszak[1], Christian Beisel[3], Mark T. A. Donoghue[4], Charles Spillane[4], Magnus Nordborg[5], Marc Rehmsmeier[5*], Claudia Köhler[1,2*]

1 Department of Biology and Zurich-Basel Plant Science Center, Swiss Federal Institute of Technology, Zurich, Switzerland, 2 Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala, Sweden, 3 Department Biosystems Science and Engineering, Swiss Federal Institute of Technology, Basel, Switzerland, 4 Genetics and Biotechnology Lab, Botany and Plant Science, National University of Ireland Galway, Aras de Brun, Ireland, 5 Gregor Mendel Institute of Molecular Plant Biology GmbH, Vienna, Austria

Abstract

Genomic imprinting is an epigenetic phenomenon leading to parent-of-origin specific differential expression of maternally and paternally inherited alleles. In plants, genomic imprinting has mainly been observed in the endosperm, an ephemeral triploid tissue derived after fertilization of the diploid central cell with a haploid sperm cell. In an effort to identify novel imprinted genes in Arabidopsis thaliana, we generated deep sequencing RNA profiles of F1 hybrid seeds derived after reciprocal crosses of Arabidopsis Col-0 and Bur-0 accessions. Using polymorphic sites to quantify allele-specific expression levels, we could identify more than 60 genes with potential parent-of-origin specific expression. By analyzing the distribution of DNA methylation and epigenetic marks established by Polycomb group (PcG) proteins using publicly available datasets, we suggest that for maternally expressed genes (MEGs) repression of the paternally inherited alleles largely depends on DNA methylation or PcG-mediated repression, whereas repression of the maternal alleles of paternally expressed genes (PEGs) predominantly depends on PcG proteins. While maternal alleles of MEGs are also targeted by PcG proteins, such targeting does not cause complete repression. Candidate MEGs and PEGs are enriched for cis-proximal transposons, suggesting that transposons might be a driving force for the evolution of imprinted genes in Arabidopsis. In addition, we find that MEGs and PEGs are significantly faster evolving when compared to other genes in the genome. In contrast to the predominant location of mammalian imprinted genes in clusters, cluster formation was only detected for few MEGs and PEGs, suggesting that clustering is not a major requirement for imprinted gene regulation in Arabidopsis.

Introduction

Genomic imprinting is an epigenetic phenomenon present in mammals and flowering plants that leads to differential expression of alleles of the same gene dependent on the parent-of-origin. Imprinted genes are differentially marked in the gametes, making maternal and paternal alleles functionally different [1]. Whereas in mammals imprinting occurs in the placenta as well as the embryo and tissues of the adult organism, most examples of imprinted genes in plants to date are confined to the endosperm [2]. Although examples of imprinted genes in the plant embryo exist [3], they seem to be rare. The endosperm is a functional analog of the mammalian placenta and serves to support embryo growth [4]. It is a triploid tissue that is derived after fertilization of the homodiploid central cell with a haploid sperm cell, whereas the second sperm cell will fertilize the haploid egg cell, leading to the formation of the diploid embryo [5].

Genome-wide studies of DNA methylation in embryo and endosperm have revealed transposon and repeat sequences to be largely hypomethylated in the endosperm compared to the embryo [6,7], with virtually all CG sequences methylated in the embryo having reduced methylation levels in the endosperm [7]. Methylation levels at CG sites are partially restored in the endosperm of mutants where the DNA glycosylase DEMETER (DME) is disrupted [7], implicating DME to be largely responsible for genome-wide CG demethylation in the endosperm. Repression of the DNA methyltransferase MET1 in the central cell is also considered to contribute to the establishment of differential DNA methylation in the endosperm [8]. Transposon insertions or local sequence duplications are known to recruit DNA methylation and to initiate silencing of neighbouring genes in vegetative tissues [9]. This process is likely to render the targeted genes to be imprinted, as the maternal alleles can escape silencing by DME-mediated DNA demethylation in the endosperm. Based on the idea that DNA demethylation will activate genes in the endosperm that are silenced in vegetative tissues, Gehring and colleagues identified five novel imprinted genes and predicted around 50 imprinted

Author Summary

Genomic imprinting poses a violation to the Mendelian rules of inheritance, which state functional equality of maternally and paternally inherited alleles. Imprinted genes are expressed dependent on their parent-of-origin, implicating an epigenetic asymmetry of maternal and paternal alleles. Genomic imprinting occurs in mammals and flowering plants. In both groups of organisms, nourishing of the progeny depends on ephemeral tissues, the placenta and the endosperm, respectively. In plants, genomic imprinting predominantly occurs in the endosperm, which is derived after fertilization of the diploid central cell with a haploid sperm cell. In this study we identify more than 60 potentially imprinted genes and show that there are different epigenetic mechanisms causing maternal and paternal-specific gene expression. We show that maternally expressed genes are regulated by DNA methylation or Polycomb group (PcG)-mediated repression, while paternally expressed genes are predominantly regulated by PcG proteins. From an evolutionary perspective, we also show that imprinted genes are associated with transposons and are more rapidly evolving than other genes in the genome. Many MEGs and PEGs encode for transcriptional regulators, implicating important functional roles of imprinted genes for endosperm and seed development.

genes in Arabidopsis, with many such genes encoding transcription factors and proteins with chromatin-related functions [6].

Although DNA methylation is widely recognized as a major mechanism for imprinted gene regulation, there are several examples suggesting that DNA methylation is not in all cases sufficient to establish imprinted gene expression. For instance, silencing of the maternal alleles of *PHERES1* (*PHE1*) and the paternal alleles of *MEDEA* (*MEA*) and *ARABIDOPSIS FORMIN HOMOLOGUE 5* (*AtFH5*) depend on repressive activity of the FERTILIZATION INDEPENDENT SEED (FIS) Polycomb group (PcG) complex [10–14]. The FIS PcG complex is a chromatin modifying complex that by trimethylating its target genes on histone H3 at lysine 27 (H3K27me3) causes gene repression [15]. MEA itself is a subunit of the FIS PcG complex and autoregulates its expression by repressing the paternal *MEA* allele [11–13], whereas activity of the maternal *MEA* allele requires DME-mediated DNA demethylation [16,17]. Similarly, imprinted expression of *PHE1* depends on both, the FIS PcG complex and DME-mediated DNA demethylation [6,18,19]. Demethylation of a helitron remnant located 2.5 kbps downstream of the *PHE1* locus as well as binding of the FIS PcG complex to the *PHE1* promoter region are required for silencing of the maternal *PHE1* alleles, suggesting long-range interactions between the repeat region and PcG proteins [19].

As demethylation of repeat elements and transposons in the endosperm is a major mechanism giving rise to imprinted gene expression, it has been proposed that imprinting arose as a by-product of a silencing mechanism targeting invading foreign DNA [6,7,20]. Another view on the evolution of imprinted genes states that imprinting arose as a consequence of a conflict over the distribution of resources from the mother to the offspring [21,22]. This theory predicts that there will be a selection for paternally active genes that can maximize the transfer of nutrients to the developing embryo, whereas the mother protects herself against the demands of the embryo by suppressing the growth induced by the paternally active genes. In line with this theory, imprinting occurs in placental mammals and flowering plants, both

contributing maternal resources to the progeny. Furthermore, many imprinted genes in mammals affect both the demand and supply of nutrients across the placenta [23]. From the few known imprinted genes in plants, some do affect endosperm growth [24–26] and there is evidence that some imprinted genes may be fast evolving [27–29]. This suggests that imprinted gene expression, although being a likely by-product of a genome defence mechanism, may confer a selective advantage.

The discussion about the origin and evolution of imprinted gene expression in plants has been restricted by the sparse knowledge of imprinted loci. In this study we report on the identification of more than 60 genes in Arabidopsis with predicted parent-of-origin specific expression, greatly extending the number of potential imprinted loci in plants. Our study also revealed that specifically maternally and paternally expressed genes are regulated by different molecular mechanisms that rely on DNA methylation and FIS PcG function, respectively. Finally, we find that imprinted genes in plants are more rapidly evolving when compared to all other genes in the genome, and we propose that transposons may have been a driving force for the evolution of imprinted gene expression in Arabidopsis.

Results

Genome-Wide Identification of Genes with Parentally-Biased Expression

We performed reciprocal crosses of the two *Arabidopsis thaliana* accessions Col-0 and Bur-0 that offer a sufficiently high number of small nucleotide polymorphisms (SNPs) to define the parent-of-origin expression of the majority of genes [30]. Seeds containing globular stage embryos were harvested at 4 days after pollination (DAP). Microscopic analysis of seeds derived from four siliques developing after reciprocal crosses of Col-0 and Bur-0 accessions did not reveal obvious developmental differences (Figure 1A), suggesting that Col-0 and Bur-0 accessions have similar properties when used as maternal plant or pollen donor. We generated mRNA-sequencing libraries of seeds derived from Col-0 × Bur-0 and Bur-0×Col-0 crosses, which we sequenced to 80-fold and 67-fold transcriptome coverage, respectively (see Materials and Methods). We identified 12041 genes (q<0.05) with maternally biased expression (maternally expressed genes, MEGs; Table S1) and 119 genes (q<0.05) with paternally biased expression (paternally expressed genes, PEGs; Table S2; see Materials and Methods for details). Within this dataset we identified seven MEGs and six PEGs that were previously predicted to be regulated by genomic imprinting [6]. We also identified the known imprinted genes *FWA* [31], *MYB3R2* [6], and *At3g25260* [6] in our MEG dataset as well as *PHE1* [10] and *At5g62110* [6] in the PEG dataset. Several known imprinted genes were not identified either due to low numbers of sequence reads (*MEA* and *FIS2*), lack of SNPs between Col-0 and Bur-0 (*MPC* [26]), or lack of significant q values (*ATFH5*, *HDG3*, *HDG8*, and *HDG9* [6]). Several of the previously identified imprinted genes (*HDG3*, *HDG8*, *HDG9*) had significantly deviating read numbers from a predicted 2m:1p ratio only in one direction of the cross but not in the reciprocal cross and therefore failed to pass our significance threshold. This suggests that genomic imprinting can be accession-dependent and many genes may be imprinted in one accession but biallelically expressed in another accession, as was previously described for different alleles of the maize *R* and *dzr1* loci [32–33]. Together, based on SNP distributions deviating from the expected 2m:1p genome ratio we successfully identified six out of twelve previously identified imprinted genes, indicating that using this approach we can successfully identify novel imprinted genes on a genome-wide scale.
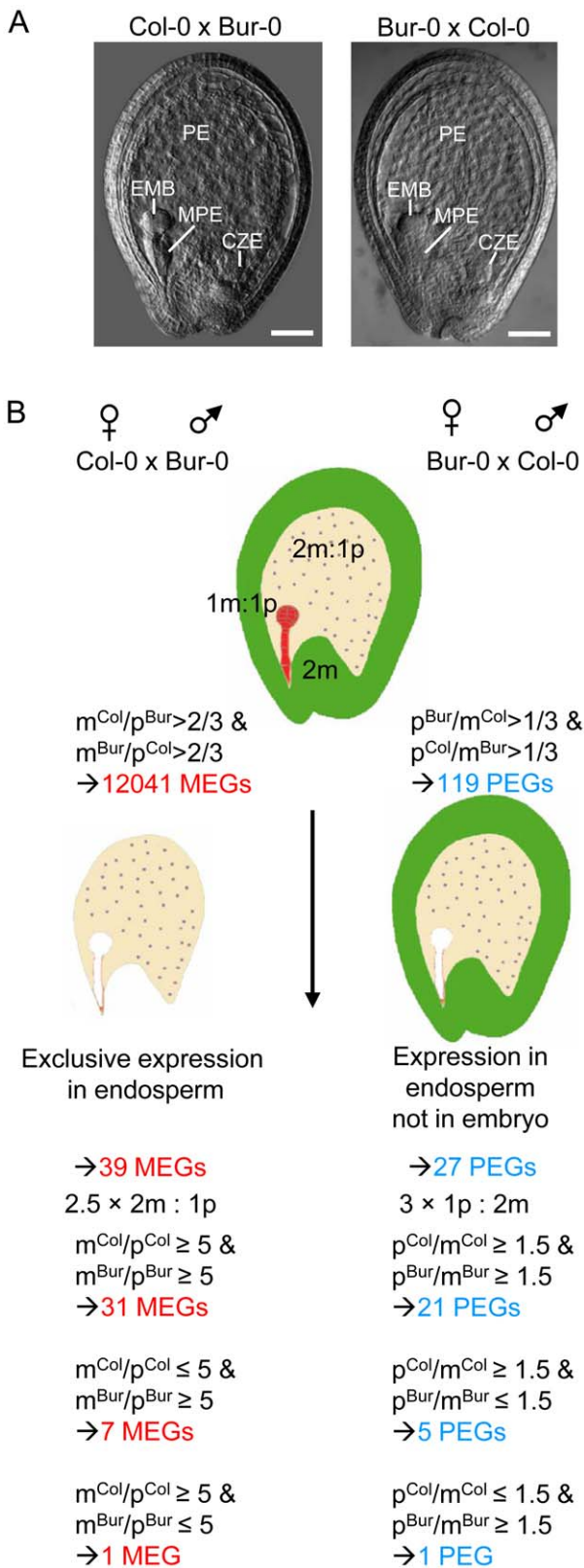
Figure 1. Scheme of Experimental Procedures Leading to the Identification of MEGs and PEGs in Arabidopsis. (A) Seeds derived after reciprocal crosses of Col-0 and Bur-0 accessions are phenotypically indistinguishable. Embyro (EMB), chalazal (CZE), micropylar (MPE) and peripheral (PE) endosperm are indicated. Scale bars, 50 μM. (B) Outline of filtering procedures leading to the identification of MEGs and PEGs. doi:10.1371/journal.pgen.1002126.g001

Identification of Imprinted Genes in the Endosperm

Genes that are regulated by genomic imprinting are expected to be expressed in zygotic tissues, excluding genes that contribute long lived RNAs from gametophytic tissues or that are expressed in the maternally-derived seed coat. As we used RNA isolated from seeds, many of the MEGs were likely to be seed coat-expressed genes and not necessarily imprinted genes. Therefore, we identified a stringent group of 400 genes that were preferentially expressed in the endosperm, selecting for genes with fivefold or greater signal log ratios (SLRs) in one of the endosperm domains compared with the seed coat and which were not significantly expressed in vegetative tissues (Table S3, Figure 1B and Materials and Methods). To avoid a bias towards strongly expressed genes, we included MEGs with low mRNA levels (read counts higher or equal to 10 and smaller or equal to 30) having only threefold or higher SLRs in one of the endosperm domains compared with the seed coat and which were not significantly expressed in vegetative tissues. After this selection we identified 39 candidate MEGs with maternally-biased expression (Table S4), among them *FWA*, *MYB3R2* and three predicted imprinted genes (*At2g19400*, *At3g23060*, *At4g00540* [6]). For simplicity, candidate MEGs and PEGs will be referred to as "MEGs" and "PEGs" throughout this manuscript. In contrast, MEGs and PEGs that passed additional PCR-based tests will be referred to as "confirmed MEGs and PEGs". We considered the possibility that some MEGs are not transcribed in the endosperm but instead contribute long-lived maternally-inherited RNAs. This would predict that the mRNA levels of these genes should be reduced in seeds at 4 DAP compared to flowers containing developed female gametophytes. Out of 39 tested genes only three genes (*At1g54280*, *At3g26590*, *At4g26140*) had lower expression levels in 4 DAP seeds than in stage 12 flowers (containing eight-nucleate/seven-celled female gametophytes [34]), implicating that maternally stored mRNAs do not extensively bias the identification of imprinted genes at 4 DAP. Selection for parentally-biased expression was based on deviating sequence reads from the expected 2m:1p ratio of maternal Col-0 (m^{Col}) to paternal Bur-0 (p^{Bur}) as well as maternal Bur-0 (m^{Bur}) and paternal Col-0 (p^{Col}) alleles. However, when analyzing the ratio of m^{Col} to p^{Col} as well as m^{Bur} to p^{Col} alleles, we noted that seven of the identified MEGs were likely to be imprinted only in the Bur-0 accession, but not in the Col-0 accession and one of the identified MEGs was likely to be imprinted only in the Col-0 accession, but not in the Bur-0 accession (Table S4), considering a fivefold higher expression of both maternal alleles over the paternal allele as maternally-biased expression. This suggests that there is considerable accession-dependency underlying the regulation of imprinted genes in Arabidopsis.

Many PEGs were strongly expressed in pollen but only weakly expressed in the endosperm (Figure S1A, S1B), implicating that transcripts loaded after fertilization from the sperm cells into the seed remained detectable in seeds at 4 DAP, as suggested by previous findings [35]. Therefore, we selected for PEGs that were present in a group of 12190 genes that we identified as being significantly expressed in the endosperm (Table S5). After this selection we obtained 38 genes (Table S6). Genes with predominant expression in the embryo are expected to mimic genes with paternally-biased expression and were excluded from the PEG list, resulting in 27 PEGs (Table S7, Figure 1B), among them previously identified genes *PHE1* [10] and *At5g62110* [6] as well as five predicted imprinted genes (*At4g11400*, *At1g48910*, *At5g50470*, *At3g19160*, *At1g23320* [6]). Considering a threefold higher expression of the paternal allele over maternal alleles as parentally-biased expression (Figure 1B), we identified five PEGs

that were predominantly paternally expressed when inherited from the Col-0 parent, but biallelically expressed when inherited from Bur-0 and one PEG that was predominantly expressed when inherited from the Bur-0 parent but biallelically expressed when inherited from Col-0 (Table S7).

We considered the possibility that some of the MEGs and PEGs are regulated by parental-specific splicing. MEGs and PEGs had on average 6–8 SNPs per gene, making it rather unlikely that for genes with numbers of SNPs this large a single-exon splice variant could lead to the statistically very significant differences in overall read numbers. Nevertheless, we analyzed for every candidate gene its female and male read distributions over all SNPs of that gene over all SNPs of that gene with Pearsons's chi-square test. All MEGs and PEGs had p-values larger than 0.05, indicating that parental-specific splicing is not a major confounding factor in our analysis.

We tested allele-specific expression of selected MEGs and PEGs by restriction-based allele-specific PCR analysis as well as sequencing analysis and found eleven out of twelve MEGs tested to be predominantly expressed from the maternal alleles in reciprocal crosses (At1g52460, At3g23060, At5g03020, At1g60970, At2g19400, At4g29570, At5g46300, At3g10590, At3g21830, At1g51000, AGL36; Figure 2A). One MEG, At1g20730, was specifically maternally expressed in the Col-0×Bur-0 cross, whereas it was biallelically expressed in the Bur-0×Col-0 cross (Figure 2B). Two MEGs with predicted accession-specific imprinting (AGL28 and AGL96) were similarly regulated (Figure 2B), indicating that these three genes are exclusively maternally expressed in Bur-0, but biallelically expressed in Col-0. We also confirmed paternal-preferential expression for eight out of twelve tested PEGs (At4g31900, At1g49290, At5g54350, At3g50720, At1g11810, At5g50470, At3g62230, At3g49770; Figure 2C). Three of the twelve tested PEGs (AGL23, At1g66630, At1g11810) were preferentially paternally expressed in one direction of the cross but biallelically expressed in the reciprocal cross, indicating a significant accession-dependency in the regulation of PEGs (Figure 2D). Biallelic expression of AGL23 in Col-0 is consistent with the previously proposed role of AGL23 for female gametophyte development in the Col-0 accession [36]. However, our data suggest that the functional roles of AGL23 differ between different Arabidopsis accessions. We furthermore confirmed the predicted accession-specific expression of At4g11940, which was paternally expressed in the cross Bur-0×Col-0, but biallelically expressed in the reciprocal cross (Figure 2D). Together, with 19 out of 24 predicted reciprocally imprinted genes being experimentally confirmed and experimental confirmation of three out of three predicted accession-dependent imprinted genes, we conclude that the majority of the newly predicted MEGs and PEGs are indeed regulated by genomic imprinting.

Among reciprocally imprinted and accession-dependent imprinted MEGs and PEGs we detected a significant enrichment of nuclear localized proteins and transcription factors (Table S8), with many of them belonging to AGL MADS-box transcription factors (AGL36, AGL28, AGL96, PHE1, AGL23) that in yeast two-hybrid studies were shown to directly or indirectly interact with each other [37] as well as with AGL62 [38] (Figure S2). AGL62 has been proposed to be a major regulator of endosperm cellularization [38], suggesting a major regulatory role of imprinting genes in timing the onset of endosperm cellularization. Furthermore, MEGs were enriched for genes encoding cytidine deaminases ($p = 3\,\mathrm{E}-4$) that in zebrafish as well as in mammals have been proposed to be required for DNA demethylation [39,40]. Whether cytidine deaminases play a similar role in the plant endosperm remains to be tested.

Dynamics of MEG and PEG Expression during Seed and Vegetative Development

Whereas MEGs were filtered based on the absence of expression in vegetative tissues, these filtering criteria were not applied for PEGs. However, expression patterns of MEGs and PEGs were very similar, both types of imprinted genes had a preference for being expressed in pollen, but were mainly excluded from vegetative tissues (Figure 3A). The expression profile of MEGs and PEGs in seeds was very comparable as well, both types of genes were not expressed in the seed coat, rarely expressed in the embryo, but most MEGs and PEGs were strongly expressed in the chalazal region of the endosperm (Figure 3B). Whereas expression of most MEGs was confined to the chalazal region of the endosperm, PEG expression was less restricted and extended to the peripheral and micropylar regions of the endosperm. MEG and PEG expression was clearly detectable in the endosperm of seeds containing preglobular stage embryos and expression in the chalazal endosperm remained detectable until seeds contained cotyledon stage embryos. Expression in the micropylar and peripheral region of the endosperm was only detectable until seeds contained heart stage embryos, after this stage expression remained confined to the chalazal endosperm. Average expression levels of MEGs and PEGs in the chalazal endosperm region were clearly above average expression levels of all genes, with expression being highest at the preglobular stage and declining towards the heart stage of seed development (Figure 3C).

Different Localization and Regulatory Impact of DNA Methylation at MEG and PEG Loci

Previous studies predicted imprinted genes based on the assumption that DNA demethylation in the central cell causes activation of maternal alleles of genes, whereas paternal alleles remain methylated and silenced [6]. Whereas this assumption should predict MEGs, it is unlikely to successfully predict PEGs. However, in our PEG dataset we found five previously predicted imprinted genes [6], indicating that DNA methylation is important, but not the only regulator of imprinted gene expression. We analyzed the CG DNA methylation status of MEGs and PEGs in vegetative tissues and in the endosperm at 7–9 DAP using previously published data [7,41]. Allele-specific DNA methylation patterns are established during gametogenesis and immediately after fertilization [42,43], implicating that the endosperm DNA methylation profile at 7–9 DAP is similar to the profile at 4 DAP (time point of this study). There are two main classes of MEGs distinguishable based on the CG DNA methylation profile: MEGs without substantial CG DNA methylation in immediate vicinity to the genic regions (Figure 4 and Figure S3A–S3C) and MEGs with high levels of CG DNA methylation surrounding genic regions (Figure 5 and Figure S3D–S3F).

The CG DNA methylation profile of PEGs was clearly distinguishable from the MEG CG profiles; almost all PEGs had CG methylation peaks on average about 700 bps up- or downstream of the coding regions, whereas coding regions and the immediate up- and downstream regions were mostly devoid of CG DNA methylation (Figure 6 and Figure S4), suggesting that low levels of DNA methylation in this region are important for keeping PEGs transcriptionally active when paternally inherited and that DNA methylation at this region is unlikely to be responsible for keeping maternal alleles of PEGs silenced. The level of CG methylation in MEGs and PEGs was reduced in the endosperm providing an explanation why MEGs and PEGs were predicted in a previous study [6].
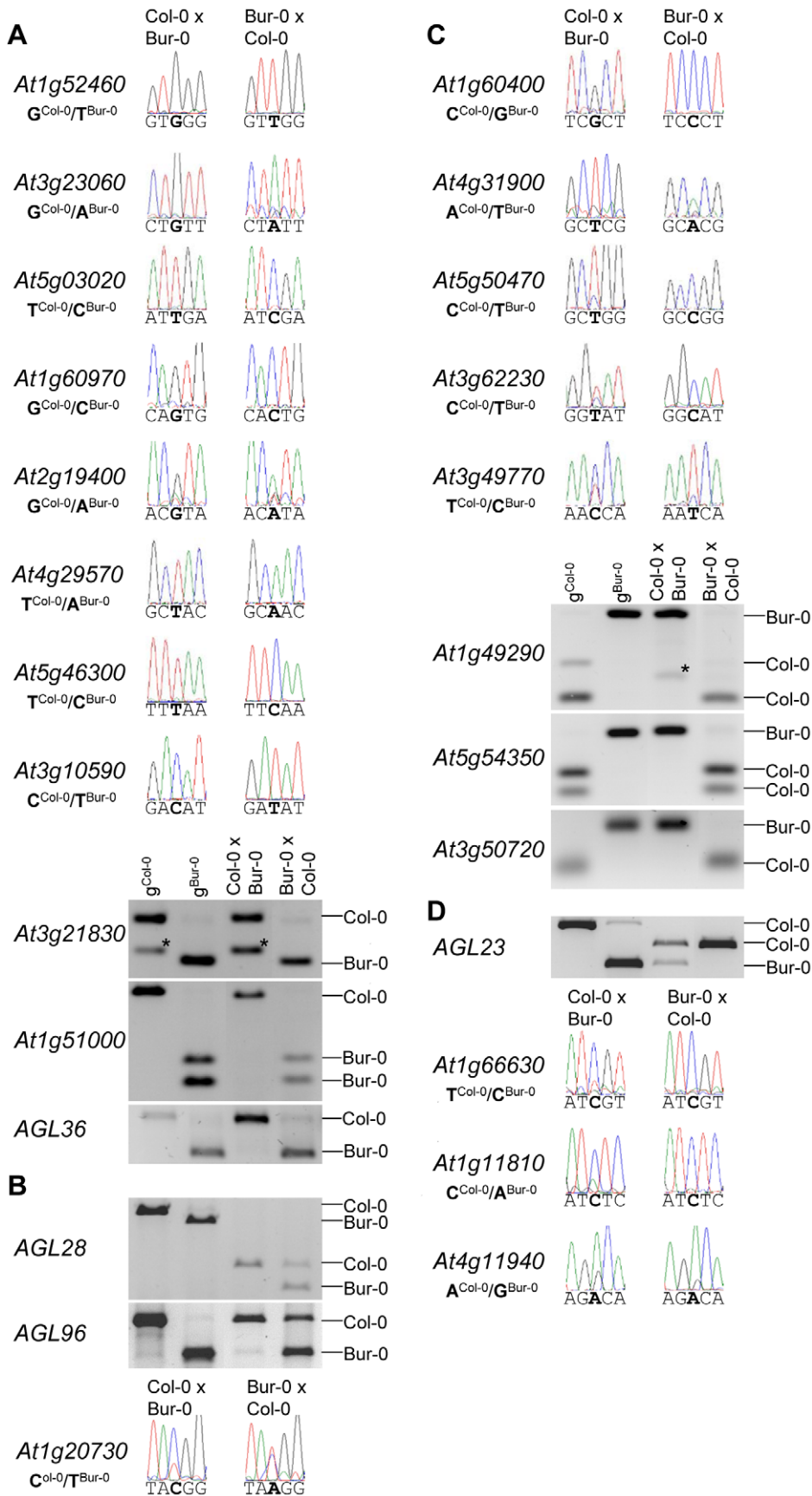
Figure 2. Allele-Specific Expression Analysis of MEGs and PEGs. Seeds of reciprocal crosses of Col-0 and Bur-0 accessions were harvested at 4 DAP and allele-specific expression was tested by restriction-based allele-specific PCR analysis or sequencing. MEGs and PEGs that are imprinted in both directions of the cross are shown in (A) and (C), MEGs and PEGs that are accession-dependently imprinted are shown in (B) and (D). Asterisks indicate unspecific PCR bands. Size differences between controls and cDNA samples are caused by the presence of introns in amplified regions.
doi:10.1371/journal.pgen.1002126.g002

Demethylation of the paternal genome by mutations in the DNA methyltransferase *MET1* has been demonstrated to cause biallelic expression of several MEGs [26,31,44,45]. Therefore, we tested allele-specific expression of confirmed MEGs in seeds derived after pollination with pollen of *met1/MET1* plants. Out of six tested MEGs that were not substantially CG methylated in the genic region, there were three genes that became either biallelically expressed upon pollination with *met1* pollen (*At1g52460*) or exclusively paternally expressed (*At5g46300*, *At3g10590*; Figure 4). The accession-dependent imprinted gene *At1g20730* which was preferentially maternally expressed in the cross Col-0×Bur-0 but biallelically expressed in the reciprocal cross became as well paternally expressed upon *met1* pollination (Figure 4). Analysis of non-CG DNA methylation revealed that within a distance of 2 kbps upstream of the transcriptional start site MEGs *At5g46300*, *At3g10590* and *At1g20730* were substantially marked by CHG or CHH methylation (Figure S5A), suggesting that non-CG DNA methylation might be involved in the repression of paternal MEG alleles. Non-CG DNA methylation levels were higher in the endosperm compared to vegetative tissues, in agreement with previous reports on frequent small interfering RNA–targeted hypermethylation of CHG and CHH target sites in the endosperm [7]. Out of five tested MEGs with substantial CG DNA methylation in the vicinity of genic regions only one MEG became reactivated upon pollination with *met1* pollen (*At3g23060*; Figure 5), suggesting that prominent CG DNA methylation marks are not a decisive criterion for DNA methylation-dependent repression of the paternal MEG alleles. Conversely, the absence of prominent CG DNA methylation in vicinity of genic regions does not exclude a regulatory role of DNA methylation on the activity status of paternal MEG alleles.

MEGs and PEGs Are Regulated by the FIS PcG Complex

As DNA methylation seemed unlikely to be responsible for repression of the paternal alleles of many MEGs as well as the maternal alleles of PEGs, we further investigated by which alternative mechanism imprinting of MEGs and PEGs is regulated. Almost all PEGs (25 out of 27 PEGs and accession-dependent PEGs) and many MEGs (31 out of 39 MEGs and accession-dependent MEGs) were PcG target genes in vegetative tissues (Tables S4, S7). In the endosperm, the average H3K27me3 levels of MEGs and PEGs were clearly increased over the H3K27me3 levels of all genes, with the H3K27me3 levels of PEGs being twice as high as the H3K27me3 levels of MEGs (Figure S6), suggesting that silencing of the maternally inherited alleles of PEGs is mediated by the FIS PcG complex. This hypothesis would predict increased expression of PEGs upon loss of FIS function. We tested this hypothesis by analyzing expression levels of PEGs in *fis2* mutants at 3 DAP and 6 DAP. Indeed, half of all PEGs and accession-dependent PEGs were significantly upregulated in the *fis2* mutant (Figure 6 and Figure S4). Furthermore, we tested allele-specific expression of four confirmed PEGs (*At4g31900*, *At1g49290*, *At5g54350*, *At3g50720*) and one accession-dependent confirmed PEG (*At1g66630*) in *fis2* and *fie* mutants lacking maternal FIS function. For all four PEGs as well as the accession-dependent PEG loss of FIS function caused activation of maternal PEG alleles, adding further support to our hypothesis (Figure 6).

Also 13 out of 39 MEGs and accession-dependent MEGs were significantly upregulated in the *fis2* mutant (Figure 4, Figure 5 and Figure S3). However, allele-specific expression analysis revealed reactivation of the paternal MEG allele in only two out of eleven tested MEGs (*At5g03020*, *At2g19400*; Figure 4 and Figure 5), indicating that increased expression of MEGs in *fis2* mutants is not necessarily a consequence of paternal MEG allele activation, but is likely caused by an increased expression of the maternal MEG alleles.

Previous studies revealed global deregulation of FIS PcG target genes in response to interploidy crosses (2n × 4n) [46,47]. Global deregulation of imprinted genes was furthermore proposed to account for interploidy seed defects [48]. Therefore, if the FIS PcG complex plays a major role in the regulation of imprinted genes, imprinted genes should become largely deregulated in response to interploidy crosses. We tested this hypothesis by analyzing MEG and PEG expression in seeds derived after pollination of diploid plants with tetraploid pollen donors. Indeed, most MEGs and PEGs that were deregulated in *fis2* were as well significantly deregulated in triploid seeds derived from interploidy crosses, adding support to this hypothesis (Figure 4, Figure 5, Figure 6, Figure S3 and Figure S4).

Together, our data reveal that maternal and paternal alleles of a subset of MEGs and maternal alleles of PEGs are regulated by the FIS PcG complex. The FIS PcG complex confers tight repression of the maternal PEG alleles and some paternal MEG alleles, whereas it mainly modulates expression of many maternal MEG alleles.

MEGs and PEGs Are Often Neighboured by Transposable Elements

Transposable elements have been implicated as the driving force for the evolution of imprinted gene expression [6,7,20]. Therefore, we addressed the question whether MEGs and PEGs have an increased likelihood to contain transposable elements in their vicinity compared to other detectable genes. Indeed, this test revealed a significantly increased likelihood for MEGs (p<0.009) as well as PEGs (p<1.3E-6). We also tested whether a particular subclass of transposable elements is enriched in the vicinity of MEGs and PEGs. Among the tested elements we noted a significant enrichment for RC/helitrons in PEGs (p<2.7 E-5). MEGs also had more RC/helitrons than expected by chance (8 versus 5), which was, however, not statistically significant (p<0.08). The presence of helitrons was previously reported within the 5′ region of imprinted genes *MEA* [49], *FWA* [31], *HDG3* and *HDG9* [6], implicating a functional association between the presence of helitrons and imprinted gene expression. Among MEGs we also noted a significant enrichment for MuDR (p = 0.01) and DNA transposable elements (p = 0.024), however, it is possible that (due to the relatively small sample size) enrichments for other elements were not detected. Correlating with a different CG DNA methylation pattern of MEGs and PEGs, the location of transposable elements in relation to the transcriptional start or stop differed between MEGs and PEGs (Figure 7B). PEGs had a much greater median distance of transposable elements in relation to the transcriptional start and stop compared to MEGs (Figure 7B), supporting previous observations of a distally located
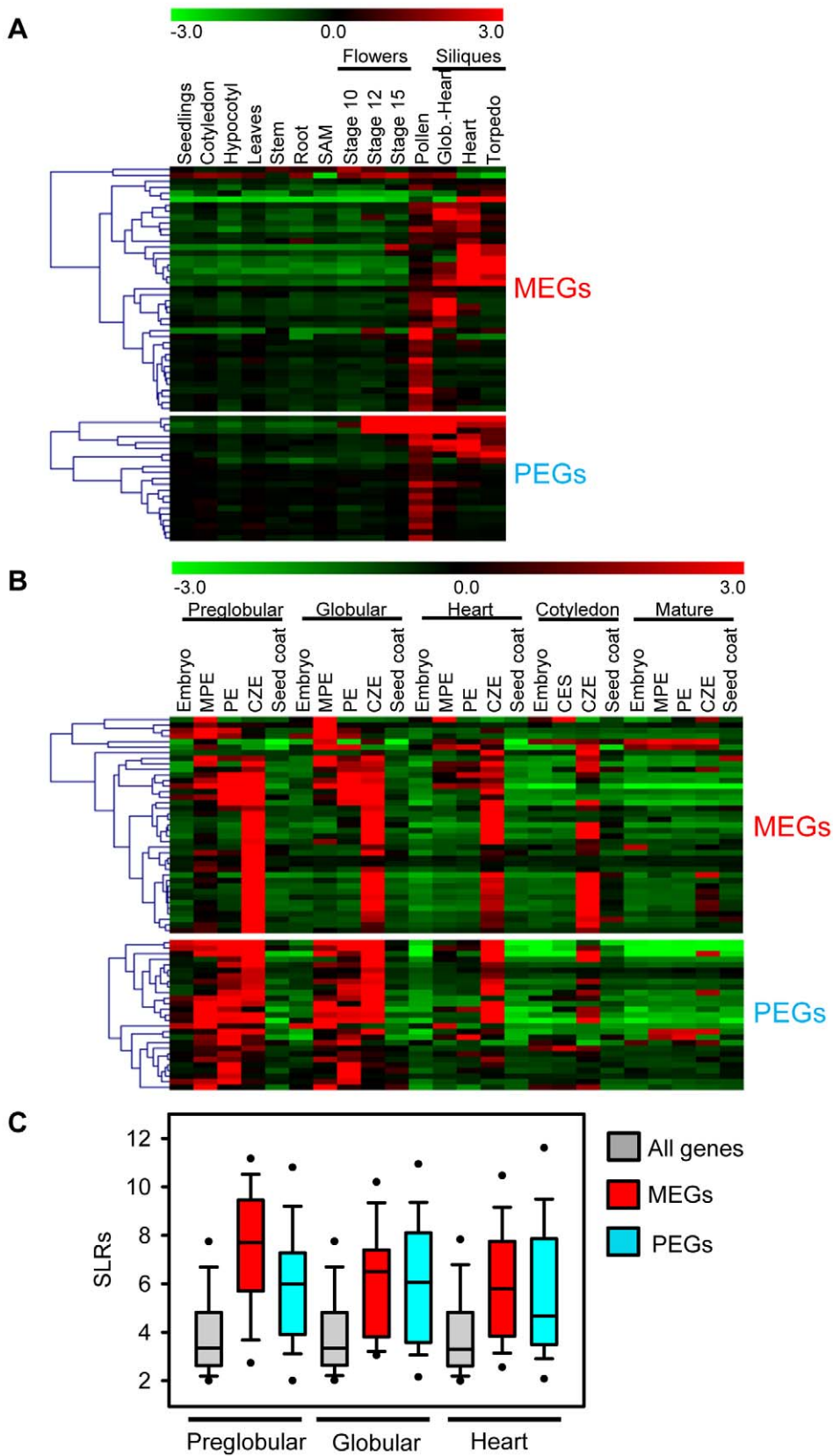
Figure 3. Expression Analysis of MEGs and PEGs in Vegetative and Seed Tissues. (A) Cluster analysis of MEGs and PEGs (including accession-dependent MEGs and PEGs) based on their expression in vegetative tissues and seeds. Each row represents a gene, and each column represents a tissue type. Tissue types are: seedlings, cotyledons, hypocotyl, leaves, stems, roots, shoot apical meristem (SAM), flowers at stages 10, 12, 15, siliques containing seeds with embryos in the globular to heart stage, heart stage and torpedo stage. Red or green indicate tissues in which a

particular gene is highly expressed or repressed, respectively. (B) Cluster analysis of MEGs and PEGs (including accession-dependent MEGs and PEGs) based on their expression in embryo, endosperm and seed coat during different stages of seed development. Each row represents a gene, and each column represents a tissue type. Tissue types are: embryos from the preglobular stage to the mature stage, micropylar (MPE), peripheral (PE) and chalazal (CZE) endosperm derived from seeds containing embryos of the preglobular stage to the mature stage, and seed coat derived from seeds containing embryos of the preglobular stage to the mature stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively. (C) Box plots of expression levels of MEGs (including accession-dependent MEGs; red) and PEGs (including accession-dependent PEGs; blue) compared to all genes (gray) in the chalazal endosperm region of seeds containing preglobular, globular and heart stage embryos. SLRs, Signal Log Ratios based on ATH1 microarray signals after RMA normalization.
doi:10.1371/journal.pgen.1002126.g003

helitron remnant being required for imprinted expression of the PEG *PHE1* [19].

MEGs and PEGs Are Faster Evolving When Compared to the Rest of the Genes in the *Arabidopsis thaliana* Genome

A number of studies of imprinted genes in both mammals and plants have found evidence that imprinted genes are rapidly evolving under positive Darwinian selection [27–29]. To determine whether either MEGs or PEGs displayed any evidence of fast evolution, pairwise dN/dS calculations were performed for the entire genome of *Arabidopsis thaliana*. Using reciprocal BLASTP searches, 19,965 orthologous pairs of proteins (gene models) were identified between *Arabidopsis thaliana* and *Arabidopsis lyrata* from the 27,235 *Arabidopsis thaliana* gene models tested. These included 27 (69.23%) of the MEGs and accession-dependent MEGs and 19 (70.37%) of the PEGs and accession-dependent PEGs identified in this study. Gene models not returning reciprocal BLAST results were not considered further for d_N/d_S analysis. Those gene models tested were then split into MEGs, PEGs and a third group representing the remaining genes tested. All three classes were then compared to determine whether there was any difference in relation to d_N/d_S values which measure rate of evolution of a protein coding-locus (Table S9).

In all three classes most genes have a d_N/d_S of less than one suggesting some level of purifying selection. However, the d_N/d_S value (as calculated here) is an average across the whole CDS sequence and masks potential heterogeneity of selection pressures across the gene. Despite this caveat, the d_N/d_S differences between imprinted and biallelically expressed genes is quite striking. For both MEGs and PEGs the reported median d_N/d_S is significantly higher than that of the background d_N/d_S for the remainder of the genome (p = 1.184e-05 and p = 2.991e-08 respectively, Wilcoxon Rank Sum), indicating that uniparentally expressed imprinted genes are fast evolving when compared to biallelically expressed genes (as represented by the background d_N/d_S values). However, although the median d_N/d_S for the MEGs is observed to be a third higher than that of the PEGs this difference is not reported as significant (p = 0.1614, Wilcoxon Rank Sum).

Also notable within the MEGs is that eight (~30%) of all the MEGs tested (i.e. *At1g61090, At3g57250, At1g51000, At5g46300, At4g29570 At1g12180, At1g52460* and *At1g07690*) reported a d_N/d_S value greater than one providing particularly strong evidence of fast evolution for these genes.

Within the PEGs, only one of the genes (~5%; *At2g20160*) displayed a d_N/d_S value greater than one. However, statistical testing did not reveal a significant difference between the number of fast evolving MEGs and PEGs (p = 0.0851, Fisher2019s exact test). Full details of the d_N/d_S analysis for both MEGs and PEGs is presented in Tables S10 and S11.

A Subset of MEGs and PEGs Are Located in Clusters

Imprinted loci in mammals are clustered over megabase regions in the genome and this clustering is essential to their imprinted regulation [50]; raising the question whether imprinted loci in

Arabidopsis are located within clusters as well. We searched for clustered MEGs and PEGs (including accession-dependent MEGs and PEGs) by applying a sliding window analysis. Using window sizes of 50 genes, significantly higher numbers of MEGs and PEGs were found to occur in clusters than expected by chance (p<0.05; Figure S7). We identified five MEG clusters containing two to three genes as well as three PEG clusters containing two genes (Figure 8A, 8B). Interestingly, most clusters contained either homologous MEGs or PEGs, or non-imprinted homologs of MEGs and PEGs (Figure 8A, 8B), implicating local sequence duplications as a driving force for the formation of imprinted genes. If so, there should be a higher incidence of imprinted genes having close sequence homologs compared to the genome-wide frequency of homologous genes. We tested this hypothesis by analyzing the number of close homologs to MEGs and PEGs and found indeed that MEGs and PEGs have an increased frequency of close homologs in comparison to the genome-wide frequency (p<0.05, Table S12), suggesting that gene duplications are in most cases connected with the formation of imprinted genes in Arabidopsis. Therefore, it seems possible that cluster formation of MEGs and PEGs is a consequence of local gene duplication and not essential for imprinted gene regulation, in agreement with the finding that only a subset of imprinted genes is localized in clusters.

Discussion

Unravelling the biological significance of genomic imprinting is crucially dependent on the identification of the majority of imprinted gene loci. In our study we succeeded in identifying more than 60 potentially imprinted loci, with a similar ratio of specifically maternally and paternally expressed imprinted genes. We successfully identified six out of twelve previously known imprinted genes, proving that our strategy can successfully identify novel imprinted genes. Six previously identified imprinted genes were not identified either because these genes are only weakly expressed at 4 DAP (*MEA* and *FIS2*) [46], they lack polymorphic sites between Col-0 and Bur-0 or because of accession-dependent imprinting. We only considered a gene to be imprinted if it had significantly deviating read numbers from the expected maternal to paternal ratio in both directions of the crosses. Those genes that passed this significance threshold were again tested for significant deviating read numbers when comparing the maternal and paternal Col-0 alleles versus the maternal and paternal Bur-0 alleles. Based on this comparison about 10% of the identified MEGs and 20% of the identified PEGs are likely to be imprinted only in one accession, however, many accession-dependent imprinted genes did not pass our first significance threshold (e.g. *HDG3, HDG8, HDG9*), suggesting that the number of accession-dependent imprinted genes is significantly higher. It has been noted that there is a difference between Arabidopsis accessions in their tolerance to interploidy crosses [51], whether accession-dependent imprinted genes are the underlying cause for this phenomenon is an attractive hypothesis.

Using publicly available microarray datasets we stringently filtered our MEG dataset for genes that are not expressed in
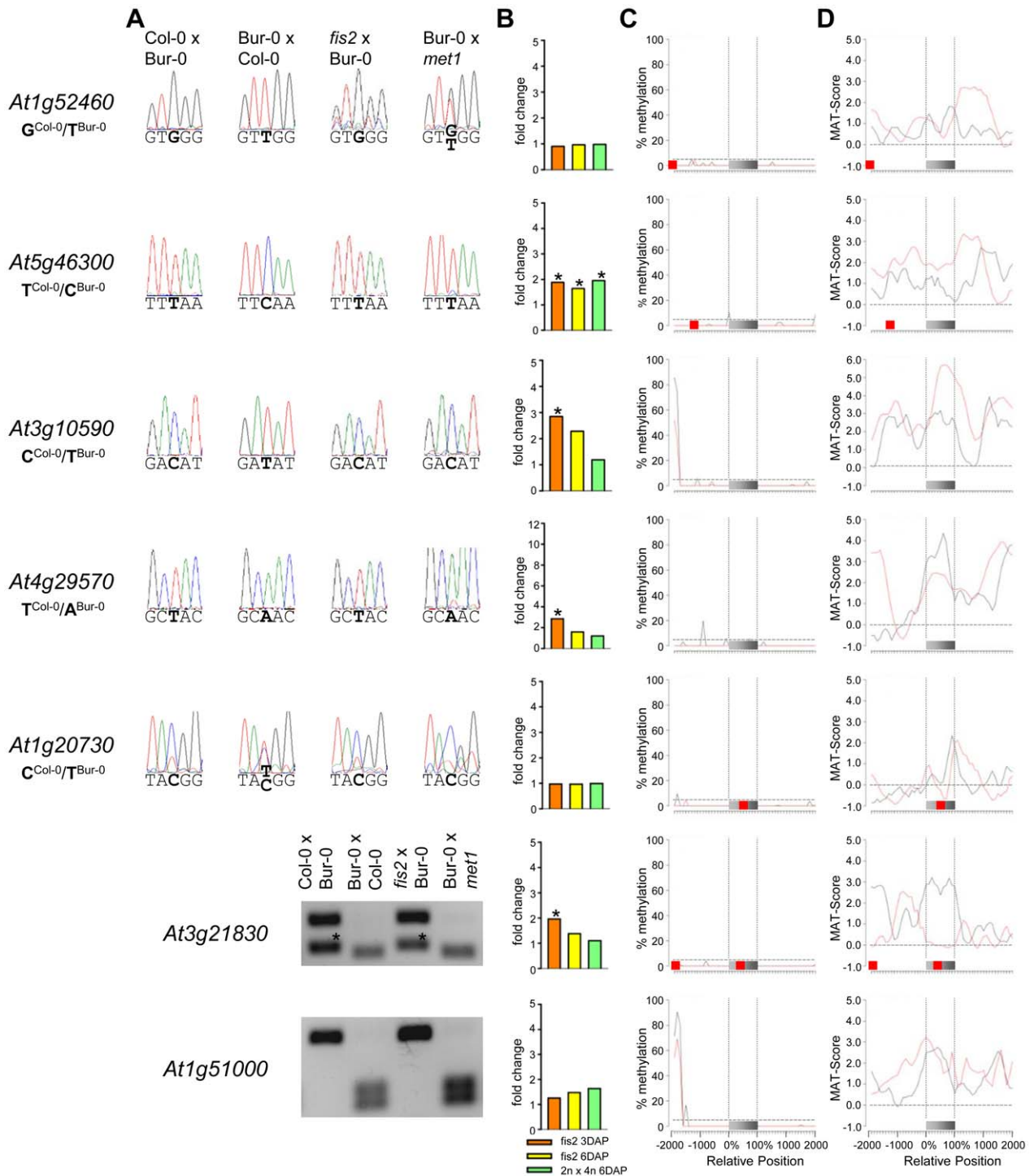
Figure 4. Impact of DNA Methylation and FIS PcG Function on the Regulation of Confirmed MEGs without Prominent Genic CG DNA Methylation. (A) Allele-specific expression analysis of indicated MEGs in seeds derived from crosses of Col-0×Bur-0, Bur-0×Col-0, *fis2*×Bur-0 and Col-0×*met1*. Seeds were harvested at 4 DAP and allele-specific expression was tested by restriction-based allele-specific PCR analysis or sequencing. Asterisks indicate unspecific PCR bands. (B) Fold-changes of MEG expression in *fis2* mutant seeds at 3 and 6 days after pollination (DAP) and from seeds derived from pollination with tetraploid pollen donors at 6 DAP compared to wild-type seeds at the corresponding time points. Data are based on ATH1 microarray signals after RMA normalization. Significantly deregulated genes are marked by an asterisk. (C) CG DNA methylation profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site. (D) H3K27me3 profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [52,64]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions

upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
doi:10.1371/journal.pgen.1002126.g004

vegetative tissues and the seed coat. This filtering allowed us to predict genes with allele-specific expression in the endosperm, but using this strategy we lost genes that are either not present on the ATH1 microarray (about 25%) or that are expressed in vegetative tissues but are still regulated by genomic imprinting. However, the vast majority of known imprinted genes are not significantly expressed during vegetative development, suggesting that we have identified a significant number of MEGs present in the genome. Although PEGs were not filtered against vegetative or seed coat expression, the majority of PEGs were similarly excluded from expression in vegetative tissues, indicating that imprinted genes in Arabidopsis have mainly endosperm-restricted functions.
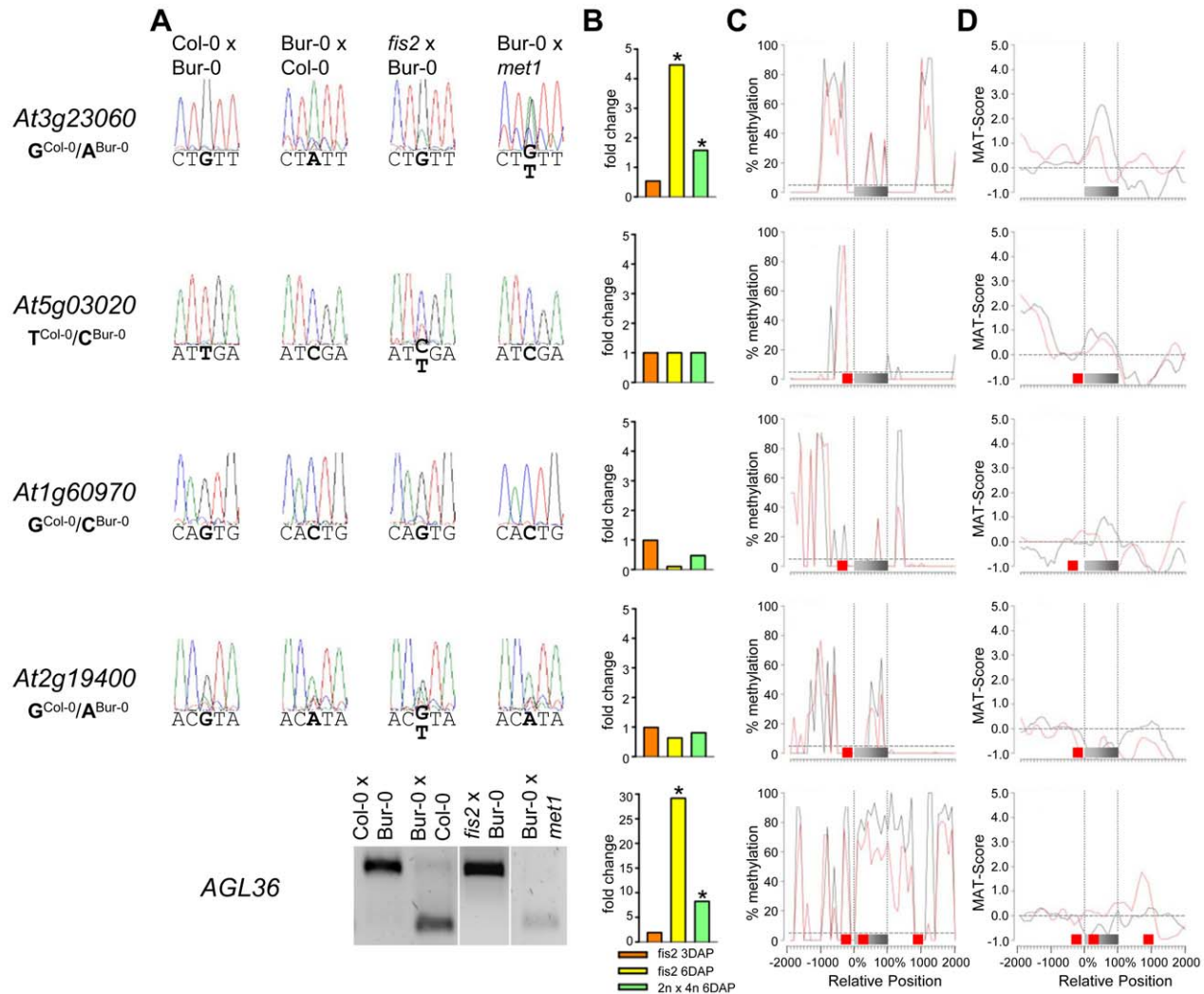

Figure 5. Impact of DNA Methylation and FIS PcG Function on the Regulation of Confirmed MEGs with Prominent Genic CG DNA Methylation. (A) Allele-specific expression analysis of indicated MEGs in seeds derived from crosses of Col-0×Bur-0, Bur-0×Col-0, fis2×Bur-0 and Col-0×met1. Seeds were harvested at 4 DAP and allele-specific expression was tested by restriction-based allele-specific PCR analysis or sequencing. (B) Fold-changes of MEG expression in fis2 mutant seeds at 3 and 6 days after pollination (DAP) and from seeds derived from pollination with tetraploid pollen donors at 6 DAP compared to wild-type seeds at the corresponding time points. Data are based on ATH1 microarray signals after RMA normalization. Significantly deregulated genes are marked by an asterisk. (C) CG DNA methylation profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site. (D) H3K27me3 profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [52,64]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
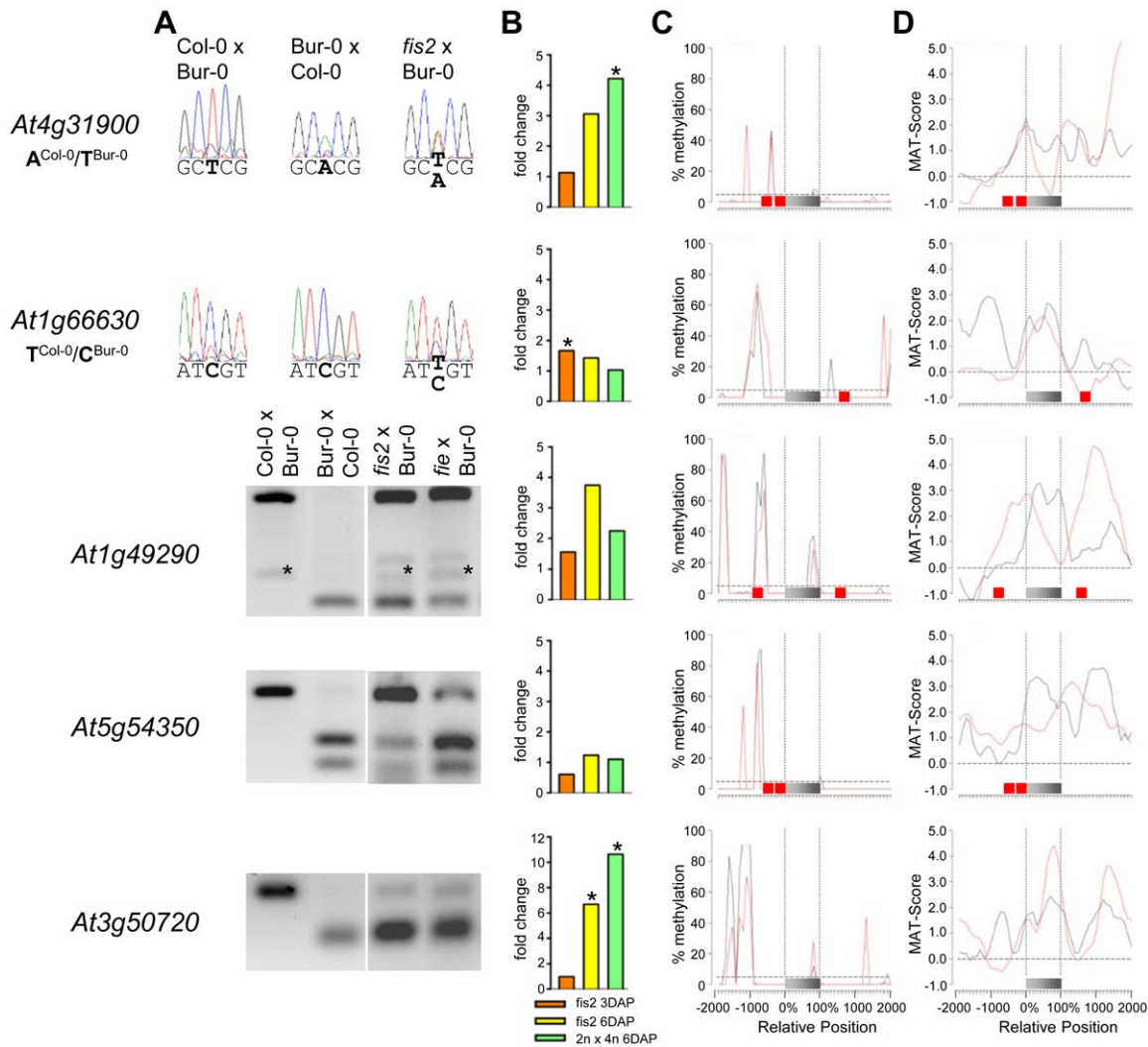doi:10.1371/journal.pgen.1002126.g005

Figure 6. Impact of FIS PcG Function on the Regulation of Confirmed PEGs. (A) Allele-specific expression analysis of indicated PEGs in seeds derived from crosses of Col-0×Bur-0, Bur-0×Col-0, fis2×Bur-0, and fie×Bur-0. Seeds were harvested at 4 DAP and allele-specific expression was tested by restriction-based allele-specific PCR analysis or sequencing. Asterisks indicate unspecific PCR bands. (B) Fold-changes of PEG expression in fis2 mutant seeds at 3 and 6 days after pollination (DAP) and from seeds derived from pollination with tetraploid pollen donors at 6 DAP compared to wild-type seeds at the corresponding time points. Data are based on ATH1 microarray signals after RMA normalization. Significantly deregulated genes are marked by an asterisk. (C) CG DNA methylation profiles of indicated PEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site. (D) H3K27me3 profiles of indicated PEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [52,64]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
doi:10.1371/journal.pgen.1002126.g006

Different Experimental Strategies Result in the Identification of Complementary Sets of Imprinted Genes

We compared the MEGs and PEGs identified in our study with MEGs and PEGs identified by a similar approach using the accession combinations Landsberg *erecta* (L*er*) and Col [45]. Whereas the majority of unfiltered MEGs identified by Hsieh and colleagues were present in our unfiltered MEG dataset (84%; 549 genes; Tables S13 and S14), only six genes were commonly identified as MEGs after filtering (Table S17; Figure S8A). The majority of MEGs defined by Hsieh and colleagues as being expressed in the endosperm were also present in our unfiltered

MEG dataset (78%, 89 genes, Figure S8A; Table S15). However, when analyzing the expression of these genes within different seed tissues we found the majority of them being strongly expressed in the seed coat (Figure S9) and did, therefore, not pass our stringent filtering criteria. Only eight out of 39 predicted MEGs identified in this study were as well present in the unfiltered dataset of Hsieh and colleagues (Figure S8A; Table S16) [45], indicating that differences in the filtering cannot sufficiently explain the differences between the identified MEG datasets. Similarly, only seven out of 119 predicted unfiltered PEGs overlapped with the unfiltered PEG dataset of Hsieh and colleagues (Table S18 and
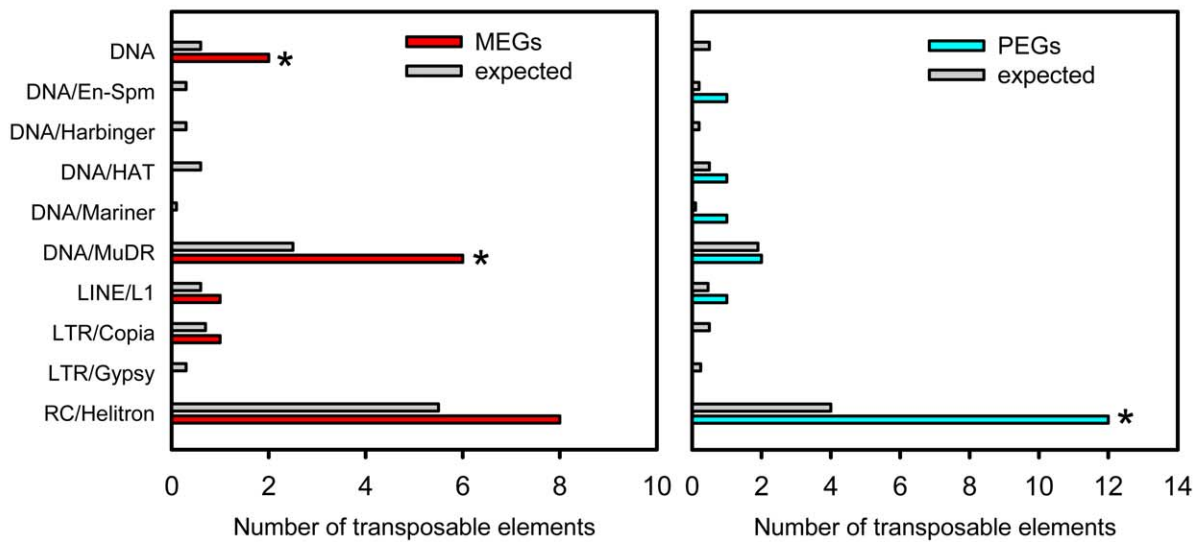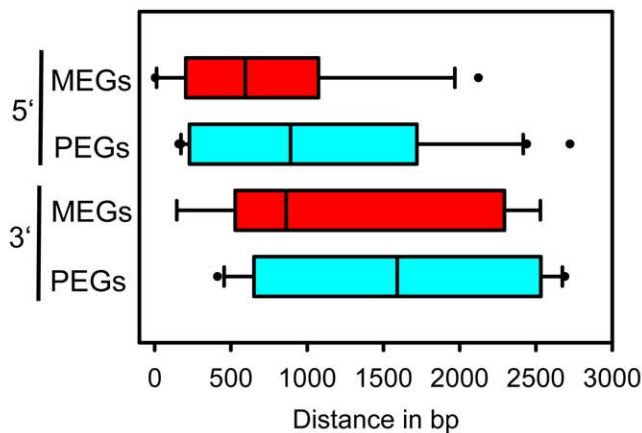
Figure 7. Types of Transposable Elements in the Vicinity of MEGs and PEGs. (A) Type of transposable elements present in MEGs (left panel) and PEGs (right panel) in comparison to their representation among detectable genes in our dataset (gray bars). TE superfamilies are as defined by TAIR (www.arabidopsis.org). (B) Distance of transposable elements in relation to the transcriptional start (5′ location) or stop (3′ location) of MEGs and PEGs.
doi:10.1371/journal.pgen.1002126.g007

Table S19, Figure S8B), supporting the view that differences in filtering strategies do not fully account for the different datasets. It thus seems likely that the different experimental setup between this study and the study by Hsieh and colleagues, including different accession combinations and different developmental stages (4 DAP versus 7–9 DAP in [45]) resulted in the identification of complementary datasets.

Regulation of MEGs and PEGs by DNA Methylation and the FIS PcG Complex

Most PEGs were devoid of CG DNA methylation around the transcriptional start site, indicating that DNA methylation is not responsible for silencing of maternal PEG alleles. Instead, we provide evidence that silencing of at least some maternal PEG alleles is mediated by the FIS PcG complex. We found that PEG loci have high H3K27me3 levels in the endosperm, and importantly, many PEGs were activated upon loss of FIS function,

which is likely contributed by reactivation of maternal PEG alleles. We also detected increased expression of MEGs upon loss of FIS function. However, this was not a necessary consequence of paternal MEG allele reactivation, but often caused by an activation of maternal MEG alleles, suggesting that endosperm hypomethylation makes maternal MEG alleles vulnerable to FIS silencing. This hypothesis is supported by previous findings from our group showing that genes and transposable elements are targeted by the FIS PcG complex in the endosperm, which are protected from PcG targeting by DNA methylation in vegetative tissues [52]. Which mechanism prevents the FIS complex from targeting paternal alleles of PEGs? We previously showed that DNA demethylation of a distally located region together with the promoter-localized FIS PcG complex are required for silencing of maternal *PHE1* alleles, suggesting that long-range interactions of sequence elements are required for efficient silencing of maternal *PHE1* alleles [19]. Here, we show that PEGs are flanked by
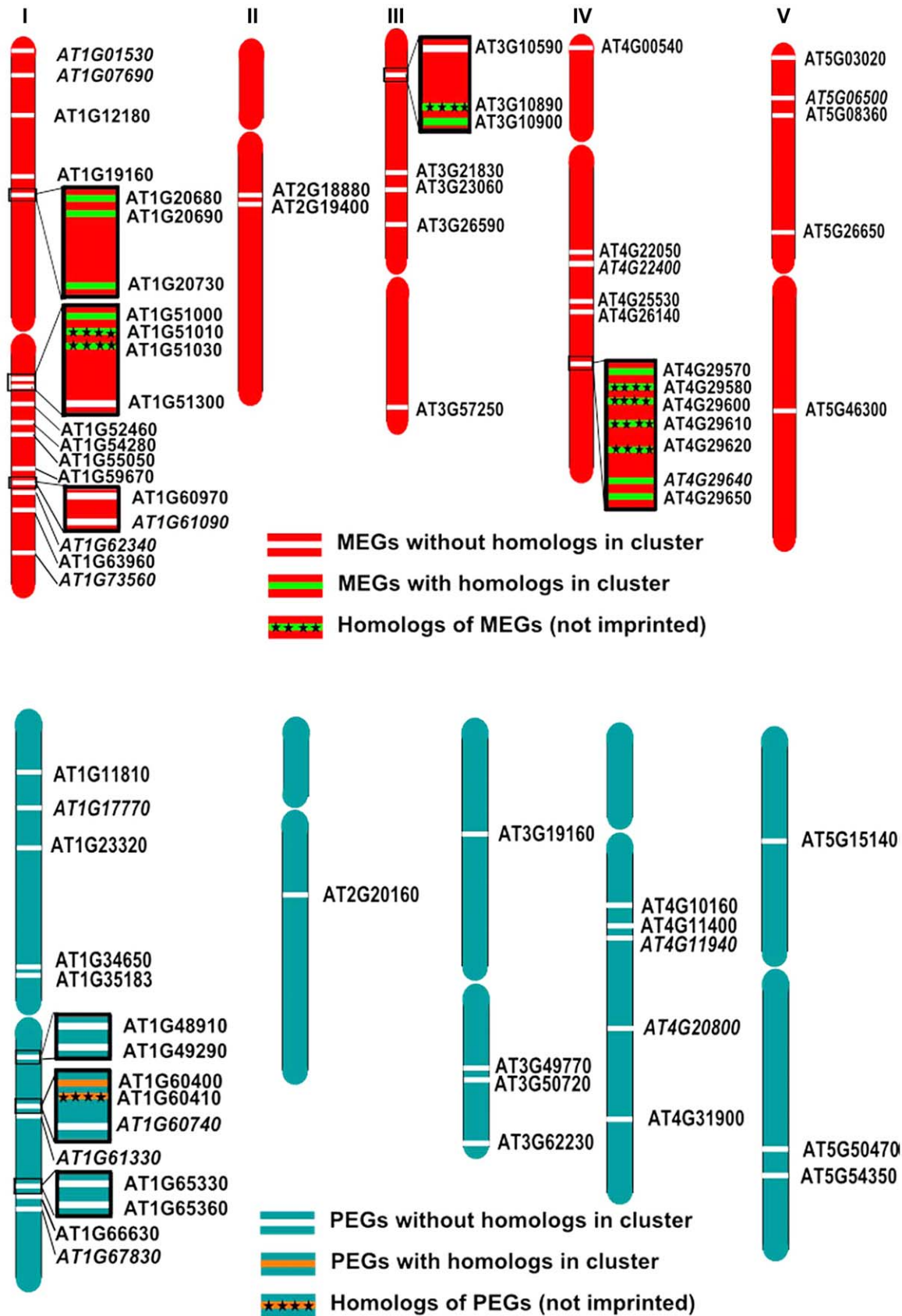
Figure 8. Some MEGs and PEGs Are Located in Clusters. Chromosomal distribution of MEGs and accession-dependent MEGs (A) and PEGs and accession-dependent PEGs (B) along the chromosomes. Accession-dependent MEGs and PEGs are italicized. Genes located in clusters are boxed. Clustered MEGs and PEGs having homologs within the cluster are highlighted in green and orange, respectively. Non-imprinted homologs of clustered MEGs and PEGs are indicated by cross signs.
doi:10.1371/journal.pgen.1002126.g008

regions of high CG DNA methylation levels, suggesting that upon DNA demethylation in the endosperm these regions are targeted by the FIS PcG complex, conferring tight repression of maternal PEG alleles. In support of this hypothesis, we found high H3K27me3 levels in PEG flanking regions, which were much higher than the H3K27me3 levels present in vegetative tissues (Figure 6, Figure S4 and Figure S6).

The impact of hypomethylation on the activity status of the paternal MEG alleles was contrasted by the lack of CG DNA methylation in the immediate vicinity of several MEGs. However, MEGs lacking CG methylation often contained substantial levels of non-CG DNA methylation (Figure S5), implicating that non-CG DNA methylation regulates paternal MEG alleles. Reactivation of these alleles upon loss of MET1 function might be a consequence of a proposed feedback between CG and non-CG DNA methylation [53,54]. Therefore, it seems possible that the paternal alleles of some MEGs are silenced by non-CG methylation established specifically in the endosperm and reactivation of the paternal alleles requires loss of non-CG methylation.

In stark contrast to silencing of MEGs and PEGs during vegetative development, many MEGs and PEGs were expressed in pollen. The vegetative cell of pollen has low levels of DNA methylation [55], suggesting that reduced levels of DNA methylation in the vegetative cell cause activation of MEGs, similar to the activation of maternal MEG alleles in the endosperm. Whether activation of PEGs in pollen is caused by reduced PcG protein activity in the vegetative cell of pollen remains to be tested.

Evolution of Imprinted Genes

Our study also sheds light on the evolution of imprinted genes, as we found a significant enrichment of transposon insertions in vicinity to MEGs and PEGs. In particular helitrons were strongly enriched in the vicinity of PEGs and were also overrepresented (albeit not statistically significant) in the vicinity of MEGs. Helitrons are eukaryotic DNA transposons which constitute >2% of the Arabidopsis genome. A striking feature of helitrons is their capacity to capture and propagate host genes, making them powerful factors shaping the evolution of genomes [56]. Although the rate of gene capture in Arabidopsis is predicted be to low compared to the major occurrence of gene capture in maize [57], it is possible that these predictions are a drastic underestimation due to rapid purging of helitron elements in Arabidopsis [58]. Thus, it is possible that helitron-mediated gene duplications which generate increased gene dosage may set the ground for imprinted gene evolution. Interestingly, we found a higher incidence of MEGs and PEGs having close sequence homologs, however, whether these duplications are a consequence of helitron activity remains to be shown.

If parental conflicts involving imprinted genes are mediated by amino-acid changes in the gene products of imprinted loci under antagonistic co-evolution, such imprinted loci may be subject to rapid evolution, possibly under positive Darwinian selection [27–29]. The identification of 60 potentially imprinted genes in Arabidopsis provided the opportunity for an initial exploration of whether the MEGs and PEGs identified displayed any evidence of rapid evolution. When the d_N/d_S values of the imprinted genes (MEGs and PEGs and accession-dependent MEGs and PEGs) were compared with the average d_N/d_S values for all other genes in the genome it is clear that imprinted genes in Arabidopsis are more rapidly evolving. Furthermore, a significant proportion of the MEGs displayed d_N/d_S values greater than 1 which is indicative of fast evolving genes. Further sequencing of these

imprinted genes in populations and outgroup species will determine whether these genes are undergoing positive Darwinian selection or are under relaxed constraints.

Imprinted genes are predominantly expressed in the endosperm, implicating a specific role of these genes during endosperm development. Alternatively, it is possible that imprinted genes are on the trajectory to become pseudogenes and therefore, are silenced during vegetative development. However, the fact that many imprinted genes have functional roles as transcription factors or have chromatin modifying activity supports a proposed functional role of imprinted genes for endosperm development. To identify these functions and to test whether MEGs and PEGs have indeed antagonistic roles in controlling endosperm growth as it has been proposed previously [21,22], will remain the challenge of future investigations.

Materials and Methods

Plant Material and Growth Conditions

The *Arabidopsis thaliana* accessions used in this study were Col-0 and Bur-0. The *fis2-5* and *met1-3* mutants (both in Col accession) were described previously [52,59]. The newly identified *fie-12* allele (GK_362D08; Col-0 accession) contains a T-DNA insertion within the third exon. The *fie-12* seed abortion ratio and mutant seed phenotypes were analyzed and found to be similar to previously published *fie* alleles (data not shown). All mutants were heterozygous and the genotype confirmed by PCR analysis. Plants were grown in a growth cabinet under long day photoperiods (16 h light and 8 h dark) at 22°C. After 10 days, seedlings were transferred to soil and plants were grown in a growth chamber at 60% humidity and daily cycles of 16 h light at 22°C and 8 h darkness at 18°C. For reciprocal crosses, designated female partners were emasculated, and the pistils hand-pollinated two days after emasculation.

Phenotypic Analysis of Seed Development in Col-0 and Bur-0 Accessions

To control for variations in seed development between Col-0 and Bur-0 accessions, siliques were harvested 4 DAP and fixed overnight in 9:1 ethanol:acetic acid. Siliques were dissected to release seeds into clearing solution (67% chloralhydrate in 8% glycerol) for overnight incubation. Microscopy imaging was performed using a Leica DM 2500 microscope using DIC optics (Leica Microsystems, Wetzlar, Germany), images were captured using a Leica DFC300 FX digital camera (Leica) and exported using Leica Application Suite Version 2.4.0.R1 (Leica Microsystems) and processed using Photoshop 7.0 (Adobe).

RNA Extraction and cDNA Synthesis

Seeds of at least 40 siliques per sample were harvested into 50 µl RNAlater (Sigma, Buchs, Switzerland) at 4 DAP. Glass beads (1.25–1.55 mm) were added, and the samples were ground unfrozen in a Silamat S5 (Ivoclar Vivadent, Ellwangen, Germany). RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For cDNA synthesis residual DNA was removed using the Qiagen RNase-free DNase Set and cDNA was synthesized using the Fermentas First strand cDNA synthesis kit (Fermentas, Burlington, Canada) according to the manufacturers instruction.

Preparation of mRNA-Sequencing Libraries

Sequencing libraries were prepared with the Illumina mRNA-Seq Sample Prep Kit (Illumina, San Diego, USA) according to the manufacturer's instructions. After adapter ligation library frag-

ments of ~250 bp were isolated from an agarose gel. The DNA was PCR amplified with Illumina primers for 15 cycles, purified and loaded on an Illumina flow cell for cluster generation. Libraries were sequenced on the Illumina Genome Analyzer II following the manufacturer's protocols.

Bioinformatic Analysis

Identification of SNP-associated reads. TAIR8 chromosome sequences were downloaded from TAIR (ftp.arabidopsis.org/home/tair/Genes/TAIR8_genome_release/tair8.at.chromosomes.fas, chromosomes 1 to5), together with the TAIR8 genome annotation (ftp.arabidopsis.org/home/tair/Genes/TAIR8_genome_release/TAIR8_gff3/TAIR8_GFF3_genes.gff). Bur-0 SNPs [30], 569,859 SNPs in total for chromosomes 1 to 5, were downloaded from the 1001 genomes website (http://1001genomes.org/data/MPI/MPIOssowski2008/releases/2008_06_05/strains/Bur-0/Bur-0_homozygous_snp_080605.txt). For each of the 569,859 Col-0/Bur-0 SNPs, we extracted a 71nt genomic window around the SNP (SNP position plus/minus 35nt) from the Col-0 reference sequence and annotated it as a Col-0 window. The nucleotide in the SNP was then replaced by the Bur-0 variant and the resulting sequence annotated as a Bur-0 window.

A suffix array was constructed from the union of Col-0 and Bur-0 windows with the mkvtree program (http://www.vmatch.de/) [60]. RNAseq reads were mapped with the vmatch program (http://www.vmatch.de/) in both forward and reverse complementary orientation (options -d and -p), allowing up to two mismatches (option -h 2), requiring the whole read to map (option -l 36), and generating maximal substring matches that are unique in the Col-0/Bur-0 window dataset (option -mum cand). This procedure resulted in 8,576,779 matches for Bur-x-Col reads (out of 102,705,076 total reads) and 6,773,239 matches for Col-x-Bur reads (out of 122,367,092 total reads). SNP windows were associated with gene ids via the TAIR8 genome annotation by overlapping with or being included in a gene region (gene start to end, ignoring exon/intron structure). SNP windows that were associated with more than one gene were discarded. From the remaining SNP windows, the grand total associated with a gene was defined as that gene's expression level. 147,349 SNP windows were matched by at least one read in one of the two reciprocal experiments, accounting for 19,161 genes, out of 25,189 genes that had at least one SNP in at least one exon, out of 28,523 annotated genes in total. Sequencing raw data generated in this study are available at GEO, accession number GSE27292.

Testing for allele-specific expression. For each gene and cross (Col-0×Bur-0 and Bur-0×Col-0), we performed a binomial one-sided test against the null-hypothesis of 2m:1p expression. The resulting p-values were the probabilities of deviation from the expected 2m:1p ratio towards either larger maternal expression or larger paternal expression under the null-hypothesis of an unbiased 2m:1p expression. The two p-values for maternal expression from the two reciprocal crosses (p1, p2) were summarized in a joint p-value based on the distribution of the second-order statistic by calculating $p = \max(p1,p2)^2$. Joint p-values for paternal expression were calculated analogously. Joint p-values, either describing reciprocal maternal expression or reciprocal paternal expression, were sorted in ascending order (from significant to insignificant), and for each joint p-value the false-discovery rate (FDR) [61] was calculated, as $q = p*n/i$, where n was the overall number of joint p-values and i was the rank of a given p-value. Genes with a q-value of 0.05 or less were selected as maternally or paternally expressed genes.

Parental-specific splicing was tested by analyzing for every candidate gene the numbers of reads across all SNPs of that gene,

using Pearson's chi-square test (R function chisq.test with parameter simulate.p.value = T).

Filtering for endosperm-expressed genes and analysis of MEGs and PEGs. Filtering for endosperm-specific gene expression was performed by using data from endosperm transcript profiles generated in the laboratories of Bob Goldberg (UCLA), John Harada (UC Davis), Brandon Le (UCLA), Anhthu Bui (UCLA), and Julie Pelletier (UC Davis) that are available under http://estdb.biology.ucla.edu/genechip/project [62]. The same data were used to generate Figure 3. Reference transcript profiles from vegetative tissues (seedlings, cotyledons, hypocotyl, leaves, stems, roots, shoot apical meristem), flowers and siliques were published by [63]. Genes were considered as preferentially expressed in the endosperm if the SLRs in one of the endosperm domains were at least fivefold higher than the SLRs of the seed coat and SLRs were below five in vegetative tissues. MEGs with low mRNA levels (read counts higher or equal 10 and smaller or equal 30) were considered as being endosperm-preferred expressed if SLRs in one of the endosperm domains were at least threefold higher compared to the seed coat and SLRs were below five in vegetative tissues (Table S3). Genes were considered as being expressed in the endosperm with SLRs>4.5 in at least one of the endosperm domains (Table S5). The transcript profiles of wild-type and *fis2* seeds at 3 and 6 DAP as well as seeds derived from interploidy crosses were published by [46,52]. H3K27me3 profiling data from vegetative tissues and the endosperm were published by [52,64]. Clustering analysis of expression profiles was performed using TM4 software [65]. DNA methylation profiles were taken from [7,41] and were visualized using R software (http://www.r-project.org/). Enrichment of GO categories (obtained from TAIR) was tested based on the hypergeometric test and multiple-testing correction according to [61] with a critical p-value of 5.0E-03. Homologous genes were identified using the blastp program from blastall (http://www.ncbi.nlm.nih.gov/staff/tao/URLAPI/blastall/) by applying matrix BLOSUM62 and a critical e-value of 0.01. Identification of MEG and PEG clusters was performed by establishing the frequency of MEGs and PEGs present in windows of a defined size using a sliding window analysis. p values were calculated from a reference distribution that was based on an identical number of randomly sampled genes. Transposable elements were identified based on information in the TAIR database (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR8_genome_release/TAIR8_Transposable_Elements.txt). Statistical testing was performed using a hypergeometric test as well as a permutation test. Both tests gave the same result.

Testing for evidence of rapid evolution of MEGs and PEGs. Reciprocal BLASTP searches were performed between *Arabidopsis thaliana* versus *Arabidopsis lyrata* (i.e. *Arabidopsis thaliana* peptides versus *Arabidopsis lyrata* peptide database and *Arabidopsis lyrata* peptides versus *Arabidopsis thaliana* peptide database) of all MEGs and PEGs listed in Tables S4 and S7, respectively. The reciprocal top hit sequences were then aligned at the peptide level using MUSCLE [66]. Using the peptide alignment as a template the reciprocal top hit CDS sequences were then aligned using the tranalign [67]. Pairwise d_N/d_S analysis was then performed on the CDS alignments using both the CodeML program (using model 0 and runmode -2) and yn00 both from the PAML package [68]. The values from CodeML are reported in the main text, values for both CodeML and yn00 are found in Tables S9, S10 and S11.

Validation of RNA-Sequencing Results

Selected loci were validated using independently prepared RNAs from reciprocal crosses between Col-0 and Bur-0. Primers used for allele specific expression analysis of selected genes are

specified in Table S20. The amplified products were either digested with indicated restriction enzymes (Table S20) and analyzed on agarose gels or by DNA sequencing.

Supporting Information

Figure S1 Expression of PEGs in Vegetative and Seed Tissues. (A) Cluster analysis of PEGs and accession-dependent PEGs based on their expression in vegetative tissues and seeds. PEGs were grouped into two mutually exclusive clusters based on their expression in pollen. The cluster containing genes with low or without expression in pollen is marked by a light blue bar; the cluster containing genes with high pollen expression is marked with a dark blue bar. Each row represents a gene, and each column represents a tissue type. Tissue types are: seedlings, cotyledons, hypocotyl, leaves, stems, roots, shoot apical meristem (SAM), flowers at stages 10, 12, 15, siliques containing seeds with embryos in the globular to heart stage, heart stage and torpedo stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively. (B) Cluster analysis of PEGs and accession-dependent PEGs based on their expression in embryo, endosperm and seed coat during different stages of seed development. PEGs were grouped into three mutually exclusive clusters based on their expression in embryo and the endosperm. The cluster containing genes with low or without expression in embryo and endosperm is marked with a light orange bar, the cluster containing genes with low expression in the embryo but high endosperm expression is marked with a dark orange bar, and the cluster containing genes with high expression in embryo but low or without expression in the endosperm is marked with a red bar. Each row represents a gene, and each column represents a tissue type. Tissue types are: embryos from the preglobular stage to the mature stage, micropylar (MPE), peripheral (PE) and chalazal (CZE) endosperm derived from seeds containing embryos of the preglobular stage to the mature stage, and seed coat derived from seeds containing embryos of the preglobular stage to the mature stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively.
(PDF)

Figure S2 Interaction Network of AGL Transcription Factors Based on Yeast Two Hybrid Interaction Data [37]. Maternally expressed MEGs are indicated in red, paternally expressed AGLs are indicated in blue. The central regulator AGL62 is depicted in orange.
(PDF)

Figure S3 CG DNA Methylation Profiles of MEGs in Vegetative Tissues and Endosperm and Expression Analysis of MEGs in *fis2* and 2n×4n Interploidy Crosses. (A–C) Analysis of MEGs without prominent CG DNA methylation levels in the vicinity of genic regions. (D–F) Analysis of MEGs with prominent CG DNA methylation levels in the vicinity of genic regions. (A, D) Fold-changes of MEG expression in *fis2* mutant seeds at 3 and 6 days after pollination (DAP) and from seeds derived from pollination with tetraploid pollen donors at 6 DAP compared to wild-type seeds at the corresponding time points. Data are based on ATH1 microarray signals after RMA normalization. Significantly deregulated genes are marked by an asterisk. (B, E) CG DNA methylation profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals

for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site. (C, F) H3K27me3 profiles of indicated MEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [52,64]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
(PDF)

Figure S4 CG DNA Methylation Profiles of PEGs in Vegetative Tissues and Endosperm and Expression Analysis of PEGs in *fis2* and 2n×4n Interploidy Crosses. (A) Fold-changes of PEG expression in *fis2* mutant seeds at 3 and 6 days after pollination (DAP) and from seeds derived from pollination with tetraploid pollen donors at 6 DAP compared to wild-type seeds at the corresponding time points. Data are based on ATH1 microarray signals after RMA normalization. Significantly deregulated genes are marked by an asterisk. (B) CG DNA methylation profiles of indicated PEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site. (C) H3K27me3 profiles of indicated PEGs in vegetative tissues (black line) or endosperm (red line) based on data published by [52,64]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
(PDF)

Figure S5 CHG and CHH Methylation Profiles of MEGs. (A, B) CHG (A) and CHH (B) DNA methylation profiles of MEGs shown in Figure 4 in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. (C, D) CHG (C) and CHH (D) DNA methylation profiles of MEGs shown in Figure S3A in vegetative tissues (black line) or endosperm (red line) based on data published by [7,41]. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Red boxes represent transposable elements. Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the DNA methylation level in vegetative tissues of TAIR8-annotated genes at the transcriptional start site.
(PDF)

Figure S6 Average H3K27me3 profiles of vegetative tissues and endosperm. Average H3K27me3 profiles of vegetative tissues (black line) or endosperm (red line) of TAIR8-annotated genes (left panels), MEGs (middle panels), and PEGs (right panels). MEGs and PEGs correspond to all genes indicated in Tables S4 and S7,

respectively. The gray bar represents the annotated gene body from transcription start (left) to transcription end (right). Profiles are shown for 5% length intervals along the gene body and for 100 bp sequence intervals for the 2-kb regions upstream and downstream of each gene. The vertical dotted lines mark the gene body. The horizontal dashed line marks the H3K27me3 level of TAIR8-annotated genes at the transcriptional start site.
(PDF)

Figure S7 Identification of windows containing significantly enriched numbers of clustered MEGs (A) and PEGs (B). The significance threshold (p = 0.05) is indicated by a red line.
(PDF)

Figure S8 Overlap of MEGs (A) and PEGs (B) identified by [45] and MEGs and PEGs identified in this study. MEGs_LC and PEGs_LC correspond to MEGs and PEGs identified by [45] using Ler/Col accessions; MEGs_BC and PEGs_BC correspond to MEGs and PEGs identified in this study using Bur-0/Col-0 accessions. Unfiltered MEGs_LC and PEGs_LC were identified by [45] using p≤0.001 and p≤0.05, respectively. Unfiltered MEGs_BC and PEGs_BC correspond to data shown in Tables S1 and S2, respectively. Filtered MEGs_BC and PEGs_BC correspond to data shown in Tables S4 and S7, respectively.
(PDF)

Figure S9 Cluster analysis of MEGs identified by [45] that overlap with unfiltered MEGs identified in this study (Table S15). Cluster analysis of MEGs was based on their expression in embryo, endosperm and seed coat during different stages of seed development. The cluster containing genes with low or without expression in seed coat is marked by a vertical orange bar. The cluster containing genes with high expression in the seed coat is marked by a vertical yellow bar. Genes present in our filtered MEG dataset (Table S7) are indicated. Each row represents a gene, and each column represents a tissue type. Tissue types are: embryos from the preglobular stage to the mature stage, micropylar (MPE), peripheral (PE) and chalazal (CZE) endosperm derived from seeds containing embryos of the preglobular stage to the mature stage, and seed coat derived from seeds containing embryos of the preglobular stage to the mature stage. Red or green indicate tissues in which a particular gene is highly expressed or repressed, respectively.
(PDF)

Table S1 Genes with maternally-biased expression (MEGs). m, Maternal, p, Paternal alleles.
(XLSX)

Table S2 Genes with paternally-biased expression (PEGs). m, Maternal, p, Paternal alleles.
(XLSX)

Table S3 Genes that are expressed in at least one stage and domain of endosperm development until heart stage and are not expressed in the seed coat and in vegetative tissues (seedlings, cotyledons, hypocotyl, leaves, stems, roots, shoot apical meristem, flowers at stages 10, 12).
(XLSX)

Table S4 MEGs and accession-dependent MEGs. m, Maternal, p, Paternal alleles. Genes marked in yellow are PcG target genes in vegetative tissues. Genes marked in orange have higher expression levels in flowers stage 12 than in siliques containing seeds with globular stage embryos (SLRs_Seeds-SLRs_Flowers). AD, accession-dependent imprinting; + imprinting confirmed; -, imprinting not confirmed.
(XLSX)

Table S5 Genes that are expressed in at least one stage and domain of endosperm development until heart stage.
(XLSX)

Table S6 Genes with paternally-biased expression (PEGs). m, Maternal, p, Paternal alleles.
(XLSX)

Table S7 PEGs and accession-dependent PEGs. m, Maternal, p, Paternal alleles. Genes marked in yellow are PcG target genes in vegetative tissues. AD, accession-dependent imprinting; + imprinting confirmed; -, imprinting not confirmed.
(XLSX)

Table S8 GO analysis of MEGs and PEGs (including accession-dependent MEGs and PEGs).
(PDF)

Table S9 Median omega, dN and dS values as calculated from pairwise alignments between *Arabidopsis thaliana* and *Arabidopsis lyrata* orthologs. Measure of spread represented as the semi interquartile range.
(PDF)

Table S10 Outputs for MEGs from Codeml and YN00 programs from the Codeml package.
(XLSX)

Table S11 Outputs for PEGs from Codeml and YN00 programs from the Codeml package.
(XLSX)

Table S12 Number of homologous genes to MEGs and PEGs (including accession-dependent MEGs and PEGs) in comparison to the genome-wide frequency of gene homologs.
(PDF)

Table S13 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of unfiltered MEGs_LC (p<0.05) and unfiltered MEGs_BC (Table S1). MEGs_LC, MEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; MEGs_BC, MEGs identified in this study using Bur-0/Col-0 accessions.
(XLSX)

Table S14 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of unfiltered MEGs_LC (p<0.001) and unfiltered MEGs_BC (Table S1). MEGs_LC, MEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; MEGs_BC, MEGs identified in this study using Bur-0/Col-0 accessions.
(XLSX)

Table S15 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of MEGs_LC (p<0.001 filtered for endosperm expression) and unfiltered MEGs_BC (Table S1). MEGs_LC, MEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; MEGs_BC, MEGs identified in this study using Bur-0/Col-0 accessions.
(XLSX)

Table S16 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of unfiltered MEGs_LC (p<0.001) and filtered MEGs_BC (Table S7). MEGs_LC, MEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; MEGs_BC, MEGs identified in this study using Bur-0/Col-0 accessions. Genes marked in yellow have been identified as accession-dependent MEGs_BC.
(XLSX)

Table S17 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of filtered MEGs_LC (p<0.001) and filtered MEGs_BC (Table S7). MEGs_LC, MEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; MEGs_BC, MEGs identified in this study using Bur-0/Col-0 accessions. Genes marked in yellow have been identified as accession-dependent MEGs_BC.
(XLSX)

Table S18 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of unfiltered PEGs_LC (p<0.05) and unfiltered PEGs_BC (Table S2). PEGs_LC, PEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; PEGs_BC, PEGs identified in this study using Bur-0/Col-0 accessions. Genes marked in red are not present on the ATH1 miroarray. Genes marked in yellow have been identified as accession-dependent PEGs_BC.
(XLSX)

Table S19 Comparative analysis of parent-of-origin specific genes identified in this study and published in [45]. Overlap of unfiltered PEGs_LC (p<0.05) and filtered PEGs_BC (Table S7). PEGs_LC, PEGs identified by Hsieh and colleagues [45] using Ler/Col-0 accessions; PEGs_BC, PEGs identified in this study using Bur-0/Col-0 accessions. Genes marked in yellow have been identified as accession-dependent PEGs_BC.
(XLSX)

Table S20 Primers and enzymes used for allele-specific expression analysis.
(PDF)

Author Contributions

Conceived and designed the experiments: PW IW CS MR CK. Performed the experiments: PW IW JS PR. Analyzed the data: PW IW JS MTAD CS MR CK. Contributed reagents/materials/analysis tools: CB CS MR MN CK. Wrote the paper: PW CS MR CK.

References

1. Köhler C, Weinhofer-Molisch I (2010) Mechanisms and evolution of genomic imprinting in plants. Heredity 105: 57–63.
2. Jullien PE, Berger F (2009) Gamete-specific epigenetic mechanisms shape genomic imprinting. Curr Opin Plant Biol 12: 637–642.
3. Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. Curr Biol 19: 1677–1681.
4. Berger F (2003) Endosperm: the crossroad of seed development. Curr Opin Plant Biol 6: 42–50.
5. Drews GN, Yadegari R (2002) Development and function of the angiosperm female gametophyte. Annu Rev Genet 36: 99–124.
6. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science 324: 1447–1451.
7. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451–1454.
8. Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, et al. (2008) Retinoblastoma and its binding partner MSI1 control imprinting in Arabidopsis. PLoS Biol 6: e194. doi:10.1371/journal.pbio.0060194.
9. Teixeira FK, Colot V (2010) Repeat elements and the Arabidopsis DNA methylation landscape. Heredity 105: 14–23.
10. Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. Nat Genet 37: 28–30.
11. Baroux C, Gagliardini V, Page DR, Grossniklaus U (2006) Dynamic regulatory interactions of Polycomb group genes: *MEDEA* autoregulation is required for imprinted gene expression in Arabidopsis. Genes Dev 20: 1081–1086.
12. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, et al. (2006) DEMETER DNA glycosylase establishes MEDEA Polycomb gene self-imprinting by allele-specific demethylation. Cell 124: 495–506.
13. Jullien PE, Katz A, Oliva M, Ohad N, Berger F (2006) Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis. Curr Biol 16: 486–492.
14. Fitz Gerald JN, Hui PS, Berger F (2009) Polycomb group-dependent imprinting of the actin regulator AtFH5 regulates morphogenesis in Arabidopsis thaliana. Development 136: 3399–3404.
15. Hennig L, Derkacheva M (2009) Diversity of Polycomb group complexes in plants: same rules, different players? Trends Genet 25: 414–423.
16. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. Cell 110: 33–42.
17. Xiao W, Gehring M, Choi Y, Margossian L, Pu H, et al. (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. Dev Cell 5: 891–901.
18. Makarevich G, Villar CB, Erilova A, Köhler C (2008) Mechanism of PHERES1 imprinting in Arabidopsis. J Cell Sci 121: 906–912.
19. Villar CB, Erilova A, Makarevich G, Trösch R, Köhler C (2009) Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. Mol Plant 2: 654–660.
20. Barlow DP (1993) Methylation and imprinting: from host defense to gene regulation? Science 260: 309–310.
21. Haig D, Westoby M (1989) Parent specific gene expression and the triploid endosperm. Am Nature 134: 147–155.
22. Trivers R, Burt A (1999) Kinship and genomic imprinting. Results Probl Cell Differ 25: 1–21.
23. Reik W, Constancia M, Fowden A, Anderson N, Dean W, et al. (2003) Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. J Physiol 547: 35–44.
24. Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, et al. (1997) Fertilization-independent seed development in Arabidopsis thaliana. Proc Natl Acad Sci USA 94: 4223–4228.
25. Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, et al. (1999) Control of fertilization-independent endosperm development by the MEDEA Polycomb gene in Arabidopsis. Proc Natl Acad Sci U S A 96: 4186–4191.
26. Tiwari S, Schulz R, Ikeda Y, Dytham L, Bravo J, et al. (2008) MATERNALLY EXPRESSED PAB C-TERMINAL, a Novel Imprinted Gene in Arabidopsis, Encodes the Conserved C-Terminal Domain of Polyadenylate Binding Proteins. Plant Cell 20: 2387–2398.
27. Spillane C, Schmid KJ, Laoueille-Duprat S, Pien S, Escobar-Restrepo JM, et al. (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. Nature 448: 349–352.
28. Miyake T, Takebayashi N, Wolf DE (2009) Possible diversifying selection in the imprinted gene, MEDEA, in Arabidopsis. Mol Biol Evol 26: 843–857.
29. O'Connell MJ, Loughran NB, Walsh TA, Donoghue MT, Schmid KJ, et al. (2010) A phylogenetic approach to test for evidence of parental conflict or gene duplications associated with protein-encoding imprinted orthologous genes in placental mammals. Mamm Genome 21: 486–498.
30. Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, et al. (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 18: 2024–2033.
31. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, et al. (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303: 521–523.
32. Kermicle J (1970) Dependence of the R-mottled aleurone phenotype in maize on the mode of sexual transmission. Genetics 66: 69–85.
33. Chaudhuri S, Messing J (1994) Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. Proc Natl Acad Sci U S A 91: 4867–4871.
34. Christensen C, King E, Jordan J, Drews GN (1997) Megagametogenesis in Arabidopsis wild type and the Gf mutant. Sexual Plant Reproduction 10: 49–64.
35. Bayer M, Nawy T, Giglione C, Galli M, Meinnel T, et al. (2009) Paternal control of embryonic patterning in Arabidopsis thaliana. Science 323: 1485–1488.

36. Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, et al. (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. Plant J 54: 1037–1048.

37. de Folter S, Immink RG, Kieffer M, Parenicova L, Henz SR, et al. (2005) Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. Plant Cell 17: 1424–1433.

38. Kang IH, Steffen JG, Portereiko MF, Lloyd A, Drews GN (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. Plant Cell 20: 635–647.

39. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, et al. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell 135: 1201–1212.

40. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, et al. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105.

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

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

43. Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. Science 330: 622–627.

44. Jullien PE, Kinoshita T, Ohad N, Berger F (2006) Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell 18: 1360–1372.

45. Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, et al. (2011) Inaugural Article: Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A.

46. Erilova A, Brownfield L, Exner V, Rosa M, Twell D, et al. (2009) Imprinting of the Polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis. PLoS Genet 5: e1000663. doi:10.1371/journal.pgen.1000663.

47. Tiwari S, Spielman M, Schulz R, Oakey RJ, Kelsey G, et al. (2010) Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. BMC Plant Biol 10: 72.

48. Jullien PE, Berger F (2010) Parental genome dosage imbalance deregulates imprinting in Arabidopsis. PLoS Genet 6: e1000885. doi:10.1371/journal.pgen.1000885.

49. Spillane C, Baroux C, Escobar-Restrepo JM, Page DR, Laoueille S, et al. (2004) Transposons and tandem repeats are not involved in the control of genomic imprinting at the MEDEA locus in Arabidopsis. Cold Spring Harb Symp Quant Biol 69: 465–475.

50. Wan LB, Bartolomei MS (2008) Regulation of imprinting in clusters: noncoding RNAs versus insulators. Adv Genet 61: 207–223.

51. Dilkes BP, Spielman M, Weizbauer R, Watson B, Burkart-Waco D, et al. (2008) The maternally expressed WRKY transcription factor TTG2 controls lethality in interploidy crosses of Arabidopsis. PLoS Biol 6: e308. doi:10.1371/journal.pbio.0060308.

52. Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C (2010) H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet 6: e1001152. doi:10.1371/journal.pgen.1001152.

53. Chan SW, Henderson IR, Zhang X, Shah G, Chien JS, et al. (2006) RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in arabidopsis. PLoS Genet 2: e83. doi:10.1371/journal.pgen.0020083.

54. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133: 523–536.

55. Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, et al. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

56. Kapitonov VV, Jurka J (2007) Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet 23: 521–529.

57. Hollister JD, Gaut BS (2007) Population and evolutionary dynamics of Helitron transposable elements in Arabidopsis thaliana. Mol Biol Evol 24: 2515–2524.

58. Sweredoski M, DeRose-Wilson L, Gaut BS (2008) A comparative computational analysis of nonautonomous helitron elements between maize and rice. BMC Genomics 9: 467.

59. Saze H, Scheid OM, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet 34: 65–69.

60. Abouelhoda M, Kurtz S, Ohlebusch E (2004) Replacing suffix trees with enhanced suffix arrays. Journal of Discrete Algorithms 2: 53–86.

61. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B 57: 289–300.

62. Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, et al. (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci U S A 107: 8063–8070.

63. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, et al. (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet 37: 501–506.

64. Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, et al. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol 5: e129. doi:10.1371/journal.pbio.0050129.

65. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.

66. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

67. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276–277.

68. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556.