

Investigations on Dynamic Regulation of
Genomic Imprinting in the *Arabidopsis*
Endosperm

Yuri Stephan van Ekelenburg

Thesis submitted for the degree of
Philosophiae Doctor (Ph.D.)
2022

Section for Genetics and Evolutionary Biology (EVOGENE)

Department of Biosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

© Yuri Stephan van Ekelenburg, 2022

*Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2526*

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: Graphics Center, University of Oslo.

Acknowledgements

This thesis has been performed under the supervision of Professor Paul E. Grini and co-supervision of Dr. Katrine N. Bjerkan and Assistant Professor Jason R. Miller.

Firstly, I would like to thank Paul for being my main supervisor, helping me through my PhD and providing a listening ear. Special thanks to Jason for guiding me through the bioinformatic analyses performed in this thesis. I would also like to especially thank Ida V. Myking for being an amazing master's student and friend.

I have been very fortunate to do my PhD at EVOGENE and would like to thank all of the members and staff for their daily contribution to the great atmosphere at this section. A special thanks to my friends and colleagues Anne, Ida, Renate, Sergio, Vegard and Verena, who have made my PhD period at Blindern a lot of fun.

Many thanks to my friends and family in Norway and the Netherlands. It has been exciting to explain and discuss what I have been doing for the past few years. Even though COVID19 has not made it easy for us to visit each other, I am very grateful that I have been able to talk to you digitally.

Finally, I would like to thank the most important person that has made this PhD possible, my partner Sissel. Thank you for your endless support and comfort when I needed that.

Oslo, March 2022.

Yuri

Table of Contents

Acknowledgements	I
Table of Contents	III
List of Abbreviations	V
List of Papers.....	VII
Summary.....	IX
Introduction.....	1
Plant sexual reproduction.....	2
Female gametophyte development.....	2
Seed development	3
Endosperm cellularization.....	5
Female gametophyte maternal effects.....	7
Genomic imprinting	8
Evolutionary origin of genomic imprinting	9
Regulatory mechanisms of genomic imprinting	11
DNA methylation in genomic imprinting	12
Histone modification in genomic imprinting	14
RNA-directed DNA Methylation.....	15
Identification of imprinted genes	17
Genomic imprinting and hybridization	19
MADS-box transcription factors and genomic imprinting	21
Aim of Study.....	23
Results and Discussion.....	25
Characterization of the <i>AGL36</i> subclade of Type I MADS-box TFs	25
Imprinting of <i>AGL36</i> in hybrids.....	27
The post-zygotic hybrid barrier is affected by environmental and genetic factors.....	28
The genomic locus of the female gametophyte maternal effect mutant <i>capulet2</i>	29
Genetic analysis of <i>ANAPHASE PROMOTING COMPLEX 6</i>	32
<i>APC6</i> is a maternally expressed imprinted gene.....	33
Expression of the ANAPHASE PROMOTING COMPLEX/CYCLOSOME subunits is parentally regulated.....	35
Isolation of endosperm-specific nuclei allowed for imprinting analysis devoid of surrounding tissue contamination.....	36
Identification of parental allele-specific expression	37
Genomic imprinting is dynamically regulated throughout seed development	39

Table of Contents

Differentiated endosperm cells possess unique imprinting profiles	43
Limited overlap between studies in identified imprinted genes	45
Imprinting is often not regulated only by MET1	47
Concluding remarks and future perspectives	49
References.....	53
Papers I - III	71

List of Abbreviations

5mC	5-methylcytosine	ESF1	EMBRYO SURROUNDING FACTOR 1
<i>A. arenosa</i>	<i>Arabidopsis arenosa</i>		
<i>A. lyrata</i>	<i>Arabidopsis lyrata</i>	ESR	Embryo Surrounding Region
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>	FANS	Fluorescence-Activated Nuclear Sorting
ac	Antipodal cell	FC	Fold Change
AGL	AGAMOUS-LIKE	FG	Female Gametophyte
AGO	ARGONAUTE	FIE	FERTILIZATION INDEPENDENT ENDOSPERM
ALE1	ABNORMAL LEAF SHAPE 1	FIS	FERTILIZATION INDEPENDENT SEED
APC/C	ANAPHASE PROMOTING COMPLEX/CYCLOSOME	FWA	FLOWERING WAGENINGEN
BEG	Biparentally Expressed Gene	GFP	GREEN FLUORESCENT PROTEIN
bp	Base pair	GO	Gene Ontology
<i>bsl1</i>	<i>baseless 1</i>	H2A	Histone 2A
<i>cap2</i>	<i>capulet2</i>	H3K27me3	Histone 3 lysine 27 trimethylation
cc	Central cell	H3K9me2	Histone 3 lysine nine dimethylation
CMT	CHROMOMETHYLASE	<i>i.e.</i>	<i>id est</i> ; that is
Col	Columbia	IGF2	INSULIN-LIKE GROWTH FACTOR 2
DAL	Developing Aleurone Layer	InDel	Insertion/Deletion
DAP	Days After Pollination	INTACT	Isolation of Nuclei Tagged in specific Cell Types
DAPI	4,6-diamidino-2-phenylindole	IRP	Informative Read Pipeline
DCL3	DICER-LIKE	<i>Ler</i>	Landsberg <i>erecta</i>
DLK1	DELTA-LIKE HOMOLOGUE 1	m:p	Maternal:Paternal
DME	DEMETER	MADS-box	MINICHROMOSOME MAINTENANCE 1, AGAMOUS, DEFICIENS and SERUM RESPONSE FACTOR
DML	DEMETER-LIKE		
DNA	Deoxyribonucleic Acid	MEA	MEDEA
DNMT1	(CYTOSINE-5)-METHYLTRANSFERASE 1	MEG	Maternally Expressed Gene
DOGL4	DOG-LIKE 4	MET	DNA METHYLTRANSFERASE
DRM	DOMAINS REARRANGED METHYLTRANSFERASE	MPC	MATERNALLY EXPRESSED PAB C-TERMINAL
<i>e.g.</i>	<i>exempli gratia</i> ; for example	MSI	MULTICOPY SUPPRESSOR OF IRA
EBN	Endosperm Balance Number		
ec	Egg cell		
EDE1	ENDOSPERM DEFECTIVE 1		
EE	Early Endosperm		
ERF17	ERF/AP2 TRANSCRIPTION FACTOR 17		

List of Abbreviations

NRPD1	NUCLEAR RNA POLYMERASE D1	ROS1	REPRESSOR OF SILENCING 1
nt	Nucleotide	RT-PCR	Real-Time Polymerase Chain Reaction
PEG	Paternally Expressed Gene	sc	Synergid cell
PHE1	PHERES 1	siRNA	Small interfering RNA
Pol	RNA POLYMERASE	SN1	Streco1
PRC2	Polycomb Repressive Complex 2	SNP	Single Nucleotide Polymorphism
RBR1	RETINOBLASTOMA RELATED 1	SUVH	SU(VAR)3-9 HOMOLOG
RdDM	RNA-directed DNA Methylation	T-DNA	Transfer DNA
RDR	RNA DEPENDENT RNA POLYMERASE	TE	Transposable Element
RGE1/ZOU	RETARDED GROWTH OF EMBRYO 1/ZHOUP1	TE1	Total Endosperm
RNA	Ribonucleic Acid	TF	Transcription Factor
		Tsu	Tsushima
		UAS	Upstream Activation Sequence
		WT	Wild-Type

List of Papers

This thesis includes the following papers.

- Paper I **Genetic variation and temperature affects hybrid barriers during interspecific hybridization.** Katrine N. Bjerkan, Karina S. Hornslien, Ida M. Johannessen, Anders K. Krabberød, Yuri S. van Ekelenburg, Maryam Kalantarian, Reza Shirzadi, Luca Comai, Anne K. Brysting, Jonathan Bramsiepe, Paul E. Grini. *the Plant Journal* 2020. **101**(1):122–140
<https://doi.org/10.1111/tpj.14523>
- Paper II **Molecular Characterization of *capulet2* reveals the importance of *ANAPHASE PROMOTING COMPLEX 6* maternal expression in Endosperm Development.** Yuri S. van Ekelenburg, Ida V. Myking, Cathal Meehan, José Gutierrez-Marcos, Paul E. Grini. *Manuscript*
- Paper III **Spatial and Temporal Regulation of Parent-of-Origin Allelic Expression in the Endosperm.** Yuri S. van Ekelenburg, Karina S. Hornslien, Tom Van Hautegeem, Matyáš Fendrych, Gert Van Isterdael, Katrine N. Bjerkan, Jason R. Miller, Moritz K. Nowack, Paul E. Grini. *Submitted Manuscript; bioRxiv*
<https://doi.org/10.1101/2022.01.29.478178>

Summary

The endosperm of flowering plants is a triploid tissue consisting of two maternal and one paternal genomes. For this reason, tight control of gene regulation is required to maintain a balanced parental expression. The endosperm is the main site of genomic imprinting, parent-of-origin specific expression of alleles. Our understanding of the function and role of imprinting in plants is still emerging and, in this thesis, genomic imprinting was investigated in the model plant *Arabidopsis* using different approaches.

The *AGL36*-clade of the Type I MADS-box transcription factor family was thoroughly examined and several members were identified as imprinted genes. Interestingly, most imprinted members of the *AGL36*-clade did not display a seed phenotype when mutated. The imprinting of *AGL36* was conserved within investigated species of the genus *Arabidopsis* and in most hybrid crosses, but imprinting was specifically lost in a hybrid between the *Arabidopsis* species *arenosa* and *thaliana*. In hybrid crosses, species barriers can be erected in the endosperm, and we show that such endosperm barriers are influenced by both environmental and genetic factors.

The *capulet2* mutation was identified to be caused by a mutation in *APC6*, a subunit of the anaphase promoting complex/cyclosome (APC/C). The genomic locus was verified by transgene complementation of *cap2* and an independent *apc6* T-DNA insertion mutant. *APC6* was furthermore demonstrated to be a maternally expressed imprinted gene, resulting in endosperm cellularization defects in *apc6* alleles, suggesting a maternally biased role of the APC/C in the endosperm.

To investigate the temporal and spatial resolution of imprinting in the endosperm, endosperm nuclei were tagged with a green fluorescent protein (GFP) and isolated using fluorescence-activated nuclear sorting (FANS). This approach allowed us to identify the dynamic regulation of genomic imprinting in a temporal and spatial manner.

Collectively, the results in this thesis show that imprinting is more dynamic than previously thought and provide novel hypotheses and methodology for the elucidation of the role and function of genomic imprinting.

Introduction

The gametes of diploid organisms possess one copy of the paternal and maternal genome that, during fertilization, fuse to form the diploid embryo that develops into the next generation. Which allele is transmitted from the female and male parent is the decisive factor that determines the genotype of the offspring and this is the basis of classical Mendelian inheritance. The field of epigenetics has allowed for the finding of several types of non-Mendelian inheritance, including genomic imprinting. Genomic imprinting is an epigenetic phenomenon found in the nutrient-supplying tissue surrounding the embryo in both animals and plants, the placenta and endosperm respectively (Berger et al., 2006; Feil and Berger, 2007; Berger and Chaudhury, 2009; Nowack et al., 2010; Luo et al., 2011).

Misregulation of genomic imprinting has been shown to be responsible for several disorders in humans, including Angelman Syndrome (Mabb et al., 2011), Prader-Willi Syndrome (Nicholls et al., 1989), and male infertility (Rotondo et al., 2013). Imprinted genes have also been associated with economically important traits in mammalian livestock (Magee et al., 2014), and selection for these traits can be enhanced by fundamental knowledge of genomic imprinting. As such, genomic imprinting in mammals has been thoroughly studied to elucidate the mechanisms behind this phenomenon.

The societal relevance of the study of genomic imprinting in plants lies in the importance of the endosperm as a key source of food for the increasingly growing human population (Li and Berger, 2012). For centuries, humans have domesticated crops to select for favorable traits, such as seed size (Gepts, 2004). The understanding of the genetic and epigenetic regulation of genes responsible for modulating such traits could significantly enhance crop development. Statistical models have estimated the human population to reach a range from 7 to 11 billion in 2100 (Vollset et al., 2020). In order to maintain a sufficient food supply for an increasing population, it is vital that crop agriculture keeps developing to be more efficient. In this thesis, the main focus will be on genomic imprinting in plants as a fundamental basis for elucidating the understanding of this epigenetic phenomenon.

Plant sexual reproduction

The life cycle of plants involves an alternation between a haploid gametophyte and diploid sporophyte stage (Grossniklaus and Schneitz, 1998; Horst and Reski, 2016). Contrary to animal gametogenesis the haploid products of meiosis, *i.e.* the male microspore and female megaspore, undergo mitotic divisions to produce the gametes (Li and Ma, 2002). In the pollen grain, the microspore undergoes mitosis and differentiates into the vegetative cell and the generative cell. The larger vegetative cell exits the cell cycle and engulfs the generative cell, which then undergoes another round of mitosis to produce two sperm cells. The vegetative cell and the two sperm cells together form the male gametophyte (McCormick, 1993; Borg et al., 2009; Twell, 2011). In the ovule, the functional megaspore develops into the female gametophyte after three rounds of syncytial mitosis followed by cellularization. The female gametophyte, also called the embryo sac, consists of seven cells, two synergid cells, three antipodal cells, one egg cell, and the homodiploid central cell (Drews et al., 1998; Yadegari and Drews, 2004).

Angiosperms, such as *Arabidopsis thaliana* (*A. thaliana*), are characterized by a double fertilization event, which distinguishes them from gymnosperms in the seed-bearing clade (Russell, 1992; Baroux et al., 2002). During the double fertilization event, two male gametes, the sperm cells, fuse with two female gametes, the egg cell and central cell. The two sperm cells travel from the stigma through the pollen tube to the embryo sac where the egg cell and the central cell are fertilized by one sperm cell each (Hulskamp et al., 1995; Faure et al., 2002; Berger et al., 2008). The fertilized egg cell develops into the diploid embryo (1:1 maternal:paternal; 1m:1p), whereas the fertilized central cell develops the triploid endosperm (2m:1p). Together with the maternal sporophytic seed coat (2m:0p), the three distinct tissues form the seed, each composed of their own unique genotype (Nowack et al., 2010).

Female gametophyte development

The female gametophyte (FG) of *A. thaliana* develops according to the polygonum-type pattern and is formed by two consecutive processes, megasporogenesis and megagametogenesis (Reiser and Fischer, 1993). During megasporogenesis, the diploid megaspore mother cell undergoes meiosis that gives rise to four haploid megaspores, of which three degenerate (Bajon et al., 1999). The spatial location of each megaspore determines which one survives and in most angiosperms this is the megaspore closest to the chalazal region (Reiser and Fischer, 1993; Cheng et al., 2013). The surviving megaspore, called the functional megaspore, then

proceeds to megagametogenesis. Megagametogenesis (Figure 1) has been dissected into eight stages (FG1 - FG8), where FG1 coincides with the presence of the single functional megaspore and FG8 presents the final morphology of the three-celled female gametophyte (Christensen et al., 1997; Skinner and Sundaresan, 2018).

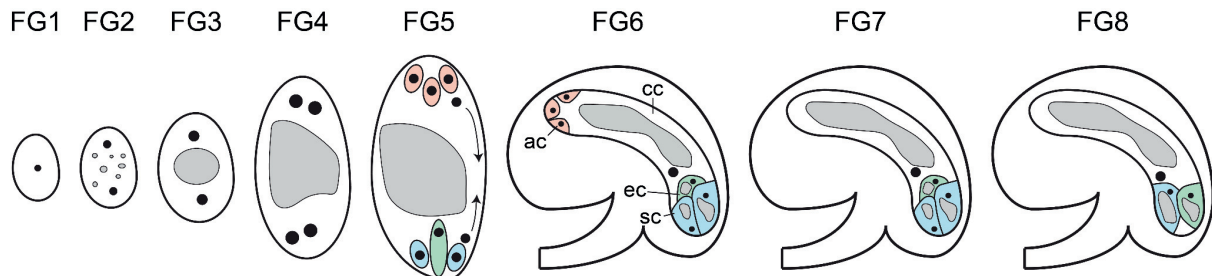


Figure 1: Illustration of gametogenesis of the female gametophyte in *Arabidopsis thaliana*. Megagametogenesis is initiated with the first mitotic division of the functional megaspore and the emergence of small vacuoles (gray) (FG2). The small vacuoles merge and form the central vacuole, separating the two megaspore daughter nuclei (FG3) that then undergo another round of mitosis forming a four-nucleate female gametophyte (FG4). Directly after the final round of mitosis (eight-nucleate female gametophyte), two nuclei (one from each polar end) migrate towards the center of the female gametophyte. Simultaneously, the remaining six nuclei start to differentiate and cellularize into three antipodal cells (orange), the egg cell (green), and two synergid cells (blue) (FG5). The two migrating nuclei fuse in the center and form the homodiploid nucleus of the central cell (white) (FG6). Around this stage, the antipodal cells are presumed to begin degeneration and the four-celled female gametophyte remains (FG7). Upon entry of the pollen tube, one of the synergid cells degenerates (FG8). ac: antipodal cells; cc: central cell; ec: egg cell; sc: synergid cell.

The synergid cells are crucial for the guidance of the pollen tube to the ovule and the release of sperm cells in the female gametophyte (Punwani and Drews, 2008; Kessler and Grossniklaus, 2011; Takeuchi and Higashiyama, 2012). Upon entry, one of the synergid cells interacts with the pollen tube, receiving its contents, after which this synergid cell degenerates (Sandaklie-Nikolova et al., 2007; Hamamura et al., 2011; Leydon et al., 2015). While the function of the egg cell, central cell, and synergid cells has become well understood, the function of the antipodal cells remains elusive, although a role has been proposed in the transfer of nutrients from the mother plant and for signaling between different compartments of the female gametophyte (Chettoor and Evans, 2015).

Seed development

After fertilization, there is a strong coordination and molecular signaling interaction between the endosperm, embryo, and seed coat (Nowack et al., 2010; Lafon-Placette and Köhler, 2014; Figueiredo et al., 2016; Ingram, 2020; Doll and Ingram, 2022). This is exemplified by several

endosperm-specific genes, such as *EMBRYO SURROUNDING FACTOR 1 (ESF1)*, *ABNORMAL LEAF SHAPE 1 (ALE1)*, and *RETARDED GROWTH OF EMBRYO 1 (RGE1)/ZHOUP1 (ZOU)*, which have been shown to regulate embryo development (Tanaka et al., 2001; Kondou et al., 2008; Yang et al., 2008; Costa et al., 2014). Furthermore, cellularization of the endosperm has been shown to be important for embryo development, as mutants of *FERTILIZATION INDEPENDENT SEED 2 (FIS2)* and *ENDOSPERM DEFECTIVE 1 (EDE1)* do not undergo endosperm cellularization resulting in arrested embryo development (Chaudhury et al., 1997; Pignocchi et al., 2009; Hehenberger et al., 2012; Lafon-Placette and Köhler, 2014).

Following the double fertilization event, the fertilized egg cell forms the zygote, and embryogenesis is initiated. An asymmetric division results in the precursor cells of the embryo proper and suspensor (Goldberg et al., 1994). Crucial for embryo proper development is correct morphologic pattern formation by additional asymmetric divisions up until the 32-cell stage, also known as the globular stage (Moukhtar et al., 2019; Armenta-Medina et al., 2021). At the same time, the suspensor differentiates, connects the embryo proper with the maternal tissue for nutrient transfer, and positions the embryo proper into the endosperm (Kawashima and Goldberg, 2010; Babu et al., 2013). The suspensor undergoes programmed cell death upon further maturation of the embryo proper (Bozhkov et al., 2005). At the two sides of the embryo, periclinal divisions result in the growth of the cotyledons and this marks the transition of the embryo from globular to heart stage (Boscá et al., 2011). The cotyledons keep growing until the maturation stage and the embryo enters dormancy.

While the embryo continuously develops, the endosperm has three distinct phases: the syncytial phase, the cellular phase, and the maturation phase (Berger et al., 2006). During the syncytial phase, endosperm nuclei divide without cell division (Brown et al., 2003; Li and Berger, 2012). The initial divisions are synchronous and subsequently regions are established as asynchronous nuclear division patterns develop (Figure 2a) (Brown et al., 2003; Boissard-Lorig et al., 2001; Picard et al., 2021). When the embryo reaches the globular stage, the cellular phase is initiated. Cellularization of the endosperm starts at the micropylar region and gradually progresses towards the chalazal endosperm (Brown et al., 1999; Boissard-Lorig et al., 2001; Hehenberger et al., 2012).

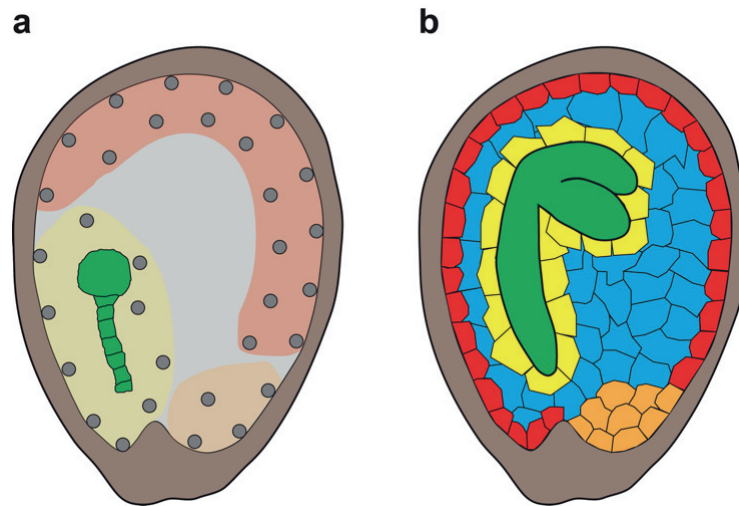


Figure 2: Illustration of seed compartments during the syncytial and cellularized stage of the endosperm. The seed consists of the seed coat (brown), (green), and the endosperm. **a)** At the syncytial stage, the endosperm nuclei (gray) form the micropylar (yellow), the peripheral (red) and the chalazal (orange) regions at the onset of asynchronous nuclear division. **b)** At the cellularized stage, the endosperm nuclei differentiate further into the embryo surrounding region (yellow), the developing aleurone layer (red), the chalazal endosperm (orange) and the central endosperm (blue).

Cell walls are formed in a periclinal manner until the endosperm is completely cellularized, except for the chalazal endosperm which never cellularizes (Berger, 2003; Olsen, 2004). The final maturation phase encompasses seed coat rupture by the embryo during germination and the transition from nutrients supplied by the mother plant to nutrient uptake by the embryo itself (Domínguez and Cejudo, 2014). Besides morphological differences, the syncytial and cellular phases also have distinct functions. The syncytial phase is characterized by the accumulation and storage of nutrients coming from the mother plant, and the cellular phase is characterized by a gradual degradation of the endosperm, which releases the previously stored nutrients as provision for the embryo (Nowack et al., 2010). This is often referred to as a sink-source switch, where the endosperm initially acts as a nutrient sink and eventually becomes the source of nutrients for the embryo until it is completely absorbed (Li and Berger, 2012; Lafon-Placette and Köhler, 2014).

Endosperm cellularization

During cellularization, the endosperm undergoes cell differentiation resulting in various subregions (hereafter referred to as domains) within the endosperm (Figure 2b) and it has become evident that gene expression is uniquely different for each domain (Brown et al., 1999; Boisnard-Lorig et al., 2001; Belmonte et al., 2013; Del Toro-De León and Köhler, 2019; Picard et al., 2021).

The chalazal endosperm, located directly at the interface with the maternal tissue, distinguishes itself from the rest of the endosperm as it remains in a syncytial state and is thought to be the entry point into the endosperm for nutrients coming from the mother plant (Nguyen et al., 2000; Olsen, 2004). Detailed microscopic analyses revealed two morphologically distinct types of tissue present in the chalazal endosperm, nodules and cysts (Nguyen et al., 2000; Olsen, 2004). The chalazal nodules are large multinucleate regions that localize at the border of the chalazal endosperm. The chalazal cysts are dense polyploid regions that localize adjacent to the maternal tissue (Baroux et al., 2004). Transcriptome analysis revealed that these two tissues, which are not separated by cell walls or cell membranes, have distinct expression profiles (Picard et al., 2021). Furthermore, the chalazal cyst showed enrichment for expression of genes involved in phloem sucrose unloading, consistent with the predicted function, *i.e.* nutrient transfer from the mother (Picard et al., 2021).

The embryo surrounding region (ESR) (Opsahl-Ferstad et al., 1997) consists of endosperm cells located close to the developing embryo and therefore has an important role in embryo-endosperm molecular communication (Yang et al., 2008; Doll et al., 2020). Additionally, the cuticle, a hydrophobic barrier between the embryo and endosperm that protects the embryo and later the seedling from water loss, is formed under the control of the ESR (Tanaka et al., 2001). After cellularization, and after degradation of the suspensor, cells of the ESR region start to undergo programmed cell death which releases the stored nutrients for uptake by the developing embryo (Denay et al., 2014; Van Hautegeem et al., 2015). Endosperm cells in the central endosperm that are adjacent to the ESR gradually become part of this domain as the embryo grows and requires supplemental nutrients (Van Hautegeem et al., 2015).

The developing aleurone layer (DAL) is the only endosperm region which does not undergo programmed cell death during embryo development (Domínguez and Cejudo, 2014; Van Hautegeem et al., 2015). In mature, dormant seeds, the aleurone layer has been shown to be crucial in maintaining dormancy (Debeaujon et al., 2000; Bethke et al., 2007).

Several genes have been identified to be important for endosperm cellularization, as mutants of these genes show disrupted cellularization. This includes *borgia* (Guitton et al., 2004), *agl62* (Kang et al., 2008), and members of the FIS-Polycomb Repressive Complex 2 (FIS-PRC2) complex (Grossniklaus et al., 1998; Luo et al., 1999; Ohad et al., 1999; Köhler et al., 2003a). Failed endosperm cellularization results in embryo abortion (Scott et al., 1998; Hehenberger et

al., 2012; Wolff et al., 2015). Interestingly, in interspecific crosses the establishment of a post-zygotic hybrid barrier shows a disrupted endosperm cellularization phenotype, connecting the endosperm as the origin of this barrier (Lafon-Placette et al., 2017). To identify disrupted cellularization patterning, endosperm markers have been used, such as the enhancer trap marker line KS117 (Ingouff et al., 2005) and members of the MADS-box TF family (Bemer et al., 2010). Altogether, endosperm cellularization is crucial for embryo development and hybridization and therefore understanding how this process is regulated is essential.

Female gametophyte maternal effects

Development of the embryo and endosperm is influenced by maternal effects, a phenomenon also observed in animals (Johnston et al., 1992; Colombo et al., 1997; Chaudhury and Berger, 2001; Baroux et al., 2008). Maternal effects refer to genetic and environmental factors the mother is subjected to that only influence the offspring. For example, the growth temperature of the mother plant influences the seed germination of the offspring, irrespective of the genotype (Nguyen et al., 2021). A special type of maternal effect, the gametophytic maternal effect, has been identified where genes are expressed in the egg cell and central cell, but the gene products (transcripts or proteins) only affect the zygote, endosperm, or both (Luo et al., 2014). Regarding the zygote, the extent of gametophytic maternal effects has been controversial. It has been proposed that in the early stages after fertilization, embryogenesis is under full maternal control through maternal transcripts deposited in the egg cell (Vielle-Calzada et al., 2000; Grimanelli et al., 2005). However, this has been largely contradicted by the finding of immediate *de novo* transcription of the paternal genome after fertilization (Weijers et al., 2001; Nodine and Bartel, 2012). How large the exact contribution of maternally deposited products is on embryogenesis remains to be determined. In the endosperm, however, gametophytic maternal effects could also be explained by genomic imprinting, but it remains challenging to determine the origin of transcripts directly after fertilization, *i.e.* if it is maternal carry-over or if it is *de novo* transcription (Evans and Kermicle, 2001).

Attempts to investigate female gametophyte maternal effects have been made using mutants that in a heterozygous state progress normally through female gametophyte development and double fertilization, but display a seed mutant phenotype. This has identified various gametophyte maternal effect mutants in *A. thaliana* that show defects in embryo development (Ray et al., 1996; Li et al., 2021) and mutants that show defects in both embryo and endosperm

development (Grini et al., 2002; Pagnussat et al., 2005). Furthermore, female gametophyte maternal effect mutants have also been detected in maize where distorted development was observed in only the endosperm (Gavazzi et al., 1997; Gutiérrez-Marcos et al., 2006), or in both the endosperm and the embryo (Evans and Kermicle, 2001), and in rice with defective endosperm development (Liu et al., 2018b). In most of these mutants, except *baseless 1 (bs11)* (Gutiérrez-Marcos et al., 2006), female gametophyte development showed no disruption, and defects in endosperm development were observed only after fertilization.

Heterozygous individuals from this class of mutants do not follow traditional Mendelian segregation (3:1) since half of the offspring will be affected (1:1) and homozygous offspring are rarely detected (Yadegari and Drews, 2004). This feature can significantly assist in the identification of female gametophyte maternal effect mutants using linkage-based screening, where segregation of linked markers (using visible mutant markers or antibiotic resistance markers) is affected (Feldmann et al., 1997; Bonhomme et al., 1998; Howden et al., 1998; Grini et al., 1999).

Genomic imprinting

Since the endosperm has an important function in seed development and an unbalanced parental allelic contribution, gene expression of each parental allele must be tightly regulated (Barton et al., 1984; Birchler, 1993). Genomic imprinting has been shown to regulate gene expression in a parent-of-origin specific manner. This is achieved by epigenetic modification of an allele based on its parental origin through DNA methylation or modification of histone tails established in the gametes (Rodrigues and Zilberman, 2015; Montgomery and Berger, 2021). Although genomic imprinting is classically described as epigenetic silencing of parental alleles, more evidence is emerging that most imprinted genes are not exclusively expressed from one allele but show preferential parental biased expression (Gehring et al., 2011; Hsieh et al., 2011; Satyaki and Gehring, 2017; Del Toro-De León and Köhler, 2019; Hornslien et al., 2019; Picard et al., 2021).

Over the last two decades, a large number of imprinted genes have been identified in various species, but imprinting has been primarily limited to flowering plants and mammals. The endosperm is analogous to the mammalian placenta, and it is postulated that imprinting has evolved convergently in species with a placental habit (Feil and Berger, 2007; Pires and

Grossniklaus, 2014). In mammals, imprinted genes are commonly involved in growth control, where paternally expressed genes (PEGs) stimulate growth and maternally expressed genes (MEGs) repress growth (Leighton et al., 1995). In plants, the function of imprinted genes remains largely elusive (Batista and Köhler, 2020). While imprinting in plants has been shown to be primarily confined to the endosperm, some genes have been found that display parent-of-origin preferential gene expression in the embryo of maize (Jahnke and Scholten, 2009; Waters et al., 2011; Zhang et al., 2011) and rice (Luo et al., 2011). Also in *A. thaliana*, various studies have suggested parent-of-origin specific gene expression in the embryo (Nodine and Bartel, 2012; Raissig et al., 2013; Pignatta et al., 2014; Alonso-Peral et al., 2017). However, genomic imprinting in the embryo remains controversial, as profound contamination from other tissues has been reported in embryo transcriptome datasets (Schon and Nodine, 2017).

In plants, the functional importance of genomic imprinting is not readily evident from mutant studies, as most imprinted gene mutants are not affected in seed development (Masiero et al., 2011; Shirzadi et al., 2011; Berger et al., 2012; Wolff et al., 2015). The lack of distorted seed phenotypes makes it challenging to study the function and role of genomic imprinting. Additionally, imprinted genes are difficult to distinguish from gametophyte parental effect genes as it is difficult to determine the origin of transcripts. Even though parental effect genes and imprinted genes appear to have similar outcomes, they are distinct genetic phenomena (Wolf and Wade, 2009). Parental effect genes are expressed in the gametes and the transcripts affect endosperm and embryo development post-fertilization, whereas imprinted genes result in parent-of-origin specific *de novo* expression after fertilization. Furthermore, imprinting is not always conserved between or within species (Wolff et al., 2011; Waters et al., 2013; Klosinska et al., 2016; Gehring and Satyaki, 2017; Pignatta et al., 2018; Hornslien et al., 2019), complicating the study to elucidate the evolutionary origin of genomic imprinting.

Evolutionary origin of genomic imprinting

Several theories have been raised to explain the evolutionary origin and function of genomic imprinting (Montgomery and Berger, 2021), and each of these hypotheses can identify one or more supporting examples of imprinted genes. This could suggest that imprinting of different genes has different selective pressures and different evolutionary origins (Patten et al., 2014; Rodrigues and Zilberman, 2015; Sazhenova and Lebedev, 2021).

The most distributed theory is the parental conflict or kinship hypothesis (Haig and Westoby, 1989; Haig, 2014), which proposes that parental alleles have different interests regarding their offspring. PEGs often encode proteins that stimulate growth, whereas MEGs encode proteins that limit nutrient allocation to the individual offspring, allowing reservation of nutrients for future offspring as well as of offspring siblings. The paternal allele aims for the survival of its single offspring, whereas the maternal allele wants to ensure that all the offspring of the mother plant survive. There are some differences between mammals and flowering plants regarding the application of the parental conflict theory. Although imprinting in mammals is mostly observed in the placental tissue, it is hypothesized that embryonic gene expression is the driving force for maternal nutrient acquisition in mammals (Moore and Haig, 1991), while the endosperm is mainly responsible for this in plants. Furthermore, the parental conflict theory is best applied when genes are expressed monoallelically, which is often observed in mammals (Babak et al., 2015; Hanna, 2020), but in contrast with observations in plants, where most imprinted genes show parentally biased expression (Hornslien et al., 2019).

A different perspective on genomic imprinting is provided by the gene dosage theory, which states that imprinting is merely a mechanism to regulate gene dosage levels (Dilkes and Comai, 2004; Ferguson-Smith, 2011). While the parental conflict theory heavily relies on natural selection, the gene dosage theory explains imprinting to be a more dynamically controlled regulatory mechanism, allowing different gene dosage levels to be required at different timepoints or tissues. In mammals, this theory is supported by the finding of timepoint-specific and tissue-specific dynamic regulation of imprinting of *DELTA-LIKE HOMOLOGUE 1* (*DLK1*) and *INSULIN-LIKE GROWTH FACTOR2* (*IGF2*) respectively (DeChiara et al., 1991; Ferrón et al., 2011). In plants, it has become apparent that most imprinted genes are not exclusively expressed from one allele but often show preferential parental expression favoring the explanation that gene dosage is regulated (Hornslien et al., 2019). Furthermore, in the triploid endosperm, maintenance of a correct balance in parental gene dosage levels has been shown to be important for seed development (Birchler, 1993; Leblanc et al., 2002; Lafon-Placette et al., 2017). In contrast, inducing imbalanced parental expression by generating triploid *Arabidopsis* embryos that mimic the endosperm genotype (2m:1p), did not evolve imprinting of these genes (Fort et al., 2017), indicating that tight parental dosage is less important for the embryo than for the endosperm and that regulation of balanced parental gene dosage levels might be tissue-specific.

Contrasting the gene dosage theory and parental conflict theory, which propose imprinting to have a developmental advantage, the transposon defense theory (Barlow, 1993; McDonald et al., 2005) explains imprinting as a side-effect from the silencing of transposable elements (TEs) and not directly related to natural selection (Chan et al., 2005; Gehring et al., 2009). As transposons insert themselves into new locations in the genome, DNA methylation pathways are activated to silence these elements. A side-effect of this form of TE silencing is that closely-located genes or gene regulatory sequences also become methylated. Interestingly, many imprinted genes are known to be flanked by TEs providing strong evidence that their imprinting pattern is caused by this mechanism (Gehring et al., 2009; Hsieh et al., 2009; Batista et al., 2019). However, imprinted genes that do not have TEs proximally located are not readily explained by this theory.

Regulatory mechanisms of genomic imprinting

In eukaryotic cells, the DNA is wrapped around a histone octamer, forming the unit of the nucleosome. The histones are composed of two copies of histone 2A (H2A), H2B, which form heterodimers, H3 and H4, which form homodimers, and together they form the octamer (McGinty and Tan, 2015). The N-terminus of histone proteins (histone tails), situated outside the nucleosome core particle, is accessible for post-translational modifications. Histone tail acetylation is associated with gene expression and histone tail methylation can result in both an active and repressive state of gene expression (Pecinka et al., 2020). The nucleosomal DNA itself is also susceptible to epigenetic modifications, which are associated with gene silencing in the canonical form of DNA methylation (Newell-Price et al., 2000). Both DNA methylation and histone modifications contribute to nucleosome dynamics that modulate the accessibility of genes for transcription. Because of their contribution to this process, both DNA methylation and histone tail methylation (Figure 3), have been linked to genomic imprinting (Law and Jacobsen, 2010; Batista and Köhler, 2020).

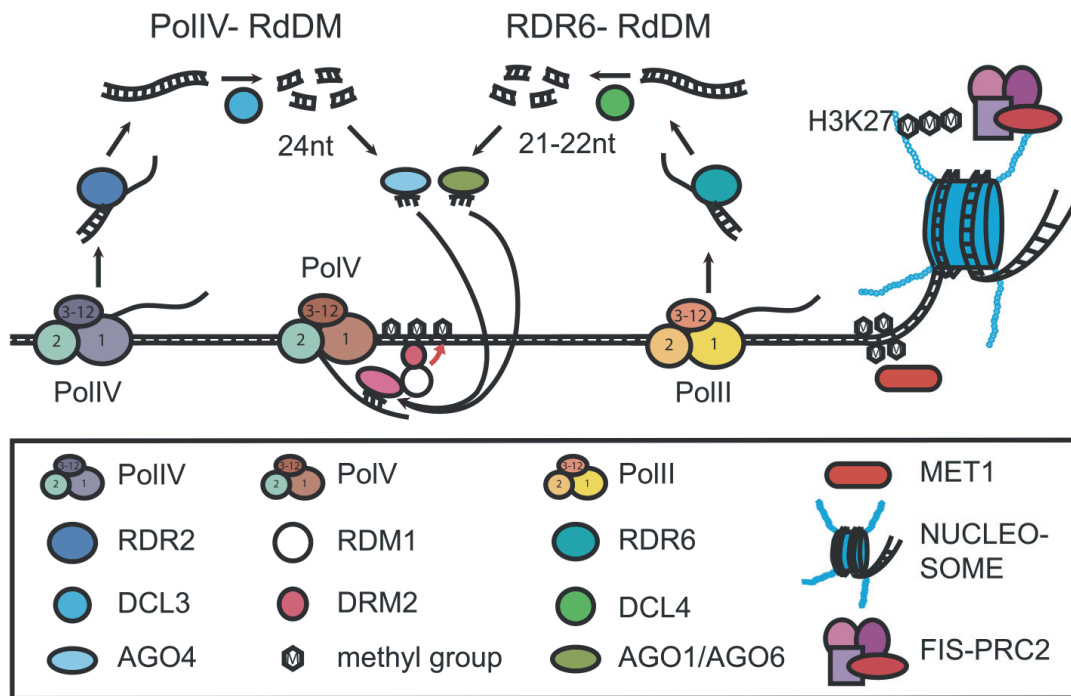


Figure 3: Illustration of the epigenetic mechanisms to regulate genomic imprinting. RNA-directed DNA Methylation (RdDM) utilizes small RNAs to guide *de novo* methylation of DNA in all sequence contexts by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2). The origin of these small RNAs differs between canonical (RNA POLYMERASE IV; PolIV) and non-canonical (RNA DEPENDENT RNA POLYMERASE; RDR6) RdDM. DNA METHYLTRANSFERASE 1 (MET1) is primarily responsible for maintaining CG methylation and the FIS-PRC2 complex mediates histone 3 lysine 27 trimethylation (H3K27me3). Figure adapted from Hornslien et al., 2019.

DNA methylation in genomic imprinting

DNA methylation comprises two different classes, *de novo* and maintenance methylation, and is performed by DNA methyltransferases. Methylation profiles are maintained on the newly replicated daughter strand based on hemimethylated profiles on the template strand. Several DNA methyltransferases have been described in *A. thaliana*, of which some have been shown to act partially redundant (Zhang and Jacobsen, 2006). In most cases, methylation marks are established on cytosine, producing 5-methylcytosine (5mC), and the sequence context determines which cytosines are targeted by the different DNA methyltransferases (Feng et al., 2010). In plants, DNA METHYLTRANSFERASE1 (MET1), the plant homolog to DNA (CYTOSINE-5)-METHYLTRANSFERASE 1 (DNMT1) of mammals, is the main maintenance DNA methyltransferase responsible for maintaining methylation of CG sites and is often linked to imprinting (Finnegan and Dennis, 1993; Law and Jacobsen, 2010). In *A. thaliana*, three *MET1* paralogs, *MET2a*, *MET2b*, and *MET3*, have been identified, of which *MET2a* and *MET2b* show expression in the central cell but their function remains largely

unknown (Jullien et al., 2012). Recently, it has been shown that *met3* mutants only slightly affect the seed transcriptome (Tirot et al., 2022).

However, MET1 is active almost ubiquitously throughout the plant life cycle and consequently, maternally expressed imprinted genes are by default methylated and silenced (Gehring and Satyaki, 2017). For these genes to be expressed in a parent-of-origin specific manner, active demethylation of the DNA needs to occur. This can be achieved by DNA glycosylase enzymes that remove and replace a methylated cytosine with an unmethylated cytosine through the base excision repair system (Lindahl, 1986). In the central cell, the DNA glycosylase DEMETER (DME) mediates the demethylation of the maternal allele (Choi et al., 2002). Furthermore, passive demethylation is mediated through reduced DNA methylation by MET1. The expression of *MET1* is repressed by a complex of MULTICOPY SUPPRESSOR OF IRA 1 (MSI1) and RETINOBLASTOMA RELATED 1 (RBR1) (Jullien et al., 2008), and the genes for both proteins are highly expressed in the central cell (Köhler et al., 2003a; Ingouff et al., 2006). *DME* is mainly expressed in the central cell, making it specific for demethylation of the maternal allele (Schoft et al., 2011; Ibarra et al., 2012). It is thought that MEGs are mainly regulated by the joint action of DME and MET1, *i.e.* maintenance of methylation of the paternal allele by MET1 in the sperm cell and demethylation of the maternal allele by DME in the central cell, and several imprinted genes have been shown to be regulated by DME/MET1, including *FLOWERING WAGENINGEN (FWA)* (Kinoshita et al., 2004) and *FIS2* (Jullien et al., 2006). More recently, however, it has been shown that *met1* mutants do not sufficiently restore the paternal expression of most MEGs (Hornslien et al., 2019), questioning the universal application of the hypothesis.

In addition to the MET DNA methyltransferases, plants possess a class of plant-specific DNA methyltransferase (chromomethylases; CMT) (Stroud et al., 2014), of which CMT3 is responsible for maintaining DNA methylation at CHG (where H is A, C, or T) sites (Lindroth et al., 2001) and CMT2 *de novo* methylates DNA in a CHH sequence context (Zemach et al., 2013). Although they are responsible for DNA methylation, CMT2 has not yet been investigated in connection to genomic imprinting and CMT3 has, when investigated, not been found to be involved in genomic imprinting (Shirzadi et al., 2011; Moreno-Romero et al., 2019). A third chromomethylase, CMT1, has been found in the *A. thaliana* genome but is predicted to be nonessential, as inactive loci have been detected in various *A. thaliana* ecotypes (Henikoff and Comai, 1998). An interesting and novel role for CMT3 to establish *de novo*

methylation that can be maintained by MET1 is emerging (Wendte et al., 2019; Papareddy et al., 2021). The *CMT3*-lacking angiosperm species *Eutrema salsugineum* was transformed with the *CMT3* gene of *A. thaliana* (*AtCMT3*), which was able to establish *de novo* DNA methylation on gene bodies specifically of *A. thaliana* orthologs (Wendte et al., 2019). While CHG methylation by CMT3 is guided by histone H3 lysine nine dimethylation (H3K9me2) (Bernatavichute et al., 2008; Du et al., 2012; Stroud et al., 2014), *de novo* methylation by CMT3 showed no prior DNA or histone methylation marks (Wendte et al., 2019). The importance of CMT3-directed *de novo* methylation of gene bodies in regards to genomic imprinting is yet to be determined.

Histone modification in genomic imprinting

The tails of histones are susceptible to epigenetic modifications that dynamically alter the state of chromatin and the main modification correlated to genomic imprinting is histone tail methylation. This is generally considered the main mechanism to silence the maternal allele of PEGs. The histone tail methylation is performed by members of the PRC2 family, a complex consisting of four core subunits that are highly conserved between animals and plants (Köhler and Hennig, 2010). PRC2 catalyzes the trimethylation of histone 3 at the amino acid lysine 27 position (H3K27me3), commonly associated as a repressive mark of gene expression (Margueron and Reinberg, 2011). In plants, it has been shown that PRC2 can be recruited by various transcription factor families to specific loci on the genome (Xiao et al., 2017; Zhou et al., 2018). Plants contain multiple genes that encode partially redundant subunits which can combine in various configurations to target different sets of genes (Chanvivattana et al., 2004; Makarevich et al., 2006). The nucleosome binding subunit FIS2 forms a protein complex that is known as FIS-PRC2 (Luo et al., 1999; Köhler and Hennig, 2010). The methylation of histones in FIS-PRC2 is mediated by the H3K27me3 methyltransferase subunit MEDEA (MEA) (Grossniklaus et al., 1998; Kiyosue et al., 1999; Chanvivattana et al., 2004), which has been shown to be dependent of FERTILIZATION INDEPENDENT ENDOSPERM (FIE), a subunit present in all PRC2 complexes (Ohad et al., 1999). The last core subunit consists of one of five WD40-containing nucleosome remodeling proteins: MSI1-5 (Köhler et al., 2003a).

The FIS-PRC2 complex, a combination of MEA, FIS2, FIE, and MSI1 (Spillane et al., 2000; Yadegari et al., 2000; Köhler et al., 2003a), has been shown to silence the maternal allele of paternally expressed genes in the central cell and as such, is involved in the regulation of

genomic imprinting (Gehring, 2013). Interestingly, genes encoding for several PRC2 subunits, including *FIS2*, *FIE*, and *MEA*, are imprinted themselves (Kinoshita et al., 1999; Luo et al., 2000). It is therefore not surprising that mutants of PRC2 subunits show distinct phenotypes in the female gametophyte and the developing seed (Grossniklaus et al., 1998; Luo et al., 2000; Spillane et al., 2000; Vinkenoog et al., 2000; Hehenberger et al., 2012). The major example of a paternally expressed gene for which imprinting of the maternal allele is regulated by FIS-PRC2, is *PHERES1* (*PHE1*), as mutants of FIS-PRC2 subunits show reactivation of the *PHE1* maternal allele (Köhler et al., 2005). The current hypothesis for the imprinting of PEGs is that the DNA of the maternal allele is hypomethylated and can therefore be targeted by FIS-PRC2 for trimethylation of H3K27. At the same time, the DNA of the paternal allele is hypermethylated and thus avoids receiving the silencing H3K27me3 mark (Rodrigues and Zilberman, 2015; Moreno-Romero et al., 2016). This dual mechanism of DNA methylation and histone methylation has been shown to regulate the imprinting of *PHE1* (Makarevich et al., 2008). While mammals globally remove histone epigenetic marks in the sperm (Braun, 2001), this is not the case for plant sperm cells (Borg and Berger, 2015). Similarly, DNA methylation marks are not erased from the entire plant genome (Calarco et al., 2012), but they are completely erased in mammalian sperm (Reik et al., 2001). Interestingly, it has become apparent that histones are specifically targeted for reprogramming of epigenetic marks in the male gametophyte (Borg et al., 2020). This process could elaborate more on how gene-specific histone-mediated silencing is alleviated by active histone reprogramming in the male gamete.

RNA-directed DNA Methylation

The RNA-directed DNA Methylation (RdDM) pathway is an elaborate plant-specific mechanism to *de novo* methylate DNA. The role of RdDM in various regulatory aspects has been well described (Matzke and Mosher, 2014; Borges and Martienssen, 2015; Cuerda-Gil and Slotkin, 2016; Wendte and Pikaard, 2017; Erdmann and Picard, 2020; Petrella et al., 2021), including genomic imprinting, and more genes are being identified to be imprinted by this pathway (Vu et al., 2013; Hornslien et al., 2019; Iwasaki et al., 2019; Kirkbride et al., 2019; Long et al., 2021).

Site-specific *de novo* DNA methylation, guided by small RNAs and irrespective of DNA sequence context, can be established by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Cao et al., 2000; Matzke and Mosher, 2014; Stroud et

al., 2014; Erdmann and Picard, 2020; Harris and Zemach, 2020). The *A. thaliana* genome harbors two *DRM2* homologs, *DRM1* and *DRM3*, of which *DRM1* shares a high similarity with *DRM2* and they are potentially acting redundantly (Cao et al., 2000; Cao and Jacobsen, 2002). Interestingly, *DRM1* is only expressed in the egg cell, whereas *DRM2* is also expressed in the central cell suggesting that *DRM1* does not participate in establishing imprinting in the endosperm (Jullien et al., 2012). Furthermore, *DRM3* has been shown to only moderately affect DNA methylation (Henderson et al., 2010; Costa-Nunes et al., 2014; Zhong et al., 2015).

Based on the origin of the small RNAs, RdDM can be classified into two main pathways, referred to as canonical and non-canonical RdDM (Erdmann and Picard, 2020). In canonical RdDM, RNA polymerase IV (PolIV) transcribes small interfering RNA (siRNA) precursors. After transcription, the siRNA precursors are processed by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) into double-stranded RNA (Law et al., 2011; Haag et al., 2012) followed by cleavage into 24-nucleotide (nt) siRNA-duplexes by DICER-LIKE 3 (DCL3) (Qi et al., 2005; Blevins et al., 2015). One of the siRNA strands is then loaded into ARGONAUTE (AGO) proteins of the AGO4-clade (AGO4, AGO6, and AGO9) (Qi et al., 2006; Wierzbicki et al., 2009; Havecker et al., 2010).

In non-canonical RdDM, siRNAs of different sizes (21, 22, or 24 nt) and origins than the canonical RdDM pathway are included (Cuerda-Gil and Slotkin, 2016). This includes the RDR6-RdDM pathway where RNA polymerase II (PolII) transcripts are made double-stranded by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6). These duplex siRNA precursors can be cleaved by DICER-LIKE 2/4 (DCL2/4) to produce 21-22 nt siRNAs that are loaded into AGO4 or AGO6 (Allen et al., 2005; Nuthikattu et al., 2013). Furthermore, a DICER-independent pathway has been described (Yang et al., 2016; Ye et al., 2016), and other non-canonical RdDM pathways have been suggested but are not well characterized. It is likely that siRNAs can also originate from other, currently unknown, mechanisms (Cuerda-Gil and Slotkin, 2016).

While the origin of siRNAs differs for both RdDM pathways, they are guided to the genome similarly. RNA polymerase V (PolV) is recruited to methylated DNA by SU(VAR)3-9 homologs 2 and 9 (SUVH2 and SUVH9 respectively) (Johnson et al., 2014; Liu et al., 2014) or to unmethylated DNA through AGO proteins guided by siRNAs (Sigman et al., 2021). The PolV-produced RNAs are proposed to act as scaffolds for the canonical and non-canonical

generated siRNA-AGO complexes (Wierzbicki et al., 2008; Böhmendorfer et al., 2016), which in turn recruit DRM2 to methylate the DNA on-site (Cao and Jacobsen, 2002; Zhong et al., 2014). Interestingly, AGO4 has also been proposed to mediate *de novo* methylation through siRNA-independent recruitment of RdDM components (Gallego-Bartolomé et al., 2019).

The RdDM pathway has mainly been shown to mediate silencing of TEs (Huettel et al., 2006; Mirouze et al., 2009; Nuthikattu et al., 2013; Zemach et al., 2013; Sigman and Slotkin, 2016; Erdmann and Picard, 2020), however, parent-of-origin specific small RNA profiles have suggested a direct link between RdDM and genomic imprinting (Mosher et al., 2009; Pignatta et al., 2014; Erdmann et al., 2017). Indeed, mutational analysis of members of the RdDM pathway indicated that RdDM silenced the paternal allele of two maternally expressed genes (Vu et al., 2013; Hornslien et al., 2019), and in a different study, the maternally expressed gene *AGL36* was shown to be upregulated, suggesting a role of RdDM in regulating this imprinted locus (Lu et al., 2012). Since then, numerous instances of RdDM-mediated imprinting of parental alleles have emerged (Iwasaki et al., 2019; Kirkbride et al., 2019). Furthermore, the antagonistic interaction between active demethylation by REPRESSOR OF SILENCING 1 (ROS1) and DNA methylation by the RdDM pathway has been shown, where ROS1 is suggested to prohibit the spread of DNA methylation from methylated TEs (Tang et al., 2016). Moreover, ROS1 has been shown to regulate demethylation of the paternal allele of *DOG-LIKE 4 (DOGL4)* as protection against hypermethylation and imprinting of the paternal allele (Zhu et al., 2018). These findings suggest that RdDM-mediated DNA methylation and ROS1-mediated demethylation dynamically regulate parent-of-origin specific gene expression.

Identification of imprinted genes

The first identified imprinted gene in flowering plants, *MEA*, was initially found as a female gametophyte maternal effect mutant where embryo and endosperm development was affected (Grossniklaus et al., 1998) and in a later study, *MEA* was found to be imprinted (Vielle-Calzada et al., 1999). As *MEA* was shown to be imprinted, other core components of the FIS-PRC2 complex were investigated, identifying *FIS2* (Luo et al., 2000; Jullien et al., 2006) as a maternally expressed imprinted gene. Furthermore, real-time polymerase chain reaction (RT-PCR) analysis of genes regulated by the FIS-PRC2 complex identified *PHE1* as the first paternally expressed imprinted gene (Köhler et al., 2003b; Köhler et al., 2005). Furthermore, analysis of mutants with a flower phenotype for parent-of-origin specific expression identified

FWA as an imprinted gene (Kinoshita et al., 2004). Then, the focus shifted towards the identification of imprinted genes by microarray analysis of crosses with misregulated parental genome balances, which identified *MATERNALLY EXPRESSED PAB C-TERMINAL (MPC)* (Tiwari et al., 2008) and *AGL36* (Shirzadi et al., 2011) as novel maternally expressed genes.

The advent of high throughput sequencing has enhanced the identification of imprinted genes in many species. This has resulted in a large number of parent-of-origin transcriptome analyses in *A. thaliana* (Gehring et al., 2011; Hsieh et al., 2011; Shirzadi et al., 2011; Wolff et al., 2011; Pignatta et al., 2014; Del Toro-De León and Köhler, 2019; Picard et al., 2021) and other species including *A. lyrata* (Klosinska et al., 2016), *Brassica napus* (Liu et al., 2018a; Rong et al., 2021), *Brassica rapa* (Yoshida et al., 2018), *Capsella rubella* (Hatorangan et al., 2016; Lafon-Placette et al., 2018), maize (Jahnke and Scholten, 2009; Waters et al., 2011; Zhang et al., 2011; Waters et al., 2013; Zhang et al., 2014), rice (Luo et al., 2011; Rodrigues et al., 2013; Yuan et al., 2017; Yang et al., 2020), wheat (Yang et al., 2018), sorghum (Zhang et al., 2016), and tomato (Florez-Rueda et al., 2016; Roth et al., 2018; Florez-Rueda et al., 2021).

However, the overlap in identified imprinted genes is rather scant between species, (Waters et al., 2013; Hatorangan et al., 2016; Klosinska et al., 2016; Gehring and Satyaki, 2017) and also between accessions of the same species (Wolff et al., 2011; Pignatta et al., 2018; Hornslien et al., 2019), questioning the conservation of genomic imprinting (Schon and Nodine, 2017; Hornslien et al., 2019). Furthermore, the number of identified MEGs and PEGs remains disputed due to the detection of widespread contamination from the surrounding tissues (Schon and Nodine, 2017). In several endosperm-specific transcriptome datasets of genomic imprinting, substantial seed coat contamination was detected. Due to the genetic nature of the seed coat, *i.e.* it is diploid and of maternal origin, contamination from this tissue significantly biases expression values towards maternal origin and therefore false positive MEGs are prone to emerge. Also, an important aspect that is often not addressed, is the variability in experimental conditions between studies, such as accession and seed developmental stage, which have shown to influence the identification of imprinted genes (Wolff et al., 2011; Xin et al., 2013; Pignatta et al., 2014; Pignatta et al., 2018).

To overcome potential contamination from other tissues in the seed, various strategies have been developed to perform endosperm-specific transcriptome analysis devoid of seed coat and embryo contamination. The isolation of nuclei tagged in specific cell types (INTACT) method

allows extraction of tissue-specific nuclei via biotin tagging (Deal and Henikoff, 2011). This has been applied to isolate endosperm-specific nuclei and subsequently used for allele-specific expression analysis (Del Toro-De León and Köhler, 2019). In a different approach, 4,6-diamidino-2-phenylindole (DAPI)-stained endosperm-specific nuclei were isolated, based on ploidy, by Fluorescence-Activated Nuclear Sorting (FANS) for endosperm transcriptome analyses (Picard et al., 2021). Furthermore, FANS-sorted endosperm nuclei were analyzed for allele-specific expression to identify genomic imprinting without contamination, showing that parent-of-origin specific expression can be specific to endosperm regions (Picard et al., 2021).

Genomic imprinting and hybridization

In hybridization, two genetically distinct but closely related species reproduce to generate a hybrid, potentially resulting in the generation of a new species. Natural selection can generate a new hybrid species that has higher fitness than either parent and this can lead to the parent species becoming extinct (Rieseberg and Willis, 2007; Coughlan and Matute, 2020). Species have evolved various strategies in order to prevent hybridization and maintain species integrity. One such measure is the erection of a seed-based post-zygotic hybrid barrier, where a hybrid fertilization product is formed but is not viable, due to distorted endosperm development (Lafon-Placette et al., 2017; Coughlan and Matute, 2020).

When two individuals of the same species but with different ploidies (*e.g.* diploid and tetraploid) hybridize, a post-zygotic barrier known as the triploid block is established, leading to distorted endosperm development and eventually seed abortion (Scott et al., 1998; Köhler et al., 2021). In crosses with a higher ploidy species as the male parent (known as paternal excess), the endosperm often shows a delay in endosperm cellularization. In contrast, when the female parent is of a higher ploidy species (known as maternal excess), the endosperm often shows precocious endosperm cellularization (Scott et al., 1998; Sekine et al., 2013). This has led to the theory that the endosperm has an effective ploidy that is species-specific and irrespective of its actual ploidy, called the endosperm balance number (EBN) (Johnston et al., 1980).

Also, during interspecific hybridization, a post-zygotic barrier can erect, resulting in defects in endosperm development (Burkart-Waco et al., 2012; Rebernig et al., 2015; Florez-Rueda et al., 2016; Lafon-Placette et al., 2017; Tonosaki et al., 2018; Wang et al., 2018). This is demonstrated by interspecific hybridization of *A. lyrata* and *A. arenosa* where crosses of

individuals with a similar ploidy resulted in disrupted endosperm cellularization (Muir et al., 2015). Interestingly, increasing the ploidy of *A. lyrata* was sufficient to bypass this barrier, suggesting that the post-zygotic barrier between *A. arenosa* and *A. lyrata* is due to a different EBN for these species (Lafon-Placette et al., 2017).

A putative role of genomic imprinting has been suggested in the erection of interspecific post-zygotic hybrid barriers (reviewed in Köhler et al., 2021). As the endosperm consists of a delicately balanced parental dosage (2m:1p), interploidy hybridization disrupts this balance resulting in misregulated parent-of-origin specific gene expression. Furthermore, interspecific hybrids often show disrupted imprinting profiles, presumably due to different species having their own regulation of genomic imprinting to maintain the correct EBN. Parental imbalance is generated when species hybridize with different imprinting patterns and several genes have been shown to switch parental expression (Josefsson et al., 2006; Burkart-Waco et al., 2015). Interestingly, inviable hybrid crosses display similar endosperm phenotypes as maternal or paternal excess, *i.e.* too early or too late endosperm cellularization respectively, highlighting the correlation between the EBN and genomic imprinting (Florez-Rueda et al., 2016; Lafon-Placette et al., 2017; Roth et al., 2019). This is supported by the finding that mutants in various imprinted genes, *fis2* (Chaudhury et al., 1997), *fie* (Vinkenoog et al., 2000), and *mea* (Grossniklaus et al., 1998), displayed delayed endosperm cellularization phenotypes. Additionally, the dosage-specific expression of various PEGs has been shown to establish the triploid block (Walia et al., 2009; Kradolfer et al., 2013; Wolff et al., 2015; Huang et al., 2017; Lafon-Placette et al., 2018). Also, the triploid block was found to be suppressed by mutants in the RdDM pathway, presumably by disruption of imprinting patterns (Wang et al., 2021). The findings described above reflect the significant importance of genomic imprinting in the establishment of the post-zygotic hybrid barrier. However, this is also challenged by the finding that expression profiles were widely shared between the lethal paternal excess cross (diploid x tetraploid) and by a paternal tetraploid RdDM mutant, *nuclear RNA polymerase D1 (nrpd1)*, that can alleviate lethality (diploid x tetraploid *nrpd1*) (Erdmann et al., 2017; Satyaki and Gehring, 2019). This suggests that paternal repression of most genes is not regulated by the RdDM pathway and only a few genes are responsible for alleviating paternal excess lethality.

MADS-box transcription factors and genomic imprinting

The family of MADS-box transcription factors (TF) has been linked to genomic imprinting (Masiero et al., 2011; Yoshida and Kawabe, 2013) and the establishment of a post-zygotic hybrid barrier (Walia et al., 2009). The MADS-box TF family is named after the conserved sequence domain found in *MINICHROMOSOME MAINTENANCE1* (*MCM1*; *Saccharomyces cerevisiae*), *AGAMOUS* (*AG*; *Arabidopsis thaliana*), *DEFICIENS* (*DEF*; *Antirrhinum majus*), and *SERUM RESPONSE FACTOR* (*SRF*; *Homo sapiens*) (Schwarz-Sommer et al., 1990). The MADS-box TFs can be divided into the classes Type I and Type II depending on various evolutionary traits (Alvarez-Buylla et al., 2000; Gramzow and Theissen, 2010). Type II MADS-box TFs are involved in major phase changes in plant development, such as vernalization, flowering, and homeotic regulation of organ identity (Becker and Theissen, 2003; Kaufmann et al., 2005; Gramzow and Theissen, 2010). The Type I MADS-box TF class can be further divided into the $M\alpha$, $M\beta$, and $M\gamma$ subclasses (Parenicová et al., 2003) and members from different subclasses often dimerize (de Folter et al., 2005). The function of Type I MADS-box TFs remains elusive although they often appear to be involved in regulation of female gametophyte and endosperm development (Köhler et al., 2005; Bemer et al., 2008; Colombo et al., 2008; Steffen et al., 2008; Roszak and Köhler, 2011; Shirzadi et al., 2011; Hehenberger et al., 2012; Chen et al., 2016; Batista et al., 2019; Paul et al., 2020; Zhang et al., 2020).

Several genes encoding MADS-box TFs have been found to be imprinted, especially those of the Type I class (Masiero et al., 2011; Yoshida and Kawabe, 2013; Zhang et al., 2018), and most of them are specifically expressed in the endosperm during seed development (Bemer et al., 2010). Mutants of these genes often do not show a pronounced seed phenotype, presumably due to redundancy, making it difficult to investigate their function. However, a function for Type I MADS-box TFs that has been proposed, is to antagonize the function of FIS-PRC2 since inactivation of several Type I MADS-box genes reduced seed failure in *fis* mutants (Pires, 2014). Furthermore, Type I MADS-box genes have shown disrupted expression levels and imprinting patterns in hybrids (Josefsson et al., 2006; Walia et al., 2009; Burkart-Waco et al., 2015; Roth et al., 2019), suggesting that they contribute to the establishment of the postzygotic hybrid barrier (Walia et al., 2009). Future investigations of Type I MADS-box genes are required to understand their role in hybridization barriers putatively involving genomic imprinting.

Aim of Study

The overarching aim of this thesis was to elucidate the role and regulation of genomic imprinting. In order to do so, three different approaches were utilized: investigation of a protein family of which many members are known to be regulated by genomic imprinting, molecular characterization, and investigation of parental bias in allelic expression of a locus causing a gametophyte maternal effect mutant and the establishment of an experimental setup allowing for the identification of imprinted genes with spatial and temporal resolution.

Since Type I MADS-box TFs have shown disrupted expression in hybrids, we aimed to determine if these genes are involved in the establishment of post-zygotic hybrid barriers. Another aim was to evaluate the conservation of imprinting in the genus *Arabidopsis*.

Furthermore, we aimed to identify the molecular identity of the female gametophyte maternal effect mutant *capulet2* and to investigate if *CAPULET2* is regulated by genomic imprinting.

Efforts to identify imprinted genes in the endosperm of many different species have been hampered by contamination from surrounding tissues. To overcome this, we pursued to establish a method for isolation of endosperm-specific nuclei to reliably identify imprinted genes. With this method, we aimed to investigate the regulation of genomic imprinting in a temporal and spatial manner.

Results and Discussion

In order to elucidate the role of genomic imprinting in plants, it is paramount to identify imprinted genes that upon mutation have detectable phenotypes in seed development. In the work presented here, the imprinting state of various Type I MADS-box TF genes was determined and mutants of imprinted family members were investigated. Additionally, the imprinting profile of the maternally expressed *AGL36* was analyzed in close relatives of *A. thaliana* and hybrid crosses (Paper I). Furthermore, a whole-genome sequencing analysis identified the genomic locus for the female gametophyte maternal effect mutant *capulet2*. Independent transfer DNA (T-DNA) insertion lines were investigated and the genomic locus was verified by transgene complementation. Moreover, it was explored whether the *CAPULET2* gene is imprinted and by which mechanism the imprinting mark is established (Paper II). Finally, genomic imprinting was investigated in a temporal and spatial manner (Paper III).

Characterization of the *AGL36* subclade of Type I MADS-box TFs

To investigate the role of imprinting in the establishment of an endosperm-based post-zygotic hybrid barrier, MADS-box TFs closely related to the maternally expressed gene *AGL36* (Shirzadi et al., 2011) were analyzed (Paper I). A phylogenetic analysis of $M\alpha$ and $M\gamma$ Type I MADS-box genes in *A. thaliana* was performed (Figure 4a). Additionally, expression of $M\alpha$ and $M\gamma$ Type I MADS-box genes was investigated by RNA sequencing of whole seed samples taken at different timepoints, ranging from one day after pollination (DAP) to twelve DAP. Expression profiles (Figure 4b) indicated that $M\alpha$ and $M\gamma$ Type I MADS-box genes are expressed similarly during seed development, consistent with the hypothesis that TFs from these classes form dimers (de Folter et al., 2005). In the $M\gamma$ class, *AGL36* is closely related to *AGL90* and *AGL34* (Figure 4a), where *AGL90* has also been shown to be imprinted (Zhang et al., 2018). More distant, but still closely related members were identified, including the paternally expressed genes *AGL37* (*PHERES1*; *PHE1*) (Wolff et al., 2011), clustering together with the biparentally expressed *AGL38* (*PHERES2*; *PHE2*) (Villar et al., 2009), and *AGL92* (Wolff et al., 2011), clustering together with *AGL86*. Together with *AGL35*, these MADS-box genes form a subclade, the *AGL36*-clade, and members of the *AGL36*-clade and the *AGL36*-interacting MADS-box TFs *AGL28* (de Folter et al., 2005; Bemer et al., 2010) and *AGL62*

(Kang et al., 2008) were selected for further analyses. *AGL62* has been shown to be biparentally expressed (Kang et al., 2008) and *AGL28* has previously shown either accession-dependent imprinting (Wolff et al., 2011) or no imprinting (Zhang et al., 2018).

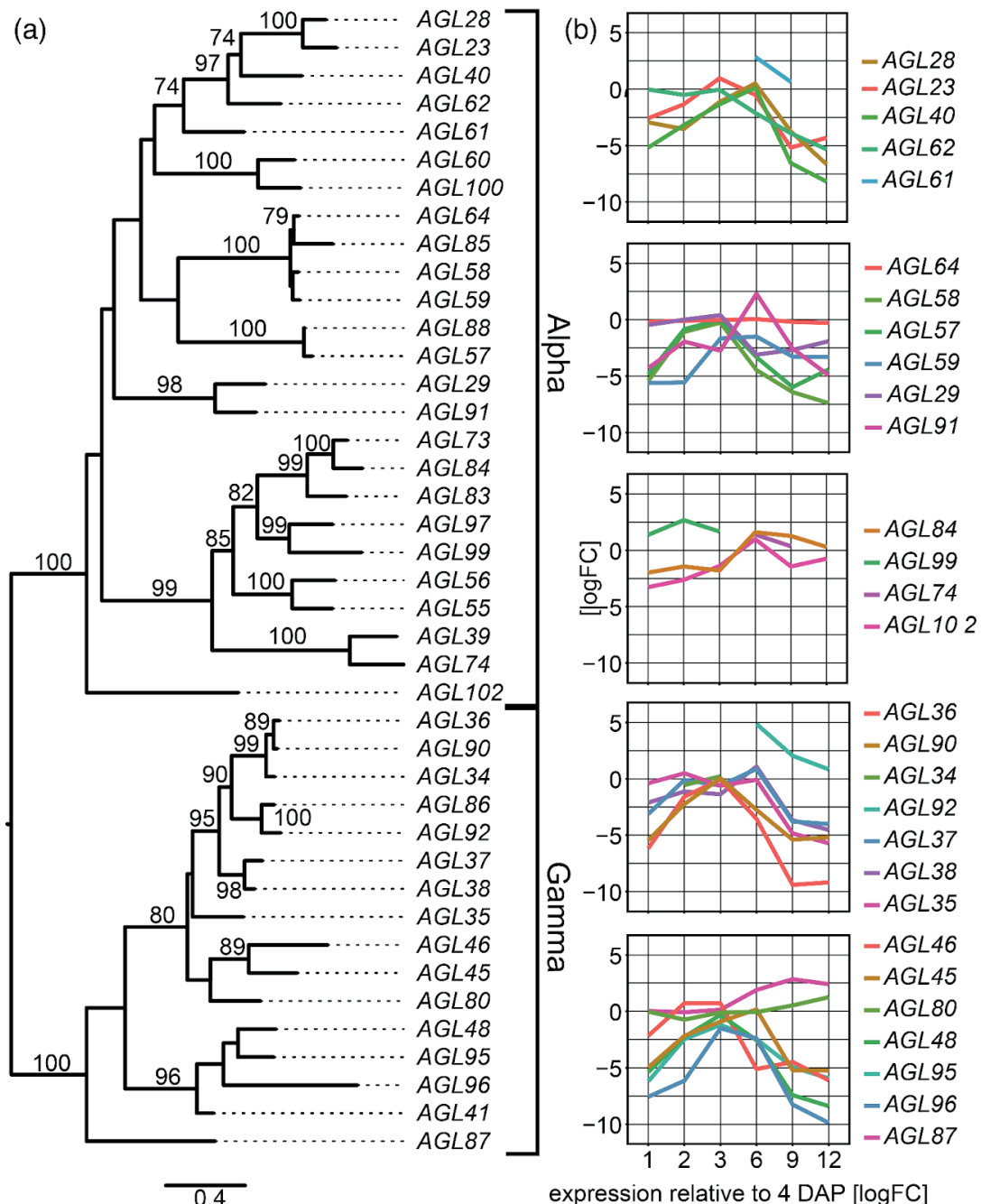


Figure 4: MADS-box Type I transcription factors share similar expression profiles during seed development. a) Phylogenetic tree of $M\alpha$ and $M\gamma$ MADS-box Type I genes in *Arabidopsis thaliana*. Scale bar represents the mean number of nucleotide substitutions per site. **b)** Gene expression profiles of $M\alpha$ and $M\gamma$ MADS-box Type I genes were classified into five groups according to the branching pattern. Gene expression profiles are relative to the expression at four days after pollination (DAP). Two biological replicates with three technical replicates were analyzed.

Members of the *AGL36*-clade are highly expressed at early stages of seed development and are collectively downregulated concurrent with endosperm cellularization, suggesting they are involved in this process (Figure 4b). However, mutants of the *AGL36*-clade members have not shown defected endosperm cellularization, except *AGL62* (Kang et al., 2008).

As *AGL36* has been shown to be a maternally expressed imprinting gene (Shirzadi et al., 2011), closely-related MADS-box genes (*AGL35*, *AGL36*, and *AGL90*) and the interacting MADS-box TF *AGL28* were investigated for parent-of-origin specific gene expression. Single nucleotide polymorphisms (SNPs) between the Columbia (Col-0) and Tsushima (Tsu-1) accessions were exploited to determine parent-of-origin allelic expression in reciprocal crosses (Paper I). In concordance with previous studies, *AGL36* (Shirzadi et al., 2011), *AGL90* (Zhang et al., 2018), and *AGL28* (Wolff et al., 2011) were maternally expressed, whereas *AGL35* (Zhang et al., 2018) was biparentally expressed. Furthermore, *AGL90* showed full paternal silencing with Tsu-1 as male parent and weak expression of paternal Col-0 in the reciprocal cross, consistent with previous reports (Hornslien et al., 2019).

Imprinting of *AGL36* in hybrids

To examine the conservation of imprinting of *AGL36* and of the influence of imprinting on the post-zygotic hybrid barrier, the investigation was extended to *A. arenosa* and *A. lyrata* subspecies *petrea* (Paper I). *AGL36-like* genes were identified in *A. arenosa* and *A. lyrata* by computational analysis, revealing that *A. arenosa* has two *AGL36-like* genes. Only one of the two *AGL36-like* genes was shown to produce transcripts and was named *AaAGL36-like*. Previous reports have indicated two *AGL36-like* genes in *A. lyrata* subspecies *petrea* (Yoshida and Kawabe, 2013), but only one *AGL36-like* gene could be verified, named *AlAGL36-like*.

In order to determine imprinting of *AGL36* in *A. arenosa*, a SNP in *AaAGL36-like* between two sister plants of the MJ09-4 population (Jørgensen et al., 2011) was identified and analyzed in cDNA from seeds from reciprocal crosses. This identified only maternal transcripts, indicating that imprinting of *AGL36* is conserved between *A. thaliana* and *A. arenosa* (Table 1; Paper I). Similarly, reciprocal crosses between *A. arenosa* and *A. lyrata* indicated that imprinting of *AGL36-like* genes is conserved in this interspecific hybrid (Table 1; Paper I).

Table 1: Investigation on the conservation of *AGL36* imprinting in the *Arabidopsis* genus. Crosses between *A. thaliana*, *A. arenosa*, and *A. lyrata* were performed to identify expression from the maternal and paternal allele to determine the conservation of imprinting of *AGL36*. Imprinting is conserved except in the *A. thaliana* x *A. arenosa* hybrid.

Parent		Expression		Conservation of <i>AGL36</i> imprinting
Male	Female	Maternal	Paternal	
<i>A. thaliana</i>	<i>A. thaliana</i>	Yes	No	Yes
<i>A. arenosa</i>	<i>A. arenosa</i>	Yes	No	Yes
<i>A. lyrata</i>	<i>A. arenosa</i>	Yes	No	Yes
<i>A. arenosa</i>	<i>A. lyrata</i>	Yes	No	Yes
<i>A. thaliana</i>	<i>A. arenosa</i>	Yes	Yes	No

However, in the *A. thaliana* x *A. arenosa* hybrid, both maternal *AtAGL36* and paternal *AaAGL36-like* alleles were detected, indicating that paternal silencing of *AaAGL36* was lifted in the hybrid seeds (Table 1; Paper I). The reciprocal cross was not investigated as *A. thaliana* is not able to fertilize *A. arenosa* (Comai et al., 2000; Bushell et al., 2003). This finding demonstrates that MEGs can also be deregulated in hybrid crosses, which has previously been shown primarily for PEGs (Josefsson et al., 2006; Burkart-Waco et al., 2015), and that conservation of genomic imprinting is species-dependent.

The post-zygotic hybrid barrier is affected by environmental and genetic factors

It has been demonstrated in rice that the temperature affects endosperm cellularization (Folsom et al., 2014) and that the expression of MADS-box genes is deregulated by heat stress (Chen et al., 2016). Additionally, different *A. thaliana* accessions have shown different strengths of the post-zygotic hybrid barrier (Burkart-Waco et al., 2012). Previous studies of post-zygotic hybrid barriers in the *Arabidopsis* genus have been performed under different temperature conditions (22°C compared to 18°C) and using different *A. arenosa* accessions (Strecno1; SN1, SN2 and MJ09-1) (Josefsson et al., 2006; Walia et al., 2009; Burkart-Waco et al., 2012; Lafon-Placette et al., 2017). To rule out temperature and/or accession-specific effects in the analysis of the post-zygotic hybrid barrier, hybrid crosses between *A. thaliana*, Col-0, and *A. arenosa* were performed at 18°C and 22°C with all four *A. arenosa* accessions, *i.e.* SN-1, SN-2, MJ09-1 and MJ09-4 (Paper I). Surprisingly, seed survival was substantially higher for hybrids with the *A. arenosa* accessions SN-1 and MJ09-4 grown at 18°C compared to 22°C. This suggests a correlation between temperature and the establishment of the post-zygotic hybrid barrier, based

on the timing of endosperm cellularization. Consistent with previous findings (Burkart-Waco et al., 2013), the embryo fails to develop from globular to heart stage at 22°C, whereas this transition was observed at 18°C. This indicates that endosperm cellularization plays a large role in establishing the hybrid barrier and that the temperature influences the severity of this barrier. The two other *A. arenosa* accessions (MJ09-1 and SN2) did not show sensitivity to temperature differences. This suggests that for the male parent, the accession genetic factor also influences the post-zygotic hybrid barrier. Furthermore, the influence of the female parent accession was investigated using the *A. thaliana* accessions Col-0, C24, Landsberg *erecta* (*Ler-1*), and Wassilewskija (*Ws-2*). We observed that only the *A. thaliana* accession *Ler-1* was not sensitive to temperature change for seed survivability, similar to MJ09-1 and SN2 as male accessions, suggesting that the temperature affects parental accessions specifically (Paper I).

Since the MADS-box TF family has been linked to the establishment of a post-zygotic hybrid barrier (Walia et al., 2009), MADS-box genes were investigated in this regard (Paper I). Homozygous single, double and triple MADS-box mutants could be obtained and no aberrant seed development was observed, except for *AGL62* as previously described (Kang et al., 2008). When these mutants were crossed to *A. arenosa* MJ09-4, seed survival was not affected in mutant hybrids compared to the wild-type (WT) hybrid (Col-0 x MJ09-4) at both 18°C and 22°C (Paper I). An exception was *agl35-1*, where germination was significantly lower and similar at both temperatures, suggesting that *AGL35* affects the post-zygotic hybrid barrier, irrespective of temperature. Altogether, this suggests that *AGL35* plays a crucial role in the establishment of the post-zygotic hybrid barrier and that the strength of this barrier is strongly affected by environmental and genetic factors.

The genomic locus of the female gametophyte maternal effect mutant *capulet2*

Most imprinted gene mutants, such as MADS-box TF genes (Paper I), do not show a detectable phenotypic difference in seed development (Masiero et al., 2011; Shirzadi et al., 2011; Berger et al., 2012; Wolff et al., 2015). Therefore, a female gametophyte maternal effect mutant of a putative imprinted gene, *capulet2* (*cap2*), could provide more insight into the function of genomic imprinting (Grini et al., 2002). The *cap2* mutant displayed a developmental arrest of the early embryo and endosperm when the mutant allele was transmitted maternally and previous genetic mapping reduced the potential genomic locus to a one mega base region on

chromosome 1 (Grini et al., 2002). In order to identify the gene identity of the mutation, whole-genome sequencing was performed of individuals carrying the mutant allele and WT individuals which were subjected to SNP analysis (Paper II). Several requirements and thresholds were applied to the SNP analysis, retaining three SNP candidates, among which *ANAPHASE PROMOTING COMPLEX 6 (APC6)*, a subunit of the ANAPHASE PROMOTING COMPLEX/CYCLOSOME (APC/C), was predicted to harbor the causative SNP for *cap2*, located in a donor splice site of *APC6*. Interestingly, a female gametophyte mutant of *APC6* has already been identified in *A. thaliana*, *nomega* (Kwee and Sundaresan, 2003), and in the rice homolog *OsAPC6* (Awasthi et al., 2012). However, these mutants showed distorted female gametophyte development whereas *cap2* displayed normal female gametophyte development (Grini et al., 2002).

Two independent *APC6* T-DNA mutant lines, *apc6-2* and *apc6-3*, were obtained and, similar to *cap2*, no homozygous mutants were detected (Paper II). Furthermore, both alleles showed a similar arrested embryo and endosperm developmental phenotype as *cap2* (Figure 5). Reciprocal crosses with Col-0 WT further indicated that the mutant seed phenotype was only observed when the *apc6* alleles were transmitted through the female parent, indicative of a gametophyte maternal effect mutant (Paper II). Interestingly, female gametophyte development was not distorted in any of the alleles (Paper II) in concordance with the previous description of *cap2* (Grini et al., 2002), but in contrast to the reported female gametophyte phenotype of *nomega* (Kwee and Sundaresan, 2003).

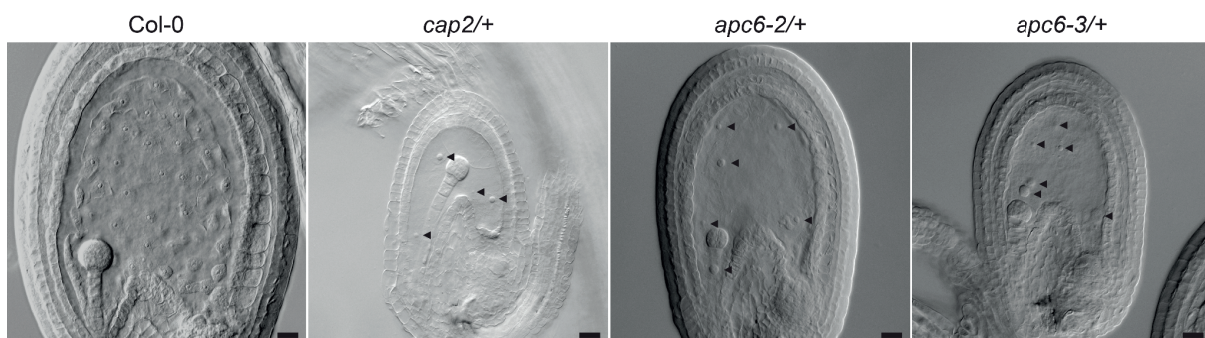


Figure 5: *cap2*, *apc6-2*, and *apc6-3* showed endosperm and embryo arrest phenotypes only when transmitted maternally. Seed phenotypes observed at the four days after pollination (DAP) stage of *cap2/+*, *apc6-2/+*, and *apc6-3/+* crossed maternally to Col-0 showed smaller seeds and arrested endosperm development. Fewer and larger endosperm nuclei were observed compared to native Col-0 endosperm nuclei. Arrowheads indicate endosperm nuclei. Scale bar = 20 μ m.

To verify the genomic locus of *cap2* in *APC6*, we transformed *cap2/+* and *apc6-3/+* with an *APC6* transgene (*pAPC6:APC6-GFP*; *APC6-GFP*). Additionally, the transgene was introgressed into *cap2* by crossing *cap2/+* paternally to two independent T2 homozygous *pAPC6:APC6-GFP* lines in a Col-0 background. All lines were propagated, genotyped and investigated for the mutant seed phenotype in the F2/T2 generation (Figure 6). A substantial reduction of seed phenotype frequency was observed for *cap2* mutants and homozygous individuals (*cap2/cap2*) could be identified, albeit only in a homozygous transgenic background. Furthermore, the mutant seed phenotype frequency of *apc6-3/+* mutants was substantially reduced in a hemizygous transgenic background, and lethality was fully alleviated in a homozygous transgenic background (Figure 6). No homozygous *apc6-3* individual was detected in the T2 generation and therefore two independent *apc6-3/+* T2 individuals in a homozygous transgenic background were propagated to the T3 generation. In this generation, full alleviation of the mutant seed phenotype was observed and homozygous individuals (*apc6-3/apc6-3*) were identified (Figure 6). These results verified successful complementation, demonstrating that *APC6* encodes *CAP2* (Paper II).

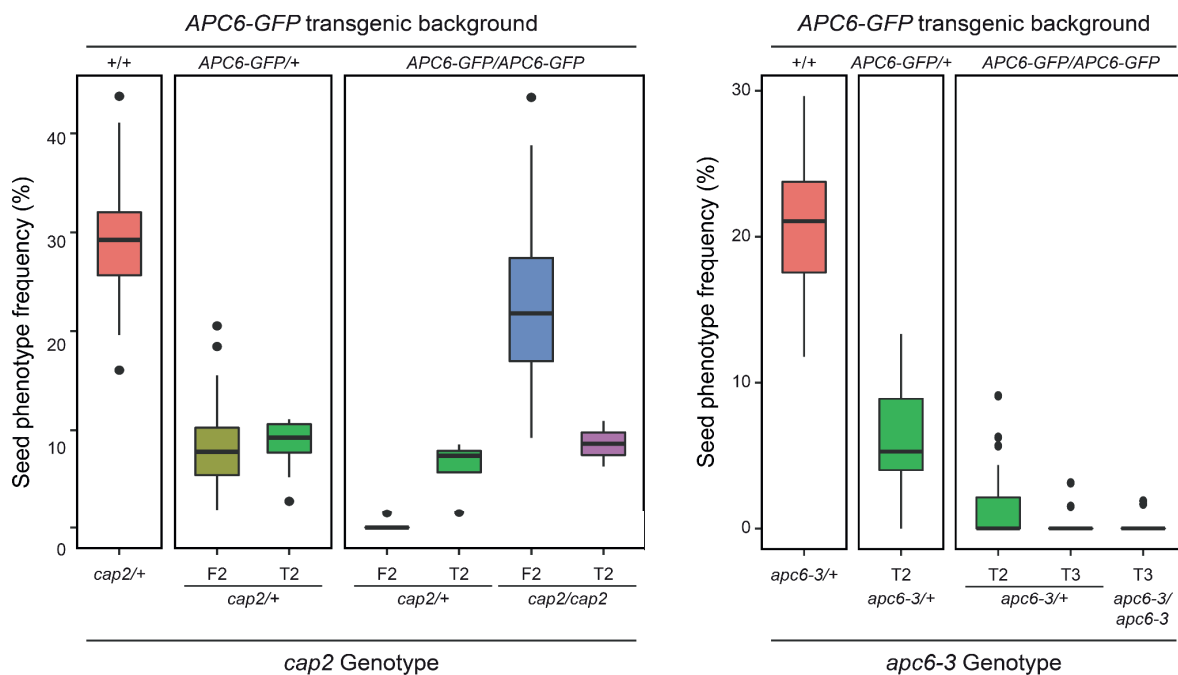


Figure 6: *cap2*, *apc6-2*, and *apc6-3* showed reduced mutant seed phenotypes after complementation with an *APC6* transgene. The *APC6* transgene was introduced into *cap2/+* and *apc6-3/+* and lines were propagated for two generations. Seed phenotype frequencies were determined and showed substantial reduction upon transgenic complementation in both a hemizygous (*APC6-GFP/+*) and homozygous (*APC6-GFP/APC6-GFP*) transgenic background, compared with native non-rescued mutants (+/+). Homozygous *cap2* and *apc6-3* individuals were identified confirming successful complementation.

Genetic analysis of *ANAPHASE PROMOTING COMPLEX 6*

With the identification of the causative SNP of *cap2* in *APC6*, an investigation of the consequences of mutations in this gene was performed (Paper II). The *cap2* mutant was previously described as fully penetrant, based only on the occurrence of the seed phenotype (Grini et al., 2002), while the *nomega* mutant has been described as partially penetrant (Kwee and Sundaresan, 2003). Since we did not observe mutant seed phenotype frequencies that corresponded to a fully penetrant mutant (*i.e.* 50% seeds would be lethal), we hypothesized partial penetrance of the mutant alleles to occur. In order to elaborate this observation, reciprocal crosses of *cap2*, *apc6-2*, and *apc6-3* with Col-0 WT were performed, and progeny was characterized for transmission of the mutant allele (Paper II). Consistent with our hypothesis, some maternal transmission was observed for *cap2*, *apc6-2*, and *apc6-3*, in contrast to previously reported (Grini et al., 2002).

To further investigate the development of mutant seeds, we performed phenotypic analysis at different developmental stages (two, four, six, and nine DAP). At two and four DAP, only the previously reported *cap2*-like phenotype was observed. However, we identified an endosperm cellularization defect at six and nine DAP (Paper II). In these seeds, the embryo appeared to have developed normally, albeit slightly delayed, while the endosperm remained syncytial. Because of the normal embryo development, we therefore suggest that the main defect in *apc6* mutants is in the endosperm. In hybrids, cellularization defects (too early, too late, or no cellularization) have previously been shown to be embryo-lethal (Paper I), although the embryo was viable in a nutrient culture after removal from the seed (Lafon-Placette et al., 2017). To investigate the correlation between defective endosperm and eventual embryo abortion, removal of the embryo from the mutant seed and placement in a nutrient culture could indicate whether the embryo of this class of seeds can survive.

Since the endosperm cellularization defective phenotype was detected only at six DAP or later, we hypothesized that *APC6* expression increases over time. Therefore, the expression of *APC6* was investigated using an RNA-seq dataset obtained from *Ler-1* x Col-0 crossed seeds at different timepoints (Paper I). An expression peak of *APC6* at six DAP was observed, coinciding with the onset of endosperm cellularization (Paper II). This finding, together with the observed cellularization defects at later developmental stages, may suggest a role for *APC6* and the APC/C in endosperm cellularization.

We determined that in all *apc6* mutants a premature stop-codon emerges (Paper II), resulting in various lengths of truncated APC6 protein. Since the insertion site of *apc6-2* is closely located to the *nomega* mutant location (Kwee and Sundaresan, 2003), it was surprising that the *apc6* mutants investigated in this thesis and *nomega* did not show similar female gametophyte phenotypes. Although these discrepancies were not investigated thoroughly, it seems unlikely that the differences were caused by genomic locations of the insertions, since *cap2*, *apc6-2*, and *apc6-3* are scattered throughout the *APC6* gene and these mutants did not show any phenotypic variation. Furthermore, a difference in maternal transmission of the mutant allele was observed between the initial study of *cap2* (Grini et al., 2002) and the *apc6* mutants studied in this thesis. Therefore, we speculate that environmental factors influence the penetrance of the mutation, since genetic factors, such as ecotype and mutation site, are alike (*cap2* and *nomega* Ler-1; *apc6-2* Col-0 and *apc6-3* Col-3). It has previously been shown that temperature substantially affects seed lethality in hybrids (Paper I). The different observations could possibly be explained by different growth temperatures (18°C in this thesis, 20°C for *cap2* (Grini et al., 2002), and 22°C for *nomega* (Kwee and Sundaresan, 2003)). It remains to be determined what exactly caused these observed differences.

APC6 is a maternally expressed imprinted gene

In order to investigate the imprinting status of *APC6*, various *A. thaliana* WT accessions (Ler-1, Col-0, C24, and Tsu-1) were crossed reciprocally. Genetic analysis indicated that SNPs were present between all accessions (except between Col-0 and Tsu-1), which were utilized in a restriction digestion analysis to identify parent-of-origin specific gene expression (Figure 7a). In all cross directions, except when C24 was crossed as the female parent, *APC6* expression can be observed as primarily maternal, indicating that *APC6* is an accession-specific MEG (Paper II).

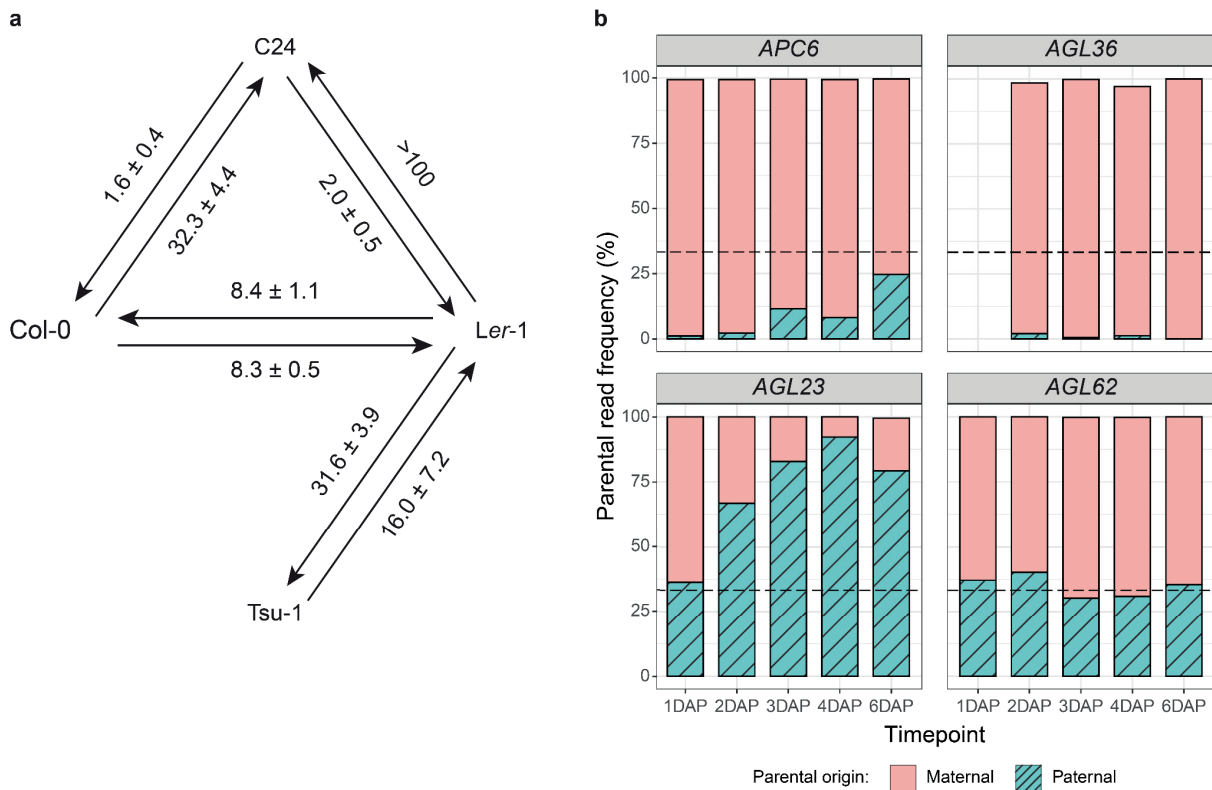


Figure 7: Single nucleotide polymorphism digestion analysis identifies *APC6* as a maternally expressed imprinted gene. **a)** Imprinting analysis of *APC6* from four days after pollination (DAP) seeds of reciprocal crosses of Col-0, C24, Ler-1, and Tsu-1 using accession-specific restriction digestion of single nucleotide polymorphisms (SNPs). Molarities from bioanalyzer fragments were determined to determine maternal:paternal expression ratio. The arrow is pointed from the female parent to the male parent. **b)** Sequenced reads obtained from Ler-1 x Col-0 RNA at one, two, three, four, and six DAP were mapped to *APC6*, *AGL36*, *AGL23*, and *AGL62*. SNP analysis was performed to determine the maternal and paternal allele frequency.

Imprinting of *APC6* was further investigated using an RNAseq dataset from Ler-1 x Col-0 seeds obtained at different timepoints (Paper I), and a SNP counting analysis was performed on reads mapped to *APC6*, *AGL36*, *AGL23*, and *AGL62* transcripts (Figure 7b). Consistent with the MEG control, *AGL36*, *APC6* showed a substantial maternal preferential bias up to the six DAP timepoint, at which the paternal read frequency reached 25%. This is to be expected, as the embryo is increasing in size and therefore the contribution from the embryonic paternal allele becomes more influential. The finding that *APC6* was mostly expressed from the maternal allele was further supported by the parental profile of the biparentally expressed *AGL62*, which showed a constant 2:1 maternal:paternal (m:p) read frequency ratio, and of the paternally expressed *AGL23*, which showed a substantially higher paternal read frequency already at two DAP. The latter indicated that the low paternal read frequency of *APC6* at two DAP was not due to general low expression from the paternal allele. An alternative possible explanation for the maternal bias of *APC6* could be that *APC6* is not expressed until six DAP

and that the maternal reads, dominating from one to four DAP, originated from the central cell as maternal carry-over. However, this seems unlikely due to dilution of the transcript abundance for each nuclear division.

The parental ratios obtained from the SNP digestion analysis (based on fragment molarities for *Ler-1* x *Col-0*) and the SNP counting on reads from RNA sequencing at four DAP are relatively similar (8m:1p and 11m:1p respectively). This strengthened the finding that *APC6* is imprinted since similar parental expression profiles were obtained using different methods. However, as mentioned earlier, contamination from the maternal seed coat or diploid embryo significantly hinders imprinting analysis (Schon and Nodine, 2017). In this thesis, whole seeds have been used for RNA sequencing (Paper I) and the SNP digestion analysis (Paper II). To accommodate potential contamination-associated bias for our imprinting analysis of *APC6*, publicly available microarray data of dissected seeds at four DAP (globular) were investigated (Belmonte et al., 2013). This data indicated that *APC6* is primarily expressed in the endosperm compared to embryo and seed coat (6:3:1 respectively). Since these tissues have distinct genotypes, we modulated an overall seed allelic frequency (1.89-fold maternal bias) for *APC6* in the scenario that both parental alleles are expressed equally, *i.e.* as a biparentally expressed gene (BEG). The SNP digestion analysis (8.3-fold or higher maternal bias) and SNP counting analysis (11.1-fold maternal bias) display substantially higher m:p ratios and strongly support the finding that *APC6* is a maternally expressed imprinted gene (Paper II).

Expression of the ANAPHASE PROMOTING COMPLEX/CYCLOSOME subunits is parentally regulated

In a broader perspective, *APC6* is part of the APC/C complex, a large E3 ubiquitin ligase protein complex crucial for the spindle assembly checkpoint and exit of mitosis and meiosis (Musacchio, 2015). The APC/C complex in *A. thaliana* consists of at least 14 highly conserved subunits and guides 26S proteasome-mediated protein degradation (Saleme et al., 2021). During metaphase of cell division, the chromosomes are attached to the spindle and the sister chromatids are held together by the cohesin ring. In a sequence of events, APC/C cleaves the cohesin ring (Lara-Gonzalez et al., 2012; Cromer et al., 2019), which allows the sister chromatids to separate and initiate the transition from metaphase to anaphase. This function of the APC/C is consistent with a previous observation in *cap2* of an arrested endosperm nucleus in metaphase (Grini et al., 2002). In addition to separating the sister chromatids, the APC/C

ubiquitylates Cyclin A and B to maintain and regulate cyclin activity (Attner and Amon, 2012). This has shown to be important for cell division as the accumulation of Cyclin A allowed cells to enter a third round of meiosis (Cromer et al., 2012) whereas a Cyclin A knock-out resulted in premature exit of meiosis I, generating diploid gametes (d'Erfurth et al., 2010). Similarly, buildup of Cyclin B has been detected in *nomega* and connects cyclin B accumulation to arrested female gametophyte development (Kwee and Sundaresan, 2003) whereas depletion of Cyclin B resulted in premature exit of mitosis (Chang et al., 2003).

Consistent with our observations of *apc6* mutants, other subunits have shown to be important for cell cycle regulation and gametophyte development. Mutants of APC/C subunits showed arrested female gametophyte development (*apc2* (Capron et al., 2003), *apc3a/apc3b* (Pérez-Pérez et al., 2008), and *apc10* (Eloy et al., 2011)) or male gametophyte development (*apc8* (Zheng et al., 2011; Xu et al., 2019) and *apc13* (Zheng et al., 2011)). In addition, overexpression of *APC8* also resulted in seed abortion similar to the knock-out, providing evidence that gene dosage levels need to be properly balanced for seed development (Zheng et al., 2011). Interestingly, similar phenotypes as for the *apc6* mutants of this thesis, *i.e.* distorted embryo and endosperm development caused by maternal gametophyte effects, were observed for *apc1* (Wang et al., 2013), *apc4* (Wang et al., 2012), and *apc11* (Guo et al., 2016). This could indicate that *APC1*, *APC4*, and *APC11* are similarly maternally expressed as *APC6* and possibly imprinted. Together with the findings in this thesis, this suggests that the APC/C complex is under strong maternal control during seed development.

Isolation of endosperm-specific nuclei allowed for imprinting analysis devoid of surrounding tissue contamination

In this thesis (Paper I and Paper II) and other studies (reviewed in Schon and Nodine, 2017), the possibility of contamination from surrounding tissues has been troublesome, although different methods have been applied to accommodate this. Therefore, we have developed a system that allows for the isolation of endosperm-specific nuclei using fluorescence-activated nuclear sorting (FANS) (Paper III). A dual component system was utilized (Weijers et al., 2003; Olvera-Carrillo et al., 2015) in which the H2A-GREEN FLUORESCENT PROTEIN (GFP) fusion protein was expressed under the control of an endosperm-specific promoter (*proMARKER*>>H2A-GFP). In selected endosperm nuclei, the expression of the transcriptional fusion activator *mGAL4-VP16* is promoted, which in turn activates *H2A-GFP* expression

through the GAL4-activated upstream activation sequence (UAS) regulatory element (Olvera-Carrillo et al., 2015). Furthermore, using promoters of genes that are expressed at different developmental stages or in different domains of the endosperm, this system can further be extended to the investigation of genomic imprinting in a temporal and spatial specific manner. For this purpose, endosperm domain or stage specific promoters were selected (Paper III) based on microarray data (Le et al., 2010) and on gene expression patterns (Winter et al., 2007). To investigate temporal-specific imprinting, two promoters were selected that were expressed in the entire endosperm at the globular stage (early endosperm; EE), representing the syncytial endosperm (Figure 2; left), and at the late cotyledon stage (total endosperm; TE1), representing the cellularized endosperm (Figure 2; right). For spatial-specific imprinting, two promoters were selected that are expressed only in certain domains of the cellularized endosperm (Figure 2; right): the embryo surrounding region (ESR) and the developing aleurone layer (DAL).

The expression vectors were transformed into the *A. thaliana* accession Col-0 which were subsequently crossed maternally to Tsu-1 (Paper III). To increase mapping sensitivity, and since Tsu-1 does not have a reference transcriptome, we generated and polished timepoint-specific (four DAP for EE and seven DAP for ESR, DAL, and TE1) gene target sequences for Col-0 and Tsu-1 from homozygous WT seeds (Paper III). Significant temporal (EE vs TE1) and spatial (ESR vs TE1) differential expression was observed (Paper III), consistent with previous findings that endosperm domains have distinct expression profiles (Belmonte et al., 2013; Del Toro-De León and Köhler, 2019; Picard et al., 2021).

Identification of parental allele-specific expression

For the identification of imprinted genes, we used and modulated the Informative Read Pipeline (IRP; (Hornslie et al., 2019)). For each marker, this pipeline identified and extracted so-called informative reads, *i.e.* read pairs that cover accession-specific SNPs or insertion/deletion (InDels). A read pair is informative if it adheres to any of the following conditions: 1) it maps identically to one target allele, but not to the other (*e.g.* 0 versus 1 SNP); 2) it maps to both alleles without InDels, but the SNP count with one target allele is smaller than half the SNP count with the other target allele (*e.g.* 2 versus 5 SNPs); 3) it maps to one allele with InDels, but it maps to the other allele without InDels and with maximal one SNP. The majority of informative reads were identified based on SNP counts compared to the presence of InDels (Table 2). This suggests that the major difference between accessions results from SNPs and

not from InDels. Furthermore, at an early stage (EE) a larger part of informative reads was selected by SNPs (89%) compared to the later stage markers (ESR, DAL, and TE1; 86%). This suggests that the type of transcript differences between accessions varies depending on the seed developmental stage.

Table 2: Selection of informative reads by SNPs or InDels. Informative reads were identified and extracted by the Informative Read Pipeline (IRP; (Hornslien et al., 2019)). Informative reads were selected based on the presence of single nucleotide polymorphisms (SNP) or insertions/deletions (InDel). EE: early endosperm; ESR: embryo surrounding region; DAL: developing aleurone layer; TE1: total endosperm.

Marker line	Informative reads selected by	
	SNP	InDel
EE	89%	11%
ESR	86%	14%
DAL	86%	14%
TE1	86%	14%

After extraction of informative reads, various filters and normalization steps were applied, as described previously (Hornslien et al., 2019). One additional filtering step was implemented due to the lack of the reciprocal cross in our experimental setup. This could potentially result in the discovery of false-positive imprinted genes because of accession-biased expression. For instance, a gene that is 10-fold higher expressed in Col-0 compared to Tsu-1 and normally shows biparental gene expression will result in much higher informative read counts that align to the Col-0 target sequence. Therefore, this gene is prone to be identified as a MEG in one direction and a PEG in the reciprocal cross. When reciprocal crosses are included, this could be easily detected as a false positive. Therefore, to exclude potential false-positive MEGs and PEGs, an accession-specific filter was established (Paper III). Differential gene expression analysis between homozygous Col-0 and Tsu-1 WT was performed (Figure 8). Genes that were identified to have an accession specific bias in expression were omitted from further analysis and only genes that showed similar expression patterns were retained for imprinting analysis (Figure 8).

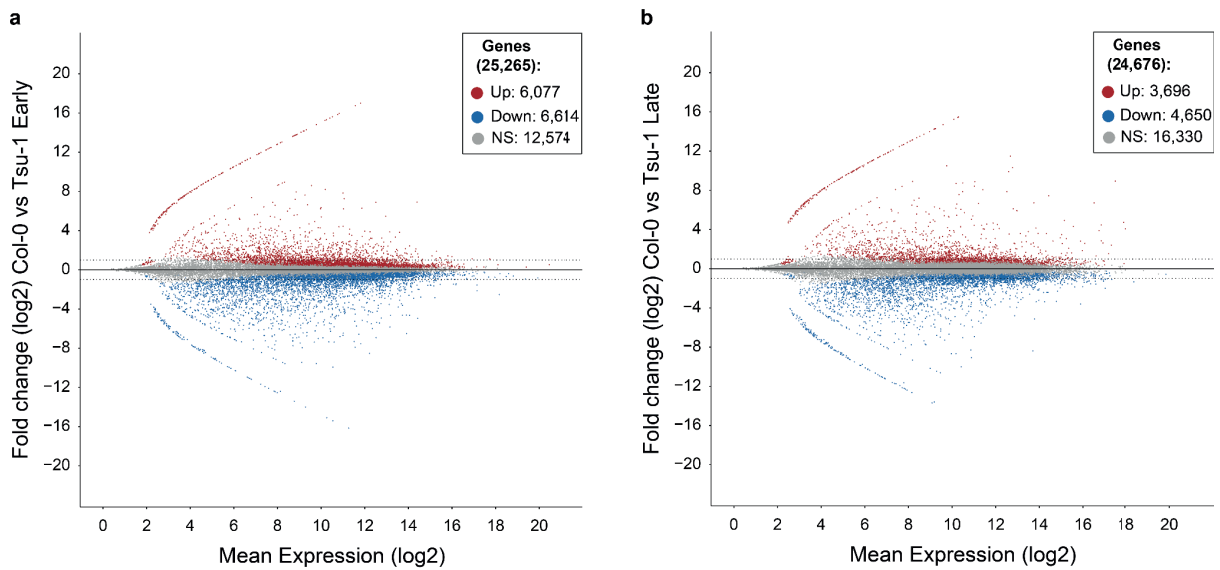


Figure 8: Ecotype-specific gene expression analysis between Col-0 and Tsu-1 at different seed stages. Differential expression between Col-0 and Tsu-1 at four days after pollination (DAP) (a) and seven DAP (b). Non-significantly regulated (NS) genes are depicted in gray and were used for the imprinting study. Dashed lines indicate fold change (log₂) of 1 and -1.

After filtering and normalization of informative reads, statistical analysis was performed to identify imprinted genes in EE, ESR, and TE1 (Paper III). A high number of MEGs and PEGs was detected and overall, more MEGs (82, 69, and 60 in EE, ESR, and TE1 respectively) than PEGs were identified (28, 22, and eleven in EE, ESR, and TE1 respectively). Only a limited overlap in imprinted genes was observed between EE and TE1 (two MEGs and no PEGs). A larger overlap was observed, as expected, between ESR and TE1 (26 MEGs and three PEGs). However, it has to be noted that there are genes that are only significantly imprinted in one domain, but that show similar parental non-significant expression bias in the other domain.

Genomic imprinting is dynamically regulated throughout seed development

To identify genes that are imprinted in one developmental stage, but biparentally expressed in the other developmental stage, the m:p fold changes (FC) of significant MEGs and PEGs in EE, TE1, or both were visualized in a scatterplot (Figure 9). As expected, albeit that most genes were significantly imprinted at only one developmental timepoint, almost half of the genes (21/56) showed similar parental expression bias, indicating that they maintain their imprinting state throughout seed development (Figure 9, diagonal green area). Several genes were observed to be imprinted at one developmental timepoint, but showed biallelic expression at the other developmental timepoint, indicating temporal-specific imprinting profiles (Paper III),

consistent with findings in maize (Zhang et al., 2011; Xin et al., 2013; Dong et al., 2017). Genes have been identified that displayed both the transition from biparental expression to imprinted expression and from imprinted expression to biparental expression (Zhang et al., 2011).

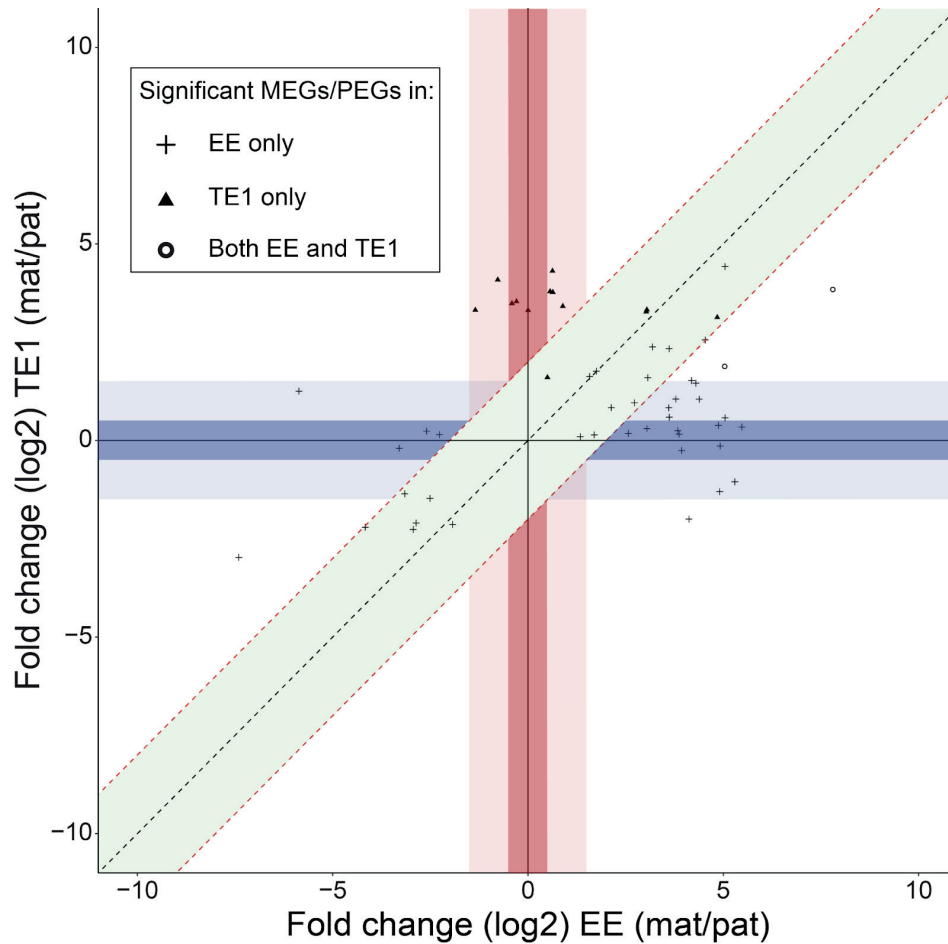


Figure 9: Identification of temporal regulation of imprinting. Scatter plot representation of the maternal:paternal fold change of significantly called imprinted genes at EE, TE1, or in both. Genes identified as MEGs or PEGs in one domain; EE (plus) or TE1 (triangle) or both domains EE and TE1 (circle) were included. Genes show similar parental expression bias (diagonal green area) or temporal expression bias (imprinted at EE, biparental at TE1: dark blue area; imprinted at TE1, biparental at EE: dark red area).

Eleven significantly imprinted genes (eight MEGs and three PEGs) in EE showed biallelic expression in TE1 (Figure 9, horizontal dark blue area) of which several have been previously identified at a similar developmental timepoint (four DAP) (Del Toro-De León and Köhler, 2019; Picard et al., 2021). This finding indicated that these genes lost their imprinting mark as the endosperm developed. The expression of these genes in EE and TE1 was investigated using data from the differential gene expression analysis (Paper III). This revealed that most MEGs in EE have an increased expression in TE1, which we correlated with the reactivation of the paternal allele (Paper III). The reactivation of the paternal allele can be achieved by removal

of the DNA methylation mark by members of the 5mC DNA glycosylase family (Penterman et al., 2007). We analyzed the expression of DNA methyltransferases and DNA glycosylases in EE and TE1 and the total normalized read counts indicated that all genes involved in these processes, except *CMT1*, *DRM1*, and *DML3*, are highly expressed (Table 3). The DNA glycosylases *DEMETER-LIKE 2 (DML2)* and *ROS1* were substantially expressed in the endosperm with *ROS1* showing >20-fold higher expression in TE1 compared to EE (Table 3). This makes *ROS1* a suitable candidate for the removal of the epigenetic mark on the paternal allele at a later stage in seed development. Interestingly, *DME*, another DNA glycosylase, was also found to be slightly upregulated (2-fold) at TE1 (Table 3). This was surprising as *DME* expression was described to be restricted to the central cell (Choi et al., 2002; Ibarra et al., 2012; Park et al., 2020).

Table 3: Expression of DNA methyltransferases and DNA glycosylases in EE and TE1. The total normalized read count indicates the expression level of each gene per domain. Fold change (log2) between EE and TE1 was determined by differential gene expression analysis (Paper III).

Gene	Total normalized reads		Fold change (log2) EE vs TE1
	EE	TE1	
<i>MET1</i>	9861	3645	1.4
<i>MET2A</i>	16349	34878	-1.1
<i>MET2B</i>	21295	32408	-0.6
<i>MET3</i>	1868	1679	0.1
<i>CMT1</i>	457	5	5.8
<i>CMT2</i>	5320	15160	-1.5
<i>CMT3</i>	12124	2503	2.2
<i>DRM1</i>	220	61	1.1
<i>DRM2</i>	520	5838	-3.4
<i>DRM3</i>	888	3906	-2.1
<i>DME</i>	26636	58804	-1.1
<i>ROS1</i>	1643	35720	-4.4
<i>DML2</i>	17637	20412	-0.2
<i>DML3</i>	159	1	5.2

However, our finding is supported by publicly available microarray data (Belmonte et al., 2013), which showed that *DME* is expressed in the endosperm and that expression is similarly upregulated at later stages (2.7-fold higher at the late cotyledon stage than at the globular stage). Whether the DME DNA glycosylase family members (*DME* and *ROS1*) are actively involved

in DNA demethylation in the endosperm at later stages of seed development remains unclear and requires further investigation.

For PEGs in EE that were biallelically expressed in TE1, differential gene expression analysis indicated that expression is not significantly different between stages (Paper III). This indicated that there was a dynamic regulation of both parental alleles, *i.e.* an increase in maternal expression and a decrease in paternal expression. In order to regulate such a dynamic process, multiple pathways are required to coalesce and, as stated earlier, different epigenetic pathways have been shown to coordinate together in the regulation of parental expression.

Interestingly, we identified three MEGs at the late TE1 stage that showed biparental expression at the syncytial EE stage (Figure 9, vertical dark red area) (Paper III). This suggested that these genes obtain their imprinting mark throughout seed development, challenging the canonical definition of imprinting which states that imprints are established in the gametes. For two genes (AT1G11940 and *ERF/AP2 TRANSCRIPTION FACTOR 17 (ERF17)*), differential gene expression analysis indicated that there is no significant difference between EE and TE1 (Paper III), indicating that there was a simultaneous dynamic regulation of parental alleles. However, we did observe that at TE1, the paternal alleles were completely silenced, whereas the maternal allele was expressed throughout seed development (Paper III), from which we speculate that the paternal alleles obtained the epigenetic silencing marks between the EE and TE1 developmental stage. The RdDM pathway is considered to regulate *de novo* DNA methylation through DRM2 and we observed that this gene is highly upregulated (>10-fold) at TE1 compared to EE (Table 3). A similar upregulated expression of *DRM2*, 4.3-fold higher at the late cotyledon stage (similar to TE1) than at the globular stage (similar to EE), was shown by publicly available microarray data (Belmonte et al., 2013). However, for this *de novo* DNA methylation to happen, the epigenetic machinery must still be able to distinguish parental alleles in order to successfully modify the paternal allele. We speculate that a secondary mechanism, possibly other epigenetic marks, must be present in the gametes that enables recognition of the paternal allele at later developmental stages and further investigations are necessary to elaborate on this possible mechanism.

Differentiated endosperm cells possess unique imprinting profiles

As a next step, we investigated the dynamic regulation of genomic imprinting between different spatial domains. Across spatial-specific endosperm markers (ESR and TE1), and when only significant imprinted genes were considered, only 29 imprinted genes overlapped. In order to determine if these were truly the only overlapping spatially imprinted genes, the m:p fold changes of significant MEGs and PEGs in the ESR, TE1, or both were visualized in a scatterplot (Figure 10). This allowed us to identify genes that are imprinted in one endosperm domain, but biparentally expressed in the other domain (Paper III). As expected, the majority of genes (80/110) showed similar parental expression bias (Figure 10, diagonal green area). This also indicated that looking only at significant imprinted genes, which indicated that only 29 genes overlapped between ESR and TE1, might provide misrepresented domain specificity.

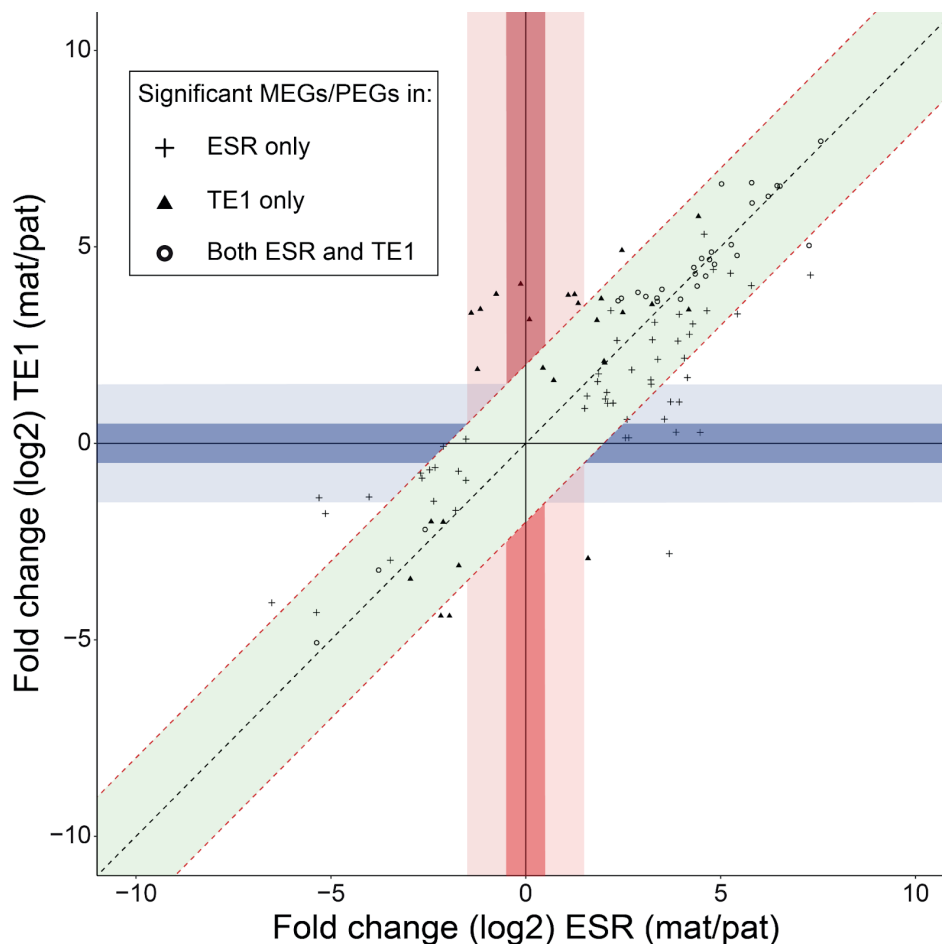


Figure 10: Identification of spatial regulation of imprinting. Scatter plot representation of the maternal:paternal fold change of significantly called imprinted genes in TE1, ESR, or in both. Genes identified as MEGs or PEGs in one domain; ESR (plus) or TE1 (triangle) or both domains ESR and TE1 (circle) were included. Genes show similar parental expression bias (diagonal green area) or spatial expression bias (imprinted at ESR, biparental at TE1: dark blue area; imprinted at TE1, biparental at ESR: dark red area).

Five significantly imprinted genes (four MEGs and one PEG) in the ESR showed biparental expression in TE1 (Figure 10, horizontal dark blue area) and one of these loci has been previously identified as imprinted (Del Toro-De León and Köhler, 2019; Hornslien et al., 2019). All of these identified imprinted genes showed comparable expression of the expressed allele in the ESR and TE1, whereas the silenced allele (paternal allele for MEGs and maternal allele for PEGs) was upregulated in TE1. This suggested that the silenced allele is specifically imprinted in the ESR, but not in the total endosperm overall (Paper III). Additionally, two significantly imprinted MEGs in TE1 showed biparental expression in the ESR (Figure 10, vertical dark red area). Indirectly, this provided evidence that genomic imprinting can be regulated in an endosperm domain-specific manner since the imprinting profile observed in the total endosperm must be caused by parental-specific expression outside the ESR (Paper III).

In order to resolve how this domain-specific imprinting was established, we investigated the expression and imprinting profile of these seven genes in the EE data. Although only four genes could be investigated, we observed both biparental as parental biased expression, suggesting that, for the genes identified, both reactivation and silencing mechanisms acted in the ESR and other domains of the endosperm (Paper III). In this experimental setup, it is needed to take into consideration that the ESR is part of the total endosperm and consequently, the expression and imprinting profile of TE1 is affected by the ESR. Although we did not quantitatively analyze the contribution of the ESR to TE1, we presumed that this is minimal and that genes, imprinted in the ESR, were masked by the expression profile in TE1. For imprinted genes in TE1 that showed biallelic expression in the ESR, we concluded that the imprinting profile of TE1 must be established in the endosperm outside the ESR.

To investigate possible mechanisms that could establish domain-specific imprinting, we investigated the expression of various DNA methyltransferases and DNA glycosylases in ESR and TE1 (Table 4). The expression of *MET2A*, *MET2B*, *MET3*, and *DME* is much higher in the ESR than in TE1, suggesting these genes may act specifically in the ESR. It has to be noted that high expression in the ESR will be visible in TE1. The expression of these genes is likely more restricted to the ESR than these values represent. Conversely, a much higher expression in TE1 compared to ESR, as can be seen for *DML2*, seems more likely to be caused by expression in a different domain in the endosperm. Domain-specific expression of DNA methyltransferases and DNA glycosylases could contribute to domain-specific establishment or alleviation of imprinting marks.

Table 4: Expression of DNA methyltransferases and DNA glycosylases in ESR and TE1. The total normalized read count represents the expression level of each gene per domain. Fold change (log2) between ESR and TE1 was determined by differential gene expression analysis (Paper III).

Gene	Total normalized reads		Fold change (log2) ESR vs TE1
	ESR	TE1	
<i>MET1</i>	2792	3645	-0.4
<i>MET2A</i>	187144	34878	2.4
<i>MET2B</i>	116436	32408	1.8
<i>MET3</i>	8701	1679	2.3
<i>CMT1</i>	0	5	-0.6
<i>CMT2</i>	7772	15160	-0.9
<i>CMT3</i>	1320	2503	-0.9
<i>DRM1</i>	132	61	0.5
<i>DRM2</i>	4963	5838	-0.2
<i>DRM3</i>	2899	3906	-0.4
<i>DME</i>	149739	58804	1.3
<i>ROS1</i>	34792	35720	0
<i>DML2</i>	7307	20412	-1.4
<i>DML3</i>	0	1	-0.3

Overall, the results described above provided strong evidence that imprinting analysis of specific endosperm domains using FANS enabled the identification of imprinted genes that would otherwise be masked by the overall biparental expression in the total endosperm (Paper III). These results also showed that regulation of imprinting profiles in subdomains of the endosperm is more complex than previously thought and that imprinting marks may be dynamically established in a spatial-specific perspective.

Limited overlap between studies in identified imprinted genes

Across all endosperm domains combined (EE, ESR, and TE1), we detected 181 MEGs and 56 PEGs, and gene ontology (GO) term enrichment analysis was performed (Paper III). In line with previous publications (Pignatta et al., 2014; Picard et al., 2021), our imprinted gene data were enriched for genes encoding for gene regulatory proteins, and more specifically, MEGs from all domains were enriched for transcription factor activity. In order to assess the conservation of identified imprinted genes, the overlap with other previously published datasets (Pignatta et al., 2014; Del Toro-De León and Köhler, 2019; Hornslien et al., 2019; Picard et al., 2021) was estimated (Paper III). Out of 237 total identified imprinted genes in this thesis,

78 (60 MEGs and 18 PEGs) were previously identified in any of the other studies (33%), and similar overlap percentages (17% - 54%) were observed between these other studies (Table 5).

Table 5: Overlap of identified imprinted genes. Total number of identified imprinted genes per study and number of imprinted genes overlapping in one or more of the other studies.

Study	Total identified	Overlapping genes	Overlap (%)
Pignatta et al., 2014	388	144	37%
Hornslie et al., 2019	317	54	17%
Del Toro-De León et al., 2019	925	223	24%
Picard et al., 2021	349	190	54%
This thesis (Paper III)	237	78	33%

This was consistent with previous observations that overlap is limited between studies within the same species (Schon and Nodine, 2017; Hornslie et al., 2019). Interestingly, the overlap in identified imprinted genes was highest between this thesis and studies that used a similar methodological approach for RNA extraction (FANS/INTACT; Paper III), strengthening the reliability of our experimental strategy for the isolation of endosperm-specific nuclei (Del Toro-De León and Köhler, 2019; Picard et al., 2021). For the identification of imprinted genes in this thesis (Paper III), it is possible that genes were excluded prior to imprinting analysis by any of the filtering steps, due to the absence of accession-specific SNPs or InDels and/or accession-biased preferential expression. The accession expression bias filtering step established in this thesis (Paper III) could indeed be used to establish the influence of the latter. When these genes, that have been identified as imprinted in any of the other studies but did not pass our filtering requirements, were not included for comparison (*i.e.* these genes were removed from the imprinted gene list of the other studies), the overlapping number of imprinted genes increased substantially, suggesting that bioinformatic filtering and thresholds affect the output of identified imprinted genes (Paper III).

Furthermore, due to the detection of widespread contamination (Schon and Nodine, 2017) and the poor overlap, the number of MEGs and PEGs has remained disputable. In order to elaborate on the low overlap between the different studies, experimental setups were compared and a large variation was observed (Paper III). Different accessions, timepoints of seed development, and growth conditions, such as temperature, could all contribute to the high variety of identified imprinted genes. In this thesis (Paper II), accession-specific imprinting of *APC6* was detected and, combined with the lack of SNPs between one accession pair (Col-0 and Tsu-1), indicated

that selection of multiple accessions may be required to reliably qualify a gene as imprinted. Furthermore, we have demonstrated temporal dynamic regulation of imprinting (Paper III), stipulating that different sets of imprinted genes are identified at different developmental seed stages. Moreover, we have shown that the temperature substantially affected hybrid seed lethality (Paper I) suggesting that temperature may have a role in the regulation of genomic imprinting. Additionally, both the temperature and accession influence the endosperm growth-rate, and therefore also affect the developmental seed stage. Instead of selecting the timepoint based on days after pollination, it may be more appropriate to select seeds based on the developmental seed stage. Lastly, different methods for RNA extraction and sequencing (whole seed transcriptome analysis vs nuclear sorting by FANS/INTACT), together with filtering and threshold parameters of the data, could contribute to the bias in the identification of imprinted genes. Overall, from the results found in this thesis, it is apparent that the identification of imprinted genes is highly sensitive to the experimental setup.

Imprinting is often not regulated only by MET1

The regulatory mechanism behind imprinting was investigated for the Type I MADS-box genes (Paper I) and *APC6* (Paper II). Mutants of components of the DNA methylation pathway (*MET1*; *CMT3*) and RdDM pathway (*NRPD1*; *DRM1*; *DRM2*) were used to determine if parental silencing of the analyzed genes is regulated by any of these pathways. The paternal allele of *AGL36* and *AGL90* was not reactivated in crosses with a hemizygous *met1-7/+* (in a Col-0 background) as a male parent and Tsu-1 as female parent, while a slight paternal reactivation for *AGL28* was observed (Paper I). This strongly indicated that imprinting of these genes is not or only in part regulated by MET1 alone. The finding that *met1-7/+* did not reactivate the paternal allele of *AGL36* was in contrast to a previous report (Shirzadi et al., 2011). A different *met1* mutant allele, *met1-4*, was used in the study where paternal reactivation of *AGL36* was detected (Shirzadi et al., 2011). Different *met1* mutant alleles have previously shown varying methylation effects (Kankel et al., 2003). It might be that *met1-7* is a weaker mutant allele than *met1-4* and does not result in a full knock-out. Furthermore, self-propagation of homozygous *met1* mutants may result in an increasingly demethylated genome (Mathieu et al., 2007; Mirouze et al., 2009). This makes the observed reactivation of the paternal allele by Shirzadi et al., 2011 more difficult to directly correlate. Therefore, and in contrast to *met1-4*, the *met1-7* allele was maintained hemizygous (Hornslien et al., 2019). It has been shown that *AGL36* is strongly upregulated in a hybrid cross with a mutant of *NRPD1*, a core subunit of

PolIV (Lu et al., 2012), although imprinting of *AGL36* was proposed to be independent of DRM2 or AGO4 (Shirzadi et al., 2011). Therefore, the role of RdDM in imprinting of *AGL36* was further investigated with the *nripd1* mutant. No reactivation of the paternal allele was observed and RT-PCR analysis indicated that *AGL36* is not upregulated in *nripd1* crossed reciprocally with Col-0 (Paper I).

The hemizygous *met1-7* mutant was used as the male parent, with *Ler-1* as the female parent, to investigate the paternal silencing of *APC6* (Paper II). No paternal reactivation was observed, suggesting that imprinting of *APC6* is not regulated by MET1. We then explored the influence of the DNA methyltransferases *DRM1*, *DRM2*, and *CMT3* on imprinting of *APC6* by crossing the hemizygous *drm1-2;drm2-2;cmt3-11* triple mutant paternally to the *Ler-1* accession (Paper II). We did not observe reactivation of the paternal allele and to accommodate for substantial gene redundancy (He et al., 2022; Liang et al., 2022), we generated a hemizygous *met1-7;drm1-2;drm2-2;cmt3-11* quadruple mutant. In a similar approach, the hemizygous quadruple mutant was crossed paternally to *Ler-1*, but no reactivation of the paternal allele was observed (Paper II). These results suggested that imprinting of *APC6* is not regulated by any of the canonical DNA methyltransferases and that other mechanisms must be involved (Paper II). Indeed, other DNA methyltransferases, *MET2A*, *MET2B*, and *CMT2*, show high expression in the early stage (EE) of endosperm development (Table 3), and further analyses are required to determine the mechanism behind the imprinting of *APC6*.

Concluding remarks and future perspectives

The overarching aim of this thesis was to enhance our understanding of the function of genomic imprinting and investigate this using different experimental approaches. We have shown that the conservation of imprinting of *AGL36* is species-dependent and we identified substantial environmental and genetic effects on the establishment of the post-zygotic hybrid barrier (Paper I). Furthermore, the female gametophyte maternal effect mutant *capulet2* was identified to be located in *APC6* which was shown to be a maternally expressed imprinted gene (Paper II). Lastly, we successfully isolated endosperm-specific nuclei devoid of contamination from surrounding tissue and identified genes that were dynamically imprinted in a temporal and spatial manner (Paper III). Collectively, the results of this thesis provide evidence for dynamic regulation of genomic imprinting in seed development and this clearly shows that the regulation of genomic imprinting is more complex and dynamic than previously thought.

Various members of the *AGL36*-clade showed parent-of-origin specific gene expression regulated by genomic imprinting, but single, double, and triple mutants of these imprinted genes did not show distorted seed development. A major finding was that imprinting of *AGL36* was conserved in the *A. lyrata* x *A. arenosa* hybrid, but that it was misregulated in the *A. thaliana* x *A. arenosa* hybrid. Interestingly, the same male parent was used in both hybrid crosses and silencing of the paternal allele was dependent on the female parent. This may indicate that imprinting is not entirely dependent on marks established in the gametophytes and that imprinting can be dynamically reprogrammed during endosperm development. Lastly, environmental and genetic factors were shown to substantially affect the post-zygotic hybrid barrier, and *agl35-1* strongly aggravated this barrier. What do these findings mean for the study of genomic imprinting and more specifically, for the Type I MADS-box TF family? Many members of the *AGL36*-clade showed to be regulated by genomic imprinting, although single mutants of *MET1* and *NRPD1* did not alleviate the imprint. However, redundancy between CG and non-CG methylation can occur (He et al., 2022; Liang et al., 2022) and therefore DNA methyltransferase redundancy could explain the lack of paternal allele reactivation. In order to fully apprehend the molecular mechanisms behind the imprinting of the *AGL36*-clade members, a larger mutant analysis is required, including additional DNA methyltransferases. Furthermore, we have shown that the temperature strongly affects the post-zygotic hybrid barrier and likewise, we speculate that the temperature influences the regulation of parent-of-origin specific gene expression. One scenario could be that the growth temperature determines

whether a parental allele is completely or partially silenced. Then, temperature-mediated regulation of genomic imprinting could also be responsible for different strengths of the post-zygotic hybrid barrier. A thorough examination of the influence of temperature on genomic imprinting and the post-zygotic hybrid barrier could provide substantial insight into this hypothesis.

We have identified that the causative SNP for the gametophyte maternal effect mutant *capulet2* is located in *APC6* and we identified this gene as a maternally expressed imprinted gene. Additionally, *apc6* mutants displayed defective endosperm cellularization and misregulated timing of endosperm cellularization often results in seed abortion. Therefore, we speculate that *APC6* and the APC/C have a role in the cellularization process of the endosperm. Yet, the question remains why a gene that is part of an essential protein complex, is parentally regulated since the paternal allele is not able to compensate for the loss of *APC6* in a maternal null mutant. Furthermore, if one subunit of this complex is maternally expressed and imprinted on the paternal allele and other subunits exhibit parental effects, are any of the other subunits also regulated by genomic imprinting? The consequence of having several subunits of such an essential protein complex regulated by genomic imprinting may be an indication of an underlying selective pressure to maintain such regulation. From an evolutionary perspective, the parental conflict theory postulates that PEGs would enhance cell proliferation. However, this does not correspond to the results found in this thesis, where *APC6* was shown to be preferentially maternally expressed. Furthermore, the transposon defense theory explains imprinting as a side-effect of TE-silencing. No transposable elements were detected in the vicinity of the *APC6* coding region (4.5 kilo base pairs (bp) upstream; The Arabidopsis Information Resource), minimizing the probability that imprinting of *APC6* is a side-effect of TE-silencing. In addition to that, regulation of TE-activity in *A. thaliana* is performed by DNA methyltransferases (Zhang and Jacobsen, 2006), which in this thesis were shown not to alleviate imprinting of *APC6*. Furthermore, other APC/C subunits, such as *APC1*, *APC4*, and *APC11*, exhibit gametophyte maternal effects which could indicate that the dosage of several subunits of this protein complex is regulated by genomic imprinting and it also suggests that expression of the APC/C is under strong maternal control. Overall, the gene dosage theory seems more likely to explain the imprinted regulation of *APC6*, since *APC6* is not fully silenced from the paternal allele, but merely repressed. The question remains why some APC/C subunits are preferentially expressed from the maternal allele, considering that they form one large protein complex. This could indicate that there is some degree of flexibility in the assembly

and composition of the APC/C depending on the tissue-specific expression of subunits (Saleme et al., 2021). It remains to be determined why knock-out of some subunits results in female gametophyte maternal effects or aborted female gametophyte development and knock-out of other subunits in aborted male gametophyte development.

In this thesis, we have successfully analyzed parent-of-origin specific gene expression on GFP-tagged endosperm-specific nuclei isolated using FANS. We have demonstrated that genomic imprinting is temporally and spatially dynamically regulated by the finding of genes that are imprinted at only one seed developmental stage (temporal) or that are imprinted in specific differentiated endosperm domains (spatial). Although we have not investigated the mechanisms behind this dynamic temporal regulation of genomic imprinting, we speculate that in the endosperm, imprinting marks can be established or removed from specific parental alleles to regulate parent-of-origin specific gene expression at different stages of seed development. This mechanism was demonstrated by the alleviation of paternal silencing of *AGL36* in the *A. thaliana* x *A. arenosa* hybrid, but not in the *A. lyrata* x *A. arenosa* hybrid or *A. arenosa* x *A. arenosa*, even though they share the same male parent (Paper I). Such a mechanism must include that a mark must be established in the gametes to allow parental alleles to be distinguished in the endosperm. Then, this mark must act as a recognition beacon to establish allele-specific imprinting. Conversely, imprinting marks could utilize similar signaling pathways to be removed from parental alleles upon seed development to alleviate parent-of-origin specific gene expression at later stages. As a consequence, a mechanism must be present in the endosperm that is able to recognize parental alleles carrying a signaling mark and, based on the results found in this thesis, a key timepoint for this seems to be endosperm cellularization. The RdDM pathway may be a candidate for establishing *de novo* DNA methylation at later stages of seed development, supported by substantial upregulation of *DRM2* in TE1 compared to EE. Investigation of this hypothesis could be accomplished by utilizing the temporally dynamically regulated imprinted genes identified in this thesis. If the parental imprint is removed in mutants of different components of the RdDM pathway, this could provide insight into how imprinting can be dynamically established in the endosperm throughout seed development. Conversely, for the removal of imprinting marks as the endosperm develops, we have shown that *DML2* and *ROS1*, DME-like DNA glycosylases, are substantially higher expressed in the endosperm at later stages of seed development compared to early stages. These are suitable candidates for further investigation to study genes that lose their imprinting mark at later stages in seed development. Furthermore, we have shown that

genes are dynamically imprinted dependent on the spatial endosperm domain. It remains unclear whether cell differentiation drives changes in genomic imprinting or whether a change in imprinting profile leads to cell differentiation. The latter could be established by signaling gradients within the syncytial endosperm resulting in altered imprinting patterns dependent on the location of endosperm nuclei. In order to investigate these aspects, the identified imprinted genes that showed spatial dynamic regulation of genomic imprinting could be utilized to follow their imprinting profile from the syncytial endosperm up to the cellularized stage. In addition, knock-outs of these genes would allow the controlled regulation of monoallelic or biallelic expression and the effect of this could be correlated to endosperm cellularization. If no difference is observed, the imprinting of the gene of interest is not essential for endosperm cellularization. The hypothesis that endosperm cellularization drives spatially dynamic regulation of genomic imprinting could be investigated using mutants that show disrupted endosperm cellularization, such as *agl62*, and determine the imprinting profiles of imprinted genes identified in this thesis.

Collectively, the results of this thesis may suggest a correlation between accession-specific imprinting and accession-specific establishment of the post-zygotic hybrid barrier. Furthermore, we have shown that post-zygotic hybrid barrier is strongly affected by environmental and genetic factors. We suspect that the temperature might affect the regulation of genomic imprinting in a similar manner and therefore a detailed analysis of this is required. Moreover, the results found here suggest that imprinting can be established dynamically in the endosperm and that canonical pathways for the establishment of imprinting do not lend sufficient explanation. Gene redundancy and possible redundancy of different regulatory pathways should be further explored in order to elucidate the exact mechanisms behind dynamic regulation of genomic imprinting.

References

- Allen E, Xie Z, Gustafson AM, Carrington JC** (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**: 207–221
- Alonso-Peral MM, Trigueros M, Sherman B, Ying H, Taylor JM, Peacock WJ, Dennis ES** (2017) Patterns of gene expression in developing embryos of Arabidopsis hybrids. *Plant J* **89**: 927–939
- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, Ribas de Pouplana L, Martínez-Castilla L, Yanofsky MF** (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci U S A* **97**: 5328–5333
- Armenta-Medina A, Gillmor CS, Gao P, Mora-Macias J, Kochian LV, Xiang D, Datla R** (2021) Developmental and genomic architecture of plant embryogenesis: from model plant to crops. *Plant Commun* **2**: 100136
- Attner MA, Amon A** (2012) Control of the mitotic exit network during meiosis. *Mol Biol Cell* **23**: 3122–3132
- Awasthi A, Paul P, Kumar S, Verma SK, Prasad R, Dhaliwal HS** (2012) Abnormal endosperm development causes female sterility in rice insertional mutant OsAPC6. *Plant Sci* **183**: 167–174
- Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, Kukurba KR, Zhang R, Li JB, van der Kooy D, et al** (2015) Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Genet* **47**: 544–549
- Babu Y, Musielak T, Henschen A, Bayer M** (2013) Suspensor length determines developmental progression of the embryo in Arabidopsis. *Plant Physiol* **162**: 1448–1458
- Bajon C, Horlow C, Motamayor JC, Sauvanet A, Robert D** (1999) Megasporogenesis in Arabidopsis thaliana L.: an ultrastructural study. *Sex Plant Reprod* **12**: 99–109
- Barlow DP** (1993) Methylation and imprinting: from host defense to gene regulation? *Science* **260**: 309–310
- Baroux C, Autran D, Gillmor CS, Grimanelli D, Grossniklaus U** (2008) The maternal to zygotic transition in animals and plants. *Cold Spring Harb Symp Quant Biol* **73**: 89–100
- Baroux C, Fransz P, Grossniklaus U** (2004) Nuclear fusions contribute to polyploidization of the gigantic nuclei in the chalazal endosperm of Arabidopsis. *Planta* **220**: 38–46
- Baroux C, Spillane C, Grossniklaus U** (2002) Evolutionary origins of the endosperm in flowering plants. *Genome Biol* **3**: reviews1026
- Barton SC, Surani MA, Norris ML** (1984) Role of paternal and maternal genomes in mouse development. *Nature* **311**: 374–376
- Batista RA, Köhler C** (2020) Genomic imprinting in plants—revisiting existing models. *Genes Dev.* **34**: 24–36
- Batista RA, Moreno-Romero J, Qiu Y, van Boven J, Santos-González J, Figueiredo DD, Köhler C** (2019) The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *Elife* **8**: e50541
- Becker A, Theissen G** (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* **29**: 464–489
- Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada CM, Munoz MD, et al** (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A* **110**: E435–44

References

- Bemer M, Heijmans K, Airoidi C, Davies B, Angenent GC** (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in *Arabidopsis*. *Plant Physiol* **154**: 287–300
- Bemer M, Wolters-Arts M, Grossniklaus U, Angenent GC** (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in *Arabidopsis* ovules. *The Plant Cell* **20**: 2088–2101
- Berger F** (2003) Endosperm: the crossroad of seed development. *Curr Opin Plant Biol* **6**: 42–50
- Berger F, Chaudhury A** (2009) Parental memories shape seeds. *Trends Plant Sci* **14**: 550–556
- Berger F, Grini PE, Schnittger A** (2006) Endosperm: an integrator of seed growth and development. *Curr Opin Plant Biol* **9**: 664–670
- Berger F, Hamamura Y, Ingouff M, Higashiyama T** (2008) Double fertilization—caught in the act. *Trends Plant Sci* **13**: 437–443
- Berger F, Vu TM, Li J, Chen B** (2012) Hypothesis: selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues. *Cold Spring Harb Symp Quant Biol* **77**: 23–29
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE** (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* **3**: e3156
- Bethke PC, Libourel IGL, Aoyama N, Chung Y-Y, Still DW, Jones RL** (2007) The *Arabidopsis* aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. *Plant Physiol* **143**: 1173–1188
- Birchler JA** (1993) Dosage analysis of maize endosperm development. *Annu Rev Genet* **27**: 181–204
- Blevins T, Podicheti R, Mishra V, Marasco M, Wang J, Rusch D, Tang H, Pikaard CS** (2015) Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *Elife* **4**: e09591
- Böhmdorfer G, Sethuraman S, Rowley MJ, Krzyszton M, Rothi MH, Bouzit L, Wierzbicki AT** (2016) Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. *Elife* **5**: e19092
- Boisnard-Lorig C, Colon-Carmona A, Bauch M, Hodge S, Doerner P, Bancharel E, Dumas C, Haseloff J, Berger F** (2001) Dynamic analyses of the expression of the HISTONE::YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. *The Plant Cell* **13**: 495–509
- Bonhomme S, Horlow C, Vezon D, de Laissardière S, Guyon A, Férault M, Marchand M, Bechtold N, Pelletier G** (1998) T-DNA mediated disruption of essential gametophytic genes in *Arabidopsis* is unexpectedly rare and cannot be inferred from segregation distortion alone. *Mol Gen Genet* **260**: 444–452
- Borg M, Berger F** (2015) Chromatin remodelling during male gametophyte development. *Plant J* **83**: 177–188
- Borg M, Brownfield L, Twell D** (2009) Male gametophyte development: a molecular perspective. *J Exp Bot* **60**: 1465–1478
- Borg M, Jacob Y, Susaki D, LeBlanc C, Buendía D, Axelsson E, Kawashima T, Voigt P, Boavida L, Becker J, et al** (2020) Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nat Cell Biol* **22**: 621–629
- Borges F, Martienssen RA** (2015) The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16**: 727–741

- Boscá S, Knauer S, Laux T** (2011) Embryonic development in *Arabidopsis thaliana*: from the zygote division to the shoot meristem. *Front Plant Sci* **2**: 93
- Bozhkov PV, Filonova LH, Suarez MF** (2005) Programmed cell death in plant embryogenesis. *Curr Top Dev Biol* **67**: 135–179
- Braun RE** (2001) Packaging paternal chromosomes with protamine. *Nat Genet* **28**: 10–12
- Brown RC, Lemmon BE, 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 RC, Lemmon BE, Nguyen H, Olsen O-A** (1999) Development of endosperm in *Arabidopsis thaliana*. *Sex Plant Reprod* **12**: 32–42
- Burkart-Waco D, Josefsson C, Dilkes B, Kozloff N, Torjek O, Meyer R, Altmann T, Comai L** (2012) Hybrid incompatibility in *Arabidopsis* is determined by a multiple-locus genetic network. *Plant Physiol* **158**: 801–812
- Burkart-Waco D, Ngo K, Dilkes B, Josefsson C** (2013) Early Disruption of Maternal–Zygotic Interaction and Activation of Defense-Like Responses in *Arabidopsis* Interspecific Crosses. *The Plant Cell* **6**: 2037–2055
- Burkart-Waco D, Ngo K, Lieberman M, Comai L** (2015) Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS One* **10**: e0117293
- Bushell C, Spielman M, Scott RJ** (2003) The basis of natural and artificial postzygotic hybridization barriers in *Arabidopsis* species. *The Plant Cell* **15**: 1430–1442
- Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, et al** (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151**: 194–205
- Cao X, Jacobsen SE** (2002) Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**: 1138–1144
- Cao X, Springer NM, Muszynski MG, Phillips RL, Kaeppler S, Jacobsen SE** (2000) Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc Natl Acad Sci U S A* **97**: 4979–4984
- Capron A, Serralbo O, Fülöp K, Frugier F, Parmentier Y, Dong A, Lecureuil A, Guerche P, Kondorosi E, Scheres B, et al** (2003) The *Arabidopsis* anaphase-promoting complex or cyclosome: molecular and genetic characterization of the APC2 subunit. *The Plant Cell* **15**: 2370–2382
- Chan SW-L, Henderson IR, Jacobsen SE** (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat Rev Genet* **6**: 351–360
- Chang DC, Xu N, Luo KQ** (2003) Degradation of cyclin B is required for the onset of anaphase in Mammalian cells. *J Biol Chem* **278**: 37865–37873
- Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon Y-H, Sung ZR, Goodrich J** (2004) Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* **131**: 5263–5276
- Chaudhury AM, Berger F** (2001) Maternal control of seed development. *Semin Cell Dev Biol* **12**: 381–386
- Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ** (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **94**: 4223–4228
- Chen C, Begcy K, Liu K, Folsom JJ, Wang Z, Zhang C, Walia H** (2016) Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol* **171**: 606–622

References

- Cheng C-Y, Mathews DE, Schaller GE, Kieber JJ** (2013) Cytokinin-dependent specification of the functional megaspore in the Arabidopsis female gametophyte. *Plant J* **73**: 929–940
- Chettoor AM, Evans MMS** (2015) Correlation between a loss of auxin signaling and a loss of proliferation in maize antipodal cells. *Front Plant Sci* **6**: 187
- Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer RL** (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* **110**: 33–42
- Christensen CA, King EJ, Jordan JR, Drews GN** (1997) Megagametogenesis in Arabidopsis wild type and the Gf mutant. *Sex Plant Reprod* **10**: 49–64
- Colombo L, Franken J, Van der Krol AR, Wittich PE, Dons HJM, Angenent GC** (1997) Downregulation of Ovule-Specific MADS Box Genes from Petunia Results in Maternally Controlled Defects in Seed Development. *The Plant Cell* **9**: 703
- Colombo M, Masiero S, Vanzulli S, Lardelli P, Kater MM, Colombo L** (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J* **54**: 1037–1048
- Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B** (2000) Phenotypic instability and rapid gene silencing in newly formed arabidopsis allotetraploids. *The Plant Cell* **12**: 1551–1568
- Costa LM, Marshall E, Tesfaye M, Silverstein KAT, Mori M, Umetsu Y, Otterbach SL, Papareddy R, Dickinson HG, Boutiller K, et al** (2014) Central Cell-Derived Peptides Regulate Early Embryo Patterning in Flowering Plants. *Science* **344**: 168–172
- Costa-Nunes P, Kim JY, Hong E, Pontes O** (2014) The cytological and molecular role of domains rearranged methyltransferase3 in RNA-dependent DNA methylation of Arabidopsis thaliana. *BMC Res Notes* **7**: 721
- Coughlan JM, Matute DR** (2020) The importance of intrinsic postzygotic barriers throughout the speciation process. *Philos Trans R Soc Lond B Biol Sci* **375**: 20190533
- Cromer L, Heyman J, Touati S, Harashima H, Araou E, Girard C, Horlow C, Wassmann K, Schnittger A, De Veylder L, et al** (2012) OSD1 promotes meiotic progression via APC/C inhibition and forms a regulatory network with TDM and CYCA1;2/TAM. *PLoS Genet* **8**: e1002865
- Cromer L, Jolivet S, Singh DK, Berthier F, De Winne N, De Jaeger G, Komaki S, Prusicki MA, Schnittger A, Guérois R, et al** (2019) Patronus is the elusive plant securin, preventing chromosome separation by antagonizing separase. *Proc Natl Acad Sci U S A* **116**: 16018–16027
- Cuerda-Gil D, Slotkin RK** (2016) Non-canonical RNA-directed DNA methylation. *Nat Plants* **2**: 16163
- Deal RB, Henikoff S** (2011) The INTACT method for cell type-specific gene expression and chromatin profiling in Arabidopsis thaliana. *Nat Protoc* **6**: 56–68
- Debeaujon I, Léon-Kloosterziel KM, Koornneef M** (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol* **122**: 403–414
- DeChiara TM, Robertson EJ, Efstratiadis A** (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**: 849–859
- Del Toro-De León G, Köhler C** (2019) Endosperm-specific transcriptome analysis by applying the INTACT system. *Plant Reprod* **32**: 55–61
- Denay G, Creff A, Moussu S, Wagnon P, Thévenin J, Gérentes M-F, Chambrier P, Dubreucq B, Ingram G** (2014) Endosperm breakdown in Arabidopsis requires heterodimers of the basic helix-loop-helix proteins ZHOUP1 and INDUCER OF CBP EXPRESSION 1. *Development* **141**: 1222–1227

- Dilkes BP, Comai L** (2004) A differential dosage hypothesis for parental effects in seed development. *The Plant Cell* **16**: 3174–3180
- Doll NM, Ingram GC** (2022) Embryo-Endosperm Interactions. *Annu Rev Plant Biol* **73**: 293–321
- Doll NM, Royek S, Fujita S, Okuda S, Chamot S, Stintzi A, Widiez T, Hothorn M, Schaller A, Geldner N, et al** (2020) A two-way molecular dialogue between embryo and endosperm is required for seed development. *Science* **367**: 431–435
- Domínguez F, Cejudo FJ** (2014) Programmed cell death (PCD): an essential process of cereal seed development and germination. *Front Plant Sci* **5**: 366
- Dong X, Zhang M, Chen J, Peng L, Zhang N, Wang X, Lai J** (2017) Dynamic and Antagonistic Allele-Specific Epigenetic Modifications Controlling the Expression of Imprinted Genes in Maize Endosperm. *Mol Plant* **10**: 442–455
- Drews GN, Lee D, Christensen CA** (1998) Genetic analysis of female gametophyte development and function. *The Plant Cell* **10**: 5–17
- Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, et al** (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**: 167–180
- Eloy NB, de Freitas Lima M, Van Damme D, Vanhaeren H, Gonzalez N, De Milde L, Hemerly AS, Beemster GTS, Inzé D, Ferreira PCG** (2011) The APC/C subunit 10 plays an essential role in cell proliferation during leaf development. *Plant J* **68**: 351–363
- Erdmann RM, Picard CL** (2020) RNA-directed DNA Methylation. *PLoS Genet* **16**: e1009034
- Erdmann RM, Satyaki PRV, Klosinska M, Gehring M** (2017) A Small RNA Pathway Mediates Allelic Dosage in Endosperm. *Cell Reports* **21**: 3364–3372
- d’Erfurth I, Cromer L, Jolivet S, Girard C, Horlow C, Sun Y, To JPC, Berchowitz LE, Copenhaver GP, Mercier R** (2010) The cyclin-A CYCA1;2/TAM is required for the meiosis I to meiosis II transition and cooperates with OSD1 for the prophase to first meiotic division transition. *PLoS Genet* **6**: e1000989
- Evans MM, Kermicle JL** (2001) Interaction between maternal effect and zygotic effect mutations during maize seed development. *Genetics* **159**: 303–315
- Faure J-E, Rotman N, Fortuné P, Dumas C** (2002) Fertilization in *Arabidopsis thaliana* wild type: developmental stages and time course. *Plant J* **30**: 481–488
- Feil R, Berger F** (2007) Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet* **23**: 192–199
- Feldmann KA, Coury DA, Christianson ML** (1997) Exceptional segregation of a selectable marker (KanR) in *Arabidopsis* identifies genes important for gametophytic growth and development. *Genetics* **147**: 1411–1422
- Feng S, Jacobsen SE, Reik W** (2010) Epigenetic reprogramming in plant and animal development. *Science* **330**: 622–627
- Ferguson-Smith AC** (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* **12**: 565–575
- Ferrón SR, Charalambous M, Radford E, McEwen K, Wildner H, Hind E, Morante-Redolat JM, Laborda J, Guillemot F, Bauer SR, et al** (2011) Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* **475**: 381–385
- Figueiredo DD, Batista RA, Roszak PJ, Hennig L, Köhler C** (2016) Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *Elife* **5**: e20542

References

- Finnegan EJ, Dennis ES** (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res* **21**: 2383–2388
- Florez-Rueda AM, Fiscalini F, Roth M, Grossniklaus U, Städler T** (2021) Endosperm and Seed Transcriptomes Reveal Possible Roles for Small RNA Pathways in Wild Tomato Hybrid Seed Failure. *Genome Biol Evol* **13**: evab107
- Florez-Rueda AM, Paris M, Schmidt A, Widmer A, Grossniklaus U, Städler T** (2016) Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive Hybrid Tomato Seeds. *Mol Biol Evol* **33**: 2935–2946
- Folsom JJ, Begcy K, Hao X, Wang D, Walia H** (2014) Rice fertilization-Independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol* **165**: 238–248
- de Folter S, Immink RGH, Kieffer M, Parenicová L, Henz SR, Weigel D, Busscher M, Kooiker M, Colombo L, Kater MM, et al** (2005) Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *The Plant Cell* **17**: 1424–1433
- Fort A, Tuteja R, Braud M, McKeown PC, Spillane C** (2017) Parental-genome dosage effects on the transcriptome of F1 hybrid triploid embryos of *Arabidopsis thaliana*. *Plant J* **92**: 1044–1058
- Gallego-Bartolomé J, Liu W, Kuo PH, Feng S, Ghoshal B, Gardiner J, Zhao JM-C, Park SY, Chory J, Jacobsen SE** (2019) Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in *Arabidopsis*. *Cell* **176**: 1068-1082.e19
- Gavazzi G, Dolfini S, Allegra D, Castiglioni P, Todesco G, Hoxha M** (1997) Dap (D effective a leuone p igmentation) mutations affect maize aleurone development. *Mol Gen Genet* **256**: 223–230
- Gehring M** (2013) Genomic imprinting: insights from plants. *Annu Rev Genet* **47**: 187–208
- Gehring M, Bubb KL, Henikoff S** (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**: 1447–1451
- Gehring M, Missirian V, Henikoff S** (2011) Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PloS One* **6**: e23687
- Gehring M, Satyaki PR** (2017) Endosperm and Imprinting, Inextricably Linked. *Plant Physiol* **173**: 143–154
- Gepts P** (2004) Crop domestication as a long-term selection experiment. *Plant Breed Rev* **24**: 1–44
- Goldberg RB, de Paiva G, Yadegari R** (1994) Plant embryogenesis: zygote to seed. *Science* **266**: 605–614
- Gramzow L, Theissen G** (2010) A hitchhiker’s guide to the MADS world of plants. *Genome Biol* **11**: 214
- Grimanelli D, Perotti E, Ramirez J, Leblanc O** (2005) Timing of the Maternal-to-Zygotic Transition during Early Seed Development in Maize. *The Plant Cell* **17**: 1061–1072
- Grini PE, Jürgens G, Hülskamp M** (2002) Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of *Arabidopsis*. *Genetics* **162**: 1911–1925
- Grini PE, Schnittger A, Schwarz H, Zimmermann I, Schwab B, Jürgens G, Hülskamp M** (1999) Isolation of ethyl methanesulfonate-induced gametophytic mutants in *Arabidopsis thaliana* by a segregation distortion assay using the multimarker chromosome 1. *Genetics* **151**: 849–863
- Grossniklaus U, Schneitz K** (1998) The molecular and genetic basis of ovule and megagametophyte development. *Semin Cell Dev Biol* **9**: 227–238
- Grossniklaus U, Vielle-Calzada JP, Hoepfner MA, Gagliano WB** (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**: 446–450

- Guitton A-E, Page DR, Chambrier P, Lionnet C, Faure J-E, Grossniklaus U, 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
- Guo L, Jiang L, Zhang Y, Lu X-L, Xie Q, Weijers D, Liu C-M** (2016) The anaphase-promoting complex initiates zygote division in *Arabidopsis* through degradation of cyclin B1. *Plant J* **86**: 161–174
- Gutiérrez-Marcos JF, Costa LM, Evans MMS** (2006) Maternal gametophytic *baseless1* is required for development of the central cell and early endosperm patterning in maize (*Zea mays*). *Genetics* **174**: 317–329
- Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS** (2012) In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. *Mol Cell* **48**: 811–818
- Haig D** (2014) Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity* **113**: 96–103
- Haig D, Westoby M** (1989) Parent-Specific Gene Expression and the Triploid Endosperm. *Am Nat* **134**: 147–155
- Hamamura Y, Saito C, Awai C, Kurihara D, Miyawaki A, Nakagawa T, Kanaoka MM, Sasaki N, Nakano A, Berger F, et al** (2011) Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in *Arabidopsis thaliana*. *Curr Biol* **21**: 497–502
- Hanna CW** (2020) Placental imprinting: Emerging mechanisms and functions. *PLoS Genet* **16**: e1008709
- Harris KD, Zemach A** (2020) Contiguous and stochastic CHH methylation patterns of plant DRM2 and CMT2 revealed by single-read methylome analysis. *Genome Biol* **21**: 194
- Hatorangan MR, Laenen B, Steige KA, Slotte T, Köhler C** (2016) Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. *The Plant Cell* **28**: 1815–1827
- Havecker ER, Wallbridge LM, Harcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC** (2010) The *Arabidopsis* RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *The Plant Cell* **22**: 321–334
- He L, Huang H, Bradai M, Zhao C, You Y, Ma J, Zhao L, Lozano-Durán R, Zhu J-K** (2022) DNA methylation-free *Arabidopsis* reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat Commun* **13**: 1335
- Hehenberger E, Kradolfer D, Köhler C** (2012) Endosperm cellularization defines an important developmental transition for embryo development. *Development* **139**: 2031–2039
- Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, Horwitz GA, Kelly KA, Pradhan S, Jacobsen SE** (2010) The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genet* **6**: e1001182
- Henikoff S, Comai L** (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics* **149**: 307–318
- Hornslieen KS, Miller JR, Grini PE** (2019) Regulation of Parent-of-Origin Allelic Expression in the Endosperm. *Plant Physiol* **180**: 1498–1519
- Horst NA, Reski R** (2016) Alternation of generations - unravelling the underlying molecular mechanism of a 165-year-old botanical observation. *Plant Biol* **18**: 549–551
- Howden R, Park SK, Moore JM, Orme J, Grossniklaus U, Twell D** (1998) Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in *Arabidopsis*. *Genetics* **149**: 621–631

References

- Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D** (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* **324**: 1451–1454
- Hsieh T-F, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, et al** (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci U S A* **108**: 1755–1762
- Huang F, Zhu Q-H, Zhu A, Wu X, Xie L, Wu X, Helliwell C, Chaudhury A, Finnegan EJ, Luo M** (2017) Mutants in the imprinted PICKLE RELATED 2 gene suppress seed abortion of fertilization independent seed class mutants and paternal excess interploidy crosses in Arabidopsis. *Plant J* **90**: 383–395
- Huetzel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJM, Matzke M** (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* **25**: 2828–2836
- Hulskamp M, Schneitz K, Pruitt RE** (1995) Genetic Evidence for a Long-Range Activity That Directs Pollen Tube Guidance in Arabidopsis. *The Plant Cell* **7**: 57–64
- Ibarra CA, Feng X, Schoft VK, Hsieh T-F, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, et al** (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360–1364
- Ingouff M, Fitz Gerald JN, Guérin C, Robert H, Sørensen MB, Van Damme D, Geelen D, Blanchoin L, Berger F** (2005) Plant formin AtFH5 is an evolutionarily conserved actin nucleator involved in cytokinesis. *Nat Cell Biol* **7**: 374–380
- Ingouff M, Jullien PE, 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
- Ingram GC** (2020) Family plot: the impact of the endosperm and other extra-embryonic seed tissues on angiosperm zygotic embryogenesis. *F1000Research* **9**: 18
- Iwasaki M, Hyvärinen L, Piskurewicz U, Lopez-Molina L** (2019) Non-canonical RNA-directed DNA methylation participates in maternal and environmental control of seed dormancy. *Elife* **8**: e37434
- Jahnke S, Scholten S** (2009) Epigenetic resetting of a gene imprinted in plant embryos. *Curr Biol* **19**: 1677–1681
- Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ, et al** (2014) SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**: 124–128
- Johnston DS, St Johnston D, Nüsslein-Volhard C** (1992) The origin of pattern and polarity in the Drosophila embryo. *Cell* **68**: 201–219
- Johnston SA, den Nijs TP, Peloquin SJ, Hanneman RE Jr** (1980) The significance of genic balance to endosperm development in interspecific crosses. *Theor Appl Genet* **57**: 5–9
- Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK** (2011) Interspecific and interploidal gene flow in Central European Arabidopsis (Brassicaceae). *BMC Evol Biol* **11**: 346
- Josefsson C, Dilkes B, Comai L** (2006) Parent-dependent loss of gene silencing during interspecies hybridization. *Curr Biol* **16**: 1322–1328
- Jullien PE, Kinoshita T, Ohad N, Berger F** (2006) Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *The Plant Cell* **18**: 1360–1372
- Jullien PE, Mosquna A, Ingouff M, Sakata T, Ohad N, Berger F** (2008) Retinoblastoma and its binding partner MS11 control imprinting in Arabidopsis. *PLoS Biol* **6**: e194
- Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F** (2012) DNA Methylation Dynamics during Sexual Reproduction in Arabidopsis thaliana. *Current Biology* **22**: 1825–1830

- Kang I-H, Steffen JG, Portereiko MF, Lloyd A, Drews GN** (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *The Plant Cell* **20**: 635–647
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ** (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**: 1109–1122
- Kaufmann K, Melzer R, Theissen G** (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* **347**: 183–198
- Kawashima T, Goldberg RB** (2010) The suspensor: not just suspending the embryo. *Trends Plant Sci* **15**: 23–30
- Kessler SA, Grossniklaus U** (2011) She's the boss: signaling in pollen tube reception. *Curr Opin Plant Biol* **14**: 622–627
- Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T** (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **303**: 521–523
- Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL** (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *The Plant Cell* **11**: 1945–1952
- Kirkbride RC, Lu J, Zhang C, Mosher RA, Baulcombe DC, Chen ZJ** (2019) Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in Arabidopsis. *Proc Natl Acad Sci U S A* **116**: 2761–2766
- Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB, 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
- Klosinska M, Picard CL, Gehring M** (2016) Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. *Nat Plants* **2**: 16145
- Köhler C, Dziasek K, Del Toro-De León G** (2021) Postzygotic reproductive isolation established in the endosperm: mechanisms, drivers and relevance. *Philos Trans R Soc Lond B Biol Sci* **376**: 20200118
- Köhler C, Hennig L** (2010) Regulation of cell identity by plant Polycomb and trithorax group proteins. *Curr Opin Genet Dev* **20**: 541–547
- Köhler C, Hennig L, Bouveret R, Gheyselinck J, Grossniklaus U, Griessem W** (2003a) Arabidopsis MS1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO J* **22**: 4804–4814
- Köhler C, Hennig L, Spillane C, Pien S, Griessem W, Grossniklaus U** (2003b) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev* **17**: 1540–1553
- Köhler C, Page DR, Gagliardini V, Grossniklaus U** (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat Genet* **37**: 28–30
- Kondou Y, Nakazawa M, Kawashima M, Ichikawa T, Yoshizumi T, Suzuki K, Ishikawa A, Koshi T, Matsui R, Muto S, et al** (2008) RETARDED GROWTH OF EMBRYO1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. *Plant Physiol* **147**: 1924–1935
- Kradolfer D, Wolff P, Jiang H, Siretskiy A, Köhler C** (2013) An imprinted gene underlies postzygotic reproductive isolation in Arabidopsis thaliana. *Dev Cell* **26**: 525–535
- Kwee H-S, Sundaresan V** (2003) The NOMEA gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in Arabidopsis. *Plant J* **36**: 853–866

References

- Lafon-Placette C, Hatorangan MR, Steige KA, Cornille A, Lascoux M, Slotte T, Köhler C** (2018) Paternally expressed imprinted genes associate with hybridization barriers in *Capsella*. *Nat Plants* **4**: 352–357
- Lafon-Placette C, Johannessen IM, Hornslien KS, Ali MF, Bjerkan KN, Bramsiepe J, Glöckle BM, Rebernik CA, Brysting AK, Grini PE, et al** (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc Natl Acad Sci U S A* **114**: E1027–E1035
- Lafon-Placette C, Köhler C** (2014) Embryo and endosperm, partners in seed development. *Curr Opin Plant Biol* **17**: 64–69
- Lara-Gonzalez P, Westhorpe FG, Taylor SS** (2012) The spindle assembly checkpoint. *Curr Biol* **22**: R966–80
- Law JA, Jacobsen SE** (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* **11**: 204–220
- Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE** (2011) SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet* **7**: e1002195
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, et al** (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci U S A* **107**: 8063–8070
- Leblanc O, Pointe C, Hernandez M** (2002) Cell cycle progression during endosperm development in *Zea mays* depends on parental dosage effects. *Plant J* **32**: 1057–1066
- Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM** (1995) Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**: 34–39
- Leydon AR, Tsukamoto T, Dunatunga D, Qin Y, Johnson MA, Palanivelu R** (2015) Pollen Tube Discharge Completes the Process of Synergid Degeneration That Is Initiated by Pollen Tube-Synergid Interaction in *Arabidopsis*. *Plant Physiol* **169**: 485–496
- Li J, Berger F** (2012) Endosperm: food for humankind and fodder for scientific discoveries. *New Phytol* **195**: 290–305
- Li W, Ma H** (2002) Gametophyte development. *Curr Biol* **12**: R718–21
- Li YJ, Yu Y, Liu X, Zhang XS, Su YH** (2021) The *Arabidopsis* MATERNAL EFFECT EMBRYO ARREST45 protein modulates maternal auxin biosynthesis and controls seed size by inducing AINTEGUMENTA. *The Plant Cell* **33**: 1907–1926
- Liang W, Li J, Sun L, Liu Y, Lan Z, Qian W** (2022) Deciphering the synergistic and redundant roles of CG and non-CG DNA methylation in plant development and transposable element silencing. *New Phytol* **233**: 722–737
- Lindahl T** (1986) DNA glycosylases in DNA repair. *Basic Life Sci* **38**: 335–340
- Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE** (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* **292**: 2077–2080
- Liu J, Li J, Liu H-F, Fan S-H, Singh S, Zhou X-R, Hu Z-Y, Wang H-Z, Hua W** (2018a) Genome-wide screening and analysis of imprinted genes in rapeseed (*Brassica napus* L.) endosperm. *DNA Res* **25**: 629–640
- Liu J, Wu X, Yao X, Yu R, Larkin PJ, Liu C-M** (2018b) Mutations in the DNA demethylase OsROS1 result in a thickened aleurone and improved nutritional value in rice grains. *Proc Natl Acad Sci U S A* **115**: 11327–11332

- Liu Z-W, Shao C-R, Zhang C-J, Zhou J-X, Zhang S-W, Li L, Chen S, Huang H-W, Cai T, He X-J** (2014) The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. *PLoS Genet* **10**: e1003948
- Long J, Walker J, She W, Aldridge B, Gao H, Deans S, Vickers M, Feng X** (2021) Nurse cell--derived small RNAs define paternal epigenetic inheritance in Arabidopsis. *Science* **373**: eabh0556
- Lu J, Zhang C, Baulcombe DC, Chen ZJ** (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. *Proc Natl Acad Sci U S A* **109**: 5529–5534
- Luo A, Shi C, Zhang L, Sun M-X** (2014) The expression and roles of parent-of-origin genes in early embryogenesis of angiosperms. *Front Plant Sci* **5**: 729
- Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A** (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci U S A* **97**: 10637–10642
- Luo M, Bilodeau P, Koltunow A, Dennis ES, Peacock WJ, Chaudhury AM** (1999) Genes controlling fertilization-independent seed development in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **96**: 296–301
- Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A** (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet* **7**: e1002125
- Mabb AM, Judson MC, Zylka MJ, Philpot BD** (2011) Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends Neurosci* **34**: 293–303
- Magee DA, Spillane C, Berkowicz EW, Sikora KM, MacHugh DE** (2014) Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Anim Genet* **45 Suppl 1**: 25–39
- Makarevich G, Leroy O, Akinci U, Schubert D, Clarenz O, Goodrich J, Grossniklaus U, Köhler C** (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis. *EMBO Rep* **7**: 947–952
- Makarevich G, Villar CBR, Erilova A, Köhler C** (2008) Mechanism of PHERES1 imprinting in Arabidopsis. *J Cell Sci* **121**: 906–912
- Margueron R, Reinberg D** (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469**: 343–349
- Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM** (2011) The Emerging Importance of Type I MADS Box Transcription Factors for Plant Reproduction. *The Plant Cell* **23**: 865–872
- Mathieu O, Reinders J, Caikovski M, Smathajitt C, Paszkowski J** (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell* **130**: 851–862
- Matzke MA, Moshier RA** (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet* **15**: 394–408
- McCormick S** (1993) Male Gametophyte Development. *The Plant Cell* **5**: 1265–1275
- McDonald JF, Matzke MA, Matzke AJ** (2005) Host defenses to transposable elements and the evolution of genomic imprinting. *Cytogenet Genome Res* **110**: 242–249
- McGinty RK, Tan S** (2015) Nucleosome structure and function. *Chem Rev* **115**: 2255–2273
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O** (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427–430

References

- Montgomery SA, Berger F** (2021) The evolution of imprinting in plants: beyond the seed. *Plant Reprod* **34**: 373–383
- Moore T, Haig D** (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* **7**: 45–49
- Moreno-Romero J, Del Toro-De León G, Yadav VK, Santos-González J, Köhler C** (2019) Epigenetic signatures associated with imprinted paternally expressed genes in the Arabidopsis endosperm. *Genome Biol* **20**: 41
- Moreno-Romero J, Jiang H, Santos-González J, Köhler C** (2016) Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *EMBO J* **35**: 1298–1311
- Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC** (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* **460**: 283–286
- Moukhtar J, Trubuil A, Belcram K, Legland D, Khadir Z, Urbain A, Palauqui J-C, Andrey P** (2019) Cell geometry determines symmetric and asymmetric division plane selection in Arabidopsis early embryos. *PLoS Comput Biol* **15**: e1006771
- Muir G, Ruiz-Duarte P, Hohmann N, Mable BK, Novikova P, Schmickl R, Guggisberg A, Koch MA** (2015) Exogenous selection rather than cytonuclear incompatibilities shapes asymmetrical fitness of reciprocal Arabidopsis hybrids. *Ecology and Evolution* **5**: 1734–1745
- Musacchio A** (2015) The Molecular Biology of Spindle Assembly Checkpoint Signaling Dynamics. *Curr Biol* **25**: R1002-18
- Newell-Price J, Clark AJ, King P** (2000) DNA methylation and silencing of gene expression. *Trends Endocrinol Metab* **11**: 142–148
- Nguyen CD, Chen J, Clark D, Perez H, Huo H (alfred)** (2021) Effects of Maternal Environment on Seed Germination and Seedling Vigor of *Petunia* × hybrid under Different Abiotic Stresses. *Plants* **10**: 581
- Nguyen H, Brown RC, Lemmon BE** (2000) The specialized chalazal endosperm in *Arabidopsis thaliana* and *Lepidium virginicum* (Brassicaceae). *Protoplasma* **212**: 99–110
- Nicholls RD, Knoll JHM, Butler MG, Karam S, Lalande M** (1989) Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *Nature* **342**: 281–285
- Nodine MD, Bartel DP** (2012) Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**: 94–97
- Nowack MK, Ungru A, Bjerkan KN, Grini PE, Schnittger A** (2010) Reproductive cross-talk: seed development in flowering plants. *Biochem Soc Trans* **38**: 604–612
- Nuthikattu S, McCue AD, Panda K, Fultz D, DeFraia C, Thomas EN, Slotkin RK** (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* **162**: 116–131
- Ohad N, Yadegari R, Margossian L, Hannon M, Michaeli D, Harada JJ, Goldberg RB, Fischer RL** (1999) Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *The Plant Cell* **11**: 407–416
- Olsen O-A** (2004) Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *The Plant Cell* **16 Suppl**: S214-27
- Olvera-Carrillo Y, Van Bel M, Van Hautegeem T, Fendrych M, Huysmans M, Simaskova M, van Durme M, Buscaill P, Rivas S, Coll NS, et al** (2015) A Conserved Core of Programmed Cell Death Indicator Genes Discriminates Developmentally and Environmentally Induced Programmed Cell Death in Plants. *Plant Physiol* **169**: 2684–2699

- Opsahl-Ferstad HG, Le Deunff E, Dumas C, Rogowsky PM** (1997) ZmEsr, a novel endosperm-specific gene expressed in a restricted region around the maize embryo. *Plant J* **12**: 235–246
- Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V** (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* **132**: 603–614
- Papareddy RK, Páldi K, Smolka AD, Hüther P, Becker C, Nodine MD** (2021) Repression of CHROMOMETHYLASE 3 prevents epigenetic collateral damage in *Arabidopsis*. *Elife* **10**: e69396
- Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, et al** (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *The Plant Cell* **15**: 1538–1551
- Park K, Lee S, Yoo H, Choi Y** (2020) DEMETER-mediated DNA Demethylation in Gamete Companion Cells and the Endosperm, and its Possible Role in Embryo Development in *Arabidopsis*. *J Plant Biol* **63**: 321–329
- Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R, Wolf JB** (2014) The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**: 119–128
- Paul P, Dhatt BK, Miller M, Folsom JJ, Wang Z, Krassovskaya I, Liu K, Sandhu J, Yu H, Zhang C, et al** (2020) MADS78 and MADS79 Are Essential Regulators of Early Seed Development in Rice. *Plant Physiol* **182**: 933–948
- Pecinka A, Chevalier C, Colas I, Kalantidis K, Varotto S, Krugman T, Michailidis C, Vallés M-P, Muñoz A, Pradillo M** (2020) Chromatin dynamics during interphase and cell division: similarities and differences between model and crop plants. *J Exp Bot* **71**: 5205–5222
- Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL** (2007) DNA demethylation in the *Arabidopsis* genome. *Proc Natl Acad Sci U S A* **104**: 6752–6757
- Pérez-Pérez JM, Serralbo O, Vanstraelen M, González C, Criqui M-C, Genschik P, Kondorosi E, Scheres B** (2008) Specialization of CDC27 function in the *Arabidopsis thaliana* anaphase-promoting complex (APC/C). *Plant J* **53**: 78–89
- Petrella R, Cucinotta M, Mendes MA, Underwood CJ, Colombo L** (2021) The emerging role of small RNAs in ovule development, a kind of magic. *Plant Reprod* **34**: 335–351
- Picard CL, Povilus RA, Williams BP, Gehring M** (2021) Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. *Nat Plants* **7**: 730–738
- Pignatta D, Erdmann RM, Scheer E, Picard CL, Bell GW, Gehring M** (2014) Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *Elife* **3**: e03198
- Pignatta D, Novitzky K, Satyaki PRV, Gehring M** (2018) A variably imprinted epiallele impacts seed development. *PLoS Genet* **14**: e1007469
- Pignocchi C, Minns GE, Nesi N, Koumproglou R, Kitsios G, Benning C, Lloyd CW, Doonan JH, Hills MJ** (2009) ENDOSPERM DEFECTIVE1 Is a Novel Microtubule-Associated Protein Essential for Seed Development in *Arabidopsis*. *The Plant Cell* **21**: 90–105
- Pires ND** (2014) Seed evolution: parental conflicts in a multi-generational household. *Biomol Concepts* **5**: 71–86
- Pires ND, Grossniklaus U** (2014) Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep* **6**: 63
- Punwani JA, Drews GN** (2008) Development and function of the synergid cell. *Sex Plant Reprod* **21**: 7–15
- Qi Y, Denli AM, Hannon GJ** (2005) Biochemical Specialization within *Arabidopsis* RNA Silencing Pathways. *Molecular Cell* **19**: 421–428

References

- Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ** (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008–1012
- Raissig MT, Bemer M, Baroux C, Grossniklaus U** (2013) Genomic imprinting in the Arabidopsis embryo is partly regulated by PRC2. *PLoS Genet* **9**: e1003862
- Ray S, Golden T, Ray A** (1996) Maternal effects of the short integument mutation on embryo development in Arabidopsis. *Dev Biol* **180**: 365–369
- Rebernik CA, Lafon-Placette C, Hatorangan MR, Slotte T, Köhler C** (2015) Non-reciprocal Interspecies Hybridization Barriers in the Capsella Genus Are Established in the Endosperm. *PLoS Genet* **11**: e1005295
- Reik W, Dean W, Walter J** (2001) Epigenetic reprogramming in mammalian development. *Science* **293**: 1089–1093
- Reiser L, Fischer RL** (1993) The Ovule and the Embryo Sac. *The Plant Cell* **5**: 1291–1301
- Rieseberg LH, Willis JH** (2007) Plant speciation. *Science* **317**: 910–914
- Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D** (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci U S A* **110**: 7934–7939
- Rodrigues JA, Zilberman D** (2015) Evolution and function of genomic imprinting in plants. *Genes Dev* **29**: 2517–2531
- Rong H, Yang W, Zhu H, Jiang B, Jiang J, Wang Y** (2021) Genomic imprinted genes in reciprocal hybrid endosperm of Brassica napus. *BMC Plant Biol* **21**: 140
- Roszak P, Köhler C** (2011) Polycomb group proteins are required to couple seed coat initiation to fertilization. *Proc Natl Acad Sci U S A* **108**: 20826–20831
- Roth M, Florez-Rueda AM, Paris M, Städler T** (2018) Wild tomato endosperm transcriptomes reveal common roles of genomic imprinting in both nuclear and cellular endosperm. *Plant J* **95**: 1084–1101
- Roth M, Florez-Rueda AM, Städler T** (2019) Differences in Effective Ploidy Drive Genome-Wide Endosperm Expression Polarization and Seed Failure in Wild Tomato Hybrids. *Genetics* **212**: 141–152
- Rotondo JC, Selvatici R, Di Domenico M, Marci R, Vesce F, Tognon M, Martini F** (2013) Methylation loss at H19 imprinted gene correlates with methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples from infertile males. *Epigenetics* **8**: 990–997
- Russell SD** (1992) Double Fertilization. *International Review of Cytology* **140**: 357–388
- Saleme M de LS, Andrade IR, Eloy NB** (2021) The Role of Anaphase-Promoting Complex/Cyclosome (APC/C) in Plant Reproduction. *Front Plant Sci* **12**: 642934
- Sandaklie-Nikolova L, Palanivelu R, King EJ, Copenhaver GP, Drews GN** (2007) Synergid cell death in Arabidopsis is triggered following direct interaction with the pollen tube. *Plant Physiol* **144**: 1753–1762
- Satyaki PRV, Gehring M** (2017) DNA methylation and imprinting in plants: machinery and mechanisms. *Crit Rev Biochem Mol Biol* **52**: 163–175
- Satyaki PRV, Gehring M** (2019) Paternally Acting Canonical RNA-Directed DNA Methylation Pathway Genes Sensitize Arabidopsis Endosperm to Paternal Genome Dosage. *The Plant Cell* **31**: 1563–1578
- Sazhenova EA, Lebedev IN** (2021) Evolutionary Aspects of Genomic Imprinting. *Mol Biol* **55**: 1–15

- Schoft VK, Chumak N, Choi Y, Hannon M, Garcia-Aguilar M, Machlicova A, Slusarz L, Mosiolek M, Park J-S, Park GT, et al** (2011) Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proc Natl Acad Sci U S A* **108**: 8042–8047
- Schon MA, Nodine MD** (2017) Widespread Contamination of Arabidopsis Embryo and Endosperm Transcriptome Data Sets. *The Plant Cell* **29**: 608–617
- Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H** (1990) Genetic Control of Flower Development by Homeotic Genes in *Antirrhinum majus*. *Science* **250**: 931–936
- Scott RJ, Spielman M, Bailey J, Dickinson HG** (1998) Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**: 3329–3341
- Sekine D, Ohnishi T, Furuumi H, Ono A, Yamada T, Kurata N, Kinoshita T** (2013) Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J* **76**: 792–799
- Shirzadi R, Andersen ED, Bjerkan KN, Gloeckle BM, Heese M, Ungru A, Winge P, Koncz C, Aalen RB, Schnittger A, et al** (2011) Genome-Wide Transcript Profiling of Endosperm without Paternal Contribution Identifies Parent-of-Origin-Dependent Regulation of AGAMOUS-LIKE36. *PLoS Genet* **7**: e1001303
- Sigman MJ, Panda K, Kirchner R, McLain LL, Payne H, Peasari JR, Husbands AY, Slotkin RK, McCue AD** (2021) An siRNA-guided ARGONAUTE protein directs RNA polymerase V to initiate DNA methylation. *Nat Plants* **7**: 1461–1474
- Sigman MJ, Slotkin RK** (2016) The First Rule of Plant Transposable Element Silencing: Location, Location, Location. *The Plant Cell* **28**: 304–313
- Skinner DJ, Sundaesan V** (2018) Recent advances in understanding female gametophyte development. *F1000Research* **7**: 804
- Spillane C, MacDougall C, Stock C, Köhler C, Vielle-Calzada JP, Nunes SM, Grossniklaus U, Goodrich J** (2000) Interaction of the Arabidopsis polycomb group proteins FIE and MEA mediates their common phenotypes. *Curr Biol* **10**: 1535–1538
- Steffen JG, Kang I-H, Portereiko MF, Lloyd A, Drews GN** (2008) AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol* **148**: 259–268
- Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE** (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nat Struct Mol Biol* **21**: 64–72
- Takeuchi H, Higashiyama T** (2012) A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in Arabidopsis. *PloS Biol* **10**: e1001449
- Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y** (2001) A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* **128**: 4681–4689
- Tang K, Lang Z, Zhang H, Zhu J-K** (2016) The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications. *Nat Plants* **2**: 16169
- Tirot L, Bonnet DMV, Jullien PE** (2022) DNA Methyltransferase 3 (MET3) is regulated by Polycomb group complex during Arabidopsis endosperm development. *Plant Reprod* **35**: 141–151
- Tiwari S, Schulz R, Ikeda Y, Dytham L, Bravo J, Mathers L, Spielman M, Guzmán P, Oakey RJ, Kinoshita T, et al** (2008) MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in Arabidopsis, encodes the conserved C-terminal domain of polyadenylate binding proteins. *The Plant Cell* **20**: 2387–2398
- Tonosaki K, Sekine D, Ohnishi T, Ono A, Furuumi H, Kurata N, Kinoshita T** (2018) Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. *Plant J* **93**: 534–544

References

- Twell D** (2011) Male gametogenesis and germline specification in flowering plants. *Sex Plant Reprod* **24**: 149–160
- Van Hautegeem T, Waters AJ, Goodrich J, Nowack MK** (2015) Only in dying, life: programmed cell death during plant development. *Trends Plant Sci* **20**: 102–113
- Vielle-Calzada JP, Baskar R, Grossniklaus U** (2000) Delayed activation of the paternal genome during seed development. *Nature* **404**: 91–94
- Vielle-Calzada JP, Thomas J, Spillane C, Coluccio A, Hoepfner MA, Grossniklaus U** (1999) Maintenance of genomic imprinting at the Arabidopsis *mea* locus requires zygotic DDM1 activity. *Genes Dev* **13**: 2971–2982
- Villar CBR, Erilova A, Makarevich G, Trösch R, Köhler C** (2009) Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. *Mol Plant* **2**: 654–660
- Vinkenoog R, Spielman M, Adams S, Fischer RL** (2000) Hypomethylation Promotes Autonomous Endosperm Development and Rescues Postfertilization Lethality in *ma* Mutants. *The Plant Cell* **12**: 2271–2282
- Vollset SE, Goren E, Yuan C-W, Cao J, Smith AE, Hsiao T, Bisignano C, Azhar GS, Castro E, Chalek J, et al** (2020) Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study. *Lancet* **396**: 1285–1306
- Vu TM, Nakamura M, Calarco JP, Susaki D, Lim PQ, Kinoshita T, Higashiyama T, Martienssen RA, Berger F** (2013) RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. *Development* **140**: 2953–2960
- Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L** (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr Biol* **19**: 1128–1132
- Wang L, Yuan J, Ma Y, Jiao W, Ye W, Yang D-L, Yi C, Jeffrey Chen Z** (2018) Rice Interploidy Crosses Disrupt Epigenetic Regulation, Gene Expression, and Seed Development. *Molecular Plant* **11**: 300–314
- Wang Y, Hou Y, Gu H, Kang D, Chen Z, Liu J, Qu L-J** (2012) The Arabidopsis APC4 subunit of the anaphase-promoting complex/cyclosome (APC/C) is critical for both female gametogenesis and embryogenesis. *Plant J* **69**: 227–240
- Wang Y, Hou Y, Gu H, Kang D, Chen Z-L, Liu J, Qu L-J** (2013) The Arabidopsis anaphase-promoting complex/cyclosome subunit 1 is critical for both female gametogenesis and embryogenesis(F). *J Integr Plant Biol* **55**: 64–74
- Wang Z, Butel N, Santos-González J, Simon L, Wärdig C, Köhler C** (2021) Transgenerational effect of mutants in the RNA-directed DNA methylation pathway on the triploid block in Arabidopsis. *Genome Biol* **22**: 141
- Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM** (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc Natl Acad Sci U S A* **110**: 19639–19644
- Waters AJ, Makarevitch I, Eichten SR, Swanson-Wagner RA, Yeh C-T, Xu W, Schnable PS, Vaughn MW, Gehring M, Springer NM** (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *The Plant Cell* **23**: 4221–4233
- Weijers D, Geldner N, Offringa R, Jürgens G** (2001) Seed development: Early paternal gene activity in Arabidopsis. *Nature* **414**: 709–710
- Weijers D, Van Hamburg J-P, Van Rijn E, Hooykaas PJJ, Offringa R** (2003) Diphtheria toxin-mediated cell ablation reveals interregional communication during Arabidopsis seed development. *Plant Physiol* **133**: 1882–1892


- Wendte JM, Pikaard CS** (2017) The RNAs of RNA-directed DNA methylation. *Biochim Biophys Acta Gene Regul Mech* **1860**: 140–148
- Wendte JM, Zhang Y, Ji L, Shi X, Hazarika RR, Shahryary Y, Johannes F, Schmitz RJ** (2019) Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation. *Elife* **8**: e47891
- Wierzbicki AT, Haag JR, Pikaard CS** (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**: 635–648
- Wierzbicki AT, Ream TS, Haag JR, Pikaard CS** (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**: 630–634
- Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ** (2007) An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS One* **2**: e718
- Wolf JB, Wade MJ** (2009) What are maternal effects (and what are they not)? *Philos Trans R Soc Lond B Biol Sci* **364**: 1107–1115
- Wolff P, Jiang H, Wang G, Santos-González J, Köhler C** (2015) Paternally expressed imprinted genes establish postzygotic hybridization barriers in *Arabidopsis thaliana*. *Elife* **4**: e10074
- Wolff P, Weinhofer I, Seguin J, Roszak P, Beisel C, Donoghue MTA, Spillane C, Nordborg M, Rehmsmeier M, Köhler C** (2011) High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* Endosperm. *PLoS Genet* **7**: e1002126
- Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, Song C, Zhuang M, Klasfeld S, He C, et al** (2017) Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat Genet* **49**: 1546–1552
- Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, et al** (2013) Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *The Plant Cell* **25**: 3212–3227
- Xu R-Y, Xu J, Wang L, Niu B, Copenhaver GP, Ma H, Zheng B, Wang Y** (2019) The *Arabidopsis* anaphase-promoting complex/cyclosome subunit 8 is required for male meiosis. *New Phytol* **224**: 229–241
- Yadegari R, Drews GN** (2004) Female gametophyte development. *The Plant Cell* **16 Suppl**: S133–41
- Yadegari R, Kinoshita T, Lotan O, Cohen G, Katz A, Choi Y, Katz A, Nakashima K, Harada JJ, Goldberg RB, et al** (2000) Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *The Plant Cell* **12**: 2367–2382
- Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, Zhang H, Zhu J-K** (2016) Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Research* **26**: 1264–1264
- Yang G, Liu Z, Gao L, Yu K, Feng M, Yao Y, Peng H, Hu Z, Sun Q, Ni Z, et al** (2018) Genomic Imprinting Was Evolutionarily Conserved during Wheat Polyploidization. *The Plant Cell* **30**: 37–47
- Yang L, Xing F, He Q, Qamar MT ul, Chen L-L, Xing Y** (2020) Conserved Imprinted Genes between Intra-Subspecies and Inter-Subspecies Are Involved in Energy Metabolism and Seed Development in Rice. *International Journal of Molecular Sciences* **21**: 9618
- Yang S, Johnston N, Talideh E, Mitchell S, Jeffree C, Goodrich J, Ingram G** (2008) The endosperm-specific ZHOUP1 gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. *Development* **135**: 3501–3509
- Ye R, Chen Z, Lian B, Rowley MJ, Xia N, Chai J, Li Y, He X-J, Wierzbicki AT, Qi Y** (2016) A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in *Arabidopsis*. *Mol Cell* **61**: 222–235

References

- Yoshida T, Kawabe A** (2013) Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in Arabidopsis species. *PLoS One* **8**: e73588
- Yoshida T, Kawanabe T, Bo Y, Fujimoto R, Kawabe A** (2018) Genome-Wide Analysis of Parent-of-Origin Allelic Expression in Endosperms of Brassicaceae Species, *Brassica rapa*. *Plant Cell Physiol* **59**: 2590–2601
- Yuan J, Chen S, Jiao W, Wang L, Wang L, Ye W, Lu J, Hong D, You S, Cheng Z, et al** (2017) Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol* **216**: 373–387
- Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D** (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**: 193–205
- Zhang M, Li N, He W, Zhang H, Yang W, Liu B** (2016) Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J* **85**: 424–436
- Zhang M, Xie S, Dong X, Zhao X, Zeng B, Chen J, Li H, Yang W, Zhao H, Wang G, et al** (2014) Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res* **24**: 167–176
- Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, et al** (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* **108**: 20042–20047
- Zhang M-X, Zhu S-S, Xu Y-C, Guo Y-L, Yang W-C, Li H-J** (2020) Transcriptional repression specifies the central cell for double fertilization. *Proc Natl Acad Sci U S A* **117**: 6231–6236
- Zhang S, Wang D, Zhang H, Skaggs MI, Lloyd A, Ran D, An L, Schumaker KS, Drews GN, Yadegari R** (2018) FERTILIZATION-INDEPENDENT SEED-Polycomb Repressive Complex 2 Plays a Dual Role in Regulating Type I MADS-Box Genes in Early Endosperm Development. *Plant Physiol* **177**: 285–299
- Zhang X, Jacobsen SE** (2006) Genetic analyses of DNA methyltransferases in Arabidopsis thaliana. *Cold Spring Harb Symp Quant Biol* **71**: 439–447
- Zheng B, Chen X, McCormick S** (2011) The anaphase-promoting complex is a dual integrator that regulates both MicroRNA-mediated transcriptional regulation of cyclin B1 and degradation of Cyclin B1 during Arabidopsis male gametophyte development. *The Plant Cell* **23**: 1033–1046
- Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE** (2014) Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* **157**: 1050–1060
- Zhong X, Hale CJ, Nguyen M, Ausin I, Groth M, Hetzel J, Vashisht AA, Henderson IR, Wohlschlegel JA, Jacobsen SE** (2015) Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. *Proc Natl Acad Sci U S A* **112**: 911–916
- Zhou Y, Wang Y, Krause K, Yang T, Dongus JA, Zhang Y, Turck F** (2018) Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. *Nat Genet* **50**: 638–644
- Zhu H, Xie W, Xu D, Miki D, Tang K, Huang C-F, Zhu J-K** (2018) DNA demethylase ROS1 negatively regulates the imprinting of DOGL4 and seed dormancy in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **115**: E9962–E9970

Papers I - III

Genetic variation and temperature affects hybrid barriers during interspecific hybridization

Katrine N. Bjerkan^{1,2}, Karina S. Hornslien¹, Ida M. Johannessen¹, Anders K. Krabberød¹, Yuri S. van Ekelenburg¹, Maryam Kalantarian¹, Reza Shirzadi¹, Luca Comai³, Anne K. Brysting^{1,2}, Jonathan Bramsiepe^{1,2} and Paul E. Grini^{1,*} 

¹EVOGENE, Department of Biosciences, University of Oslo, 0316, Oslo, Norway,
²CEES, Department of Biosciences, University of Oslo, 0316, Oslo, Norway, and
³Plant Biology and Genome Center, University of California, Davis, Davis, CA, 95616, USA

Received 5 August 2018; revised 31 July 2019; accepted 19 August 2019; published online 5 September 2019.

*For correspondence (e-mail paul.grini@ibv.uio.no).

SUMMARY

Genomic imprinting regulates parent-specific transcript dosage during seed development and is mainly confined to the endosperm. Elucidation of the function of many imprinted genes has been hampered by the lack of corresponding mutant phenotypes, and the role of imprinting is mainly associated with genome dosage regulation or allocation of resources. Disruption of imprinted genes has also been suggested to mediate endosperm-based post-zygotic hybrid barriers depending on genetic variation and gene dosage. Here, we have analyzed the conservation of a clade from the MADS-box type I class transcription factors in the closely related species *Arabidopsis arenosa*, *A. lyrata*, and *A. thaliana*, and show that *AGL36-like* genes are imprinted and maternally expressed in seeds of *Arabidopsis* species and in hybrid seeds between outbreeding species. In hybridizations between outbreeding and inbreeding species the paternally silenced allele of the *AGL36-like* gene is reactivated in the hybrid, demonstrating that also maternally expressed imprinted genes are perturbed during hybridization and that such effects on imprinted genes are specific to the species combination. Furthermore, we also demonstrate a quantitative effect of genetic diversity and temperature on the strength of the post-zygotic hybridization barrier. Markedly, a small decrease in temperature during seed development increases the survival of hybrid F1 seeds, suggesting that abiotic and genetic parameters play important roles in post-zygotic species barriers, pointing at evolutionary scenarios favoring such effects.

Keywords: Imprinting, endosperm, hybridization, post-zygotic barriers, *Arabidopsis thaliana*, *Arabidopsis arenosa*, *Arabidopsis lyrata*.

INTRODUCTION

Seed development is a sophisticated and highly regulated process that requires precise signaling events and interaction between many distinct cell types and tissues. It starts with fusion of the male and female gametes generated in the male and female gametophytes, giving rise to the embryo and endosperm that develop in parallel inside the protective seed coat. The process is initiated when a conspecific pollen grain lands on the stigma of the female reproductive organ and the pollen tube delivers two sperm cells to the female gametophyte. One sperm cell fertilizes the haploid egg cell which develops into the diploid embryo, while the other sperm cell fertilizes the homodiploid central cell generating the triploid endosperm. The endosperm is important for nutrient flow to the embryo

but also for coordinating growth of the developing seed (Nowack *et al.*, 2010).

The endosperm has two maternal genome copies and one paternal copy, and a specialized epigenetic phenomenon called genomic imprinting regulates parent-specific gene dosage during seed development, usually occurring in the endosperm (Gehring and Satyaki, 2017). Imprinting is manifested by expression of one parental allele, with concurrent silencing of the other allele. The main mechanisms for this process are DNA methylation and histone methylation (Berger *et al.*, 2006). The FERTILIZATION INDEPENDENT SEED-Polycomb Repressive Complex 2 (FIS-PRC2) mediates histone methylation while, in *A. thaliana*, DNA methylation mediated imprinting is

maintained by the DNA methyltransferase MET1 (Rodrigues and Zilberman, 2015).

A prominent gene family displaying frequent imprinting of its members is the MADS-box transcription factor (TF) family. The MADS-box TFs can be divided into type I and type II by evolutionary relationships. The type I TFs are further divided into $M\alpha$, $M\beta$, and $M\gamma$ phylogenetic subclasses and only share the highly conserved DNA-binding MADS (M) domain. The type II TFs have, in addition to the M domain, the Intervening (I), the Keratin (K) and the C-terminal (C) domains that are often referred to as the MIKC type (Parenicova *et al.*, 2003). The type II class is thought to have evolved from an ancient whole genome duplication, as orthologs are found in many other species and the genes are well distributed across all chromosomes in *A. thaliana*. The type I class TFs originate from more recent and smaller scale duplication events and, in *A. thaliana*, they are mainly concentrated on chromosomes I and V (Parenicova *et al.*, 2003; Airoidi and Davies, 2012). As a consequence, MADS-box type I orthologs are uncommon in other species (Masiero *et al.*, 2011). Imprinted genes occur frequently in the type I class, consistent with the hypothesis that recently duplicated genes are more often imprinted to regulate gene dosage (Yoshida and Kawabe, 2013). Imprinting is observed mainly in the $M\alpha$ and $M\gamma$ subgroups and, moreover, members of these two subclasses interact extensively in yeast two-hybrid assays, suggesting a common function as heterodimers (de Folter *et al.*, 2005).

Functional studies of the MADS-box type I TFs by genetic dissection, however, are hampered by genetic redundancy. Their roles have also been suggested to have restricted effect and may therefore be involved in a specific developmental processes (Nam *et al.*, 2004). Only a few type I genes have been studied phenotypically, including *AGAMOUS-LIKE (AGL) 23*, *AGL36*, *PHERES (PHE) 1 (AGL37)* and *PHE2 (AGL38)*, *DIANA (AGL61)*, *AGL62*, and *AGL80* (Köhler *et al.*, 2003; Köhler *et al.*, 2005; Bemer *et al.*, 2008; Colombo *et al.*, 2008; Kang *et al.*, 2008; Steffen *et al.*, 2008; Shirzadi *et al.*, 2011). To this end, the biological roles of many imprinted genes are still not known, but the role of imprinting is mainly associated with genome dosage regulation or allocation of resources (Haig and Westoby, 1989; Dilkes and Comai, 2004; Rodrigues and Zilberman, 2015).

Imprinting has previously been shown to be disrupted in hybrid crosses of *A. thaliana* and *A. arenosa*. The MADS-box TF *PHE1*, which is imprinted and only paternally expressed in *A. thaliana*, was upregulated in hybrid seeds and it was shown that the expressed *PHE1* was predominantly maternally expressed (Josefsson *et al.*, 2006). Disruption of the expression levels of co-adapted MADS-box TFs in hybrids may thus trigger genome-wide perturbations observed in hybrids (Roth *et al.*, 2019). Furthermore,

other MADS-box type I TFs have been shown to be highly upregulated in incompatible hybrid crosses between *A. thaliana* mothers and *A. arenosa* fathers. Using knock-out mutant lines of these genes as *A. thaliana* mother, increased viability in the incompatible hybrid seeds, suggesting that these MADS-box type I TFs partly constitute a genetic basis for the post-zygotic barrier (Walia *et al.*, 2009). Hence, investigation of the imprinting status of these genes and other known imprinted genes in *A. arenosa* and *A. lyrata* will shed light on the role and consequently the evolution of imprinting. It is disputed whether imprinting of specific genes is conserved, and whether the mechanisms behind the establishment and maintenance of imprinting between related and distant species are preserved (Waters *et al.*, 2013; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016; Chen *et al.*, 2018).

Diploid *A. arenosa* crossed as father to more than 50 accessions of *A. thaliana* displayed live seeds in the range of 1% to 30% (Burkart-Waco *et al.*, 2012). This suggests that the strength of the post-zygotic barrier can be modulated by genetic variation in accessions. Comparison of *A. arenosa* crossed to different *A. thaliana* accessions, thorough phenotyping (Burkart-Waco *et al.*, 2013) and sequencing of RNA from hybrid seeds (Burkart-Waco *et al.*, 2015), identified perturbation of the imprinting patterns of eight known paternally expressed genes. As these crosses were limited to a specific *A. arenosa* population, we hypothesize that the observed barrier is population dependent. Lafon-Placette *et al.* (2017) demonstrated that in crosses between *A. lyrata* and *A. arenosa*, the post-zygotic species barrier is due to endosperm cellularization failure. A similar study in the *Capsella* genus also indicated endosperm failure as the main seed defect in incompatible crosses (Rebernick *et al.*, 2015). A post-zygotic endosperm-based barrier has also been described for rice (Tonosaki *et al.*, 2018; Wang *et al.*, 2018) and tomato (Florez-Rueda *et al.*, 2016).

Here we have investigated the role of genetic variation in the establishment of post-zygotic endosperm-based hybrid barriers both in general, using accession and in a targeted manner, addressing specific MADS-box type I loci. We investigated the function and regulation of a conserved clade of MADS-box type I $M\gamma$ class (*AGL34*, *AGL36*, and *AGL90*) together with some of their interacting partners. To further elucidate function, we have analyzed the conservation of this clade in the closely related species *A. arenosa*, *A. lyrata*, and *A. halleri*, including the imprinting status of *AGL36-like* genes in *A. arenosa*, in *A. thaliana* crossed to *A. arenosa* and in the reciprocal cross of *A. arenosa* and *A. lyrata*. We find that *AGL36-like* genes are imprinted and maternally expressed in seeds of *Arabidopsis* species and in hybrid seeds between outbreeding species. In hybridizations between outbreeding and inbreeding species the paternally silenced allele of the

AGL36-like gene is reactivated in the hybrid, demonstrating that also maternally expressed imprinted genes are perturbed during hybridization and that such effects on imprinted genes are specific to the species combination.

Moreover, we investigated the role of temperature in hybridization of different genetic backgrounds and specific loci and find a significant positive correlation between lower temperatures and hybrid seed germination rate. We report that just a small change in temperature during seed development is sufficient to increase survival of hybrid F1 seeds, suggesting that abiotic parameters play an important role in post-zygotic, endosperm-based species barriers. Crossing mutants of the $M\gamma$ and $M\alpha$ clades, and their interacting partners to *A. arenosa* to further investigate the effect of these genes on the hybridization barrier identified that lack of *AGL35* significantly aggravated the *A. thaliana* *A. arenosa* hybrid barrier and that *AGL35* is involved in the temperature dependency of the hybrid barrier.

RESULTS

MADS-box type I $M\alpha$ and $M\gamma$ expression in seed development

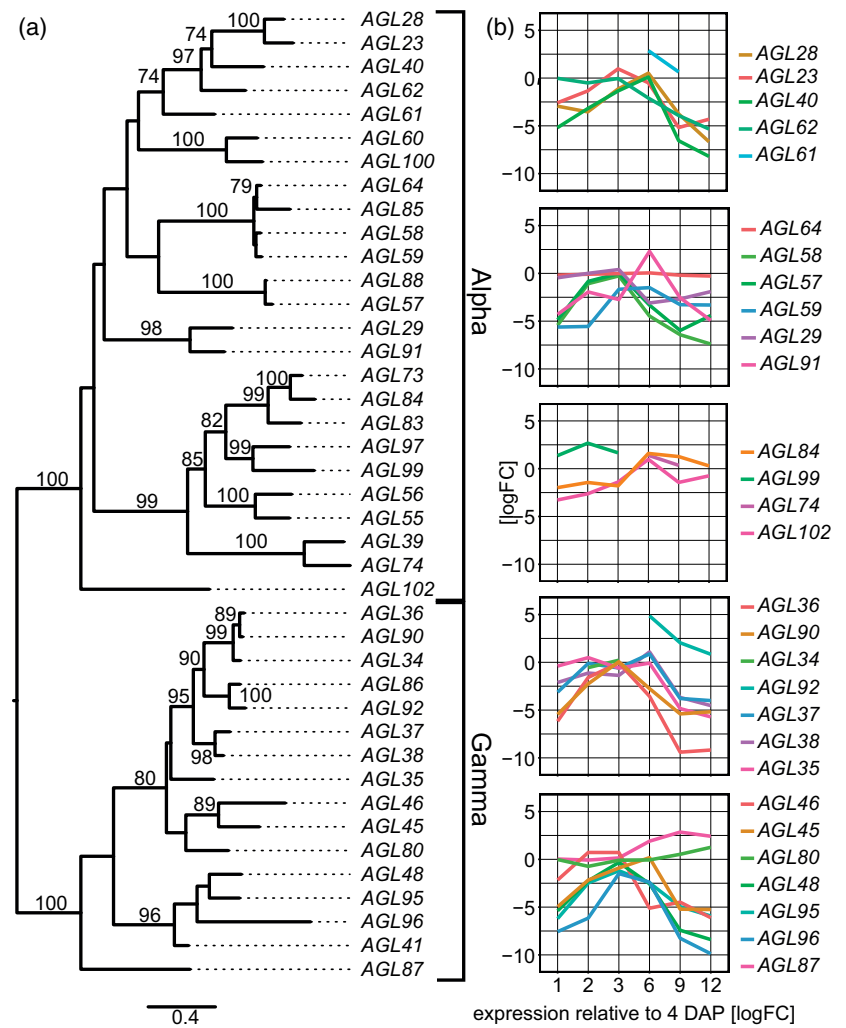
In order to investigate the role of imprinted loci in the establishment of endosperm-based hybrid barriers, we analyzed MADS-box TFs that are closely related to the paternally silenced *AGL36* (Shirzadi *et al.*, 2011). We re-analyzed the phylogeny of the $M\alpha$ and $M\gamma$ classes and assembled them in two groups with several subclades (Figure 1a). In the $M\gamma$ group, *AGL36* constitutes a subclade together with *AGL90* and *AGL34*, and the latter genes may represent recent, local gene duplication events in *A. thaliana* as orthologs are not readily identified (Masiero *et al.*, 2011). *AGL36* and *AGL90* have been shown to be imprinted (Shirzadi *et al.*, 2011, Zhang *et al.*, 2018). The wider subclade includes known imprinted, paternally expressed genes *PHE1/AGL37* and *AGL92* (Wolff *et al.*, 2011), which cluster together with *PHE2/AGL38* and *AGL86*, respectively. *PHE2* expression has previously been demonstrated to be bi-allelic (Villar *et al.*, 2009). The most distant member of the subclade is *AGL35*, which is closely linked to and located between *AGL34* and *AGL36* on chromosome 5. *AGL34*, *AGL35*, *AGL36* and *AGL90* all map in a 100 kbp cluster on chromosome 5 (Parenicova *et al.*, 2003), and this clustering makes this subclade an especially interesting case to study evolution of imprinted genes. The function of *AGL90* and *AGL34* is not known, but *AGL36* interacts with two $M\alpha$ MADS-box TFs, *AGL28* and *AGL62* (de Folter *et al.*, 2005; Bemer *et al.*, 2010) where *AGL28* has been shown not to be imprinted (Zhang *et al.*, 2018) or to display accession dependent imprinting (Wolff *et al.*, 2011). *AGL62* is biparentally expressed and required for endosperm cellularization (Kang *et al.*, 2008), and for reason of functional study we have included *AGL62* and *AGL28* in our analysis.

Next, we investigated the relative expression of all MADS-box type I TFs at seed developmental stages ranging from 1 day after pollination (DAP) to 12 DAP (Figures 1b and S1). *AGL36* expression peaked at 4 DAP and coincided with the timing of endosperm cellularization (Shirzadi *et al.*, 2011), hence an RNA sequencing-based differential expression analysis relating all stages to 4 DAP was performed. $M\alpha$ and $M\gamma$ class TFs are overrepresented in the transcriptome of the developing seed compared with the $M\beta$ class. All 16 $M\gamma$ genes and two-thirds of the 25 $M\alpha$ genes are expressed, whereas less than half of the 21 $M\beta$ genes can be identified (Figure S1b). Ordering the $M\alpha$ and $M\gamma$ expression profiles according to the branching pattern displayed a general expression trend with a peak between 4 and 6 DAP (Figure 1b). In the $M\gamma$ *AGL36* subclade, *AGL35*, *AGL36* and *AGL90* display similar profiles with increasing or unchanged expression toward 4 DAP followed by a decline. *AGL34* can only be detected in a small developmental window, but the relative expression pattern is equivalent to *AGL36* and *AGL90* at these stages (Figures 1b and S1b). This supports findings by Zhang *et al.* (2018) and indicates that *AGL34* is not a pseudogene, as previously postulated (Bemer *et al.*, 2010). A similar pattern is found in $M\alpha$ subclades, including *AGL28*. A decrease of *AGL62* levels was observed after the expression maxima observed in $M\alpha$ and $M\gamma$ classes (Figure 1b). The $M\alpha$ *AGL62* is required for correct timing of endosperm cellularization (Kang *et al.*, 2008), and thus plays a putative role in the establishment of endosperm-based hybrid barriers (Lafon-Placette *et al.*, 2017). The $M\alpha$ MADS-box TF subclass is hypothesized to form dimers with the $M\gamma$ -type (de Folter *et al.*, 2005), and taken together, the co-occurring $M\alpha$ and $M\gamma$ expression patterns may indicate a possible role for these TFs in the establishment of cellularization-based post-zygotic hybrid barriers.

Imprinting and regulation of $M\gamma$ and $M\alpha$ MADS-box genes

Dosage imbalance caused by imprinted genes has been proposed as a cause for hybrid failure in both plants and animals (Dilkes and Comai, 2004; Wolf *et al.*, 2014; Brekke *et al.*, 2016). To this end, we successively re-analyzed parent-of-origin expression of the wider *AGL36* subclade (Figure 1a), including the interacting $M\alpha$ *AGL28* (Figure 2). Using Col-0 and Tsu-1 accession-specific single nucleotide polymorphisms (SNP), we analyzed *AGL28*, *AGL35*, *AGL36* and *AGL90* in 4 DAP Col-0 Tsu-1 hybrid seeds (Figure 2a). Maternal bias from the seed coat could be excluded, as all transcripts were previously shown to be enriched >8-fold 4 DAP in the peripheral endosperm (*AGL36*, *AGL90*) compared to all other seed tissues or >8-fold and >5-fold enriched in the chalazal endosperm (*AGL35* and *AGL28*, respectively) (Belmonte *et al.*, 2013; Hornslien *et al.*, 2019). *AGL34* was not expressed at a sufficient level in 4 DAP Tsu-1 and was thus omitted. Gene-specific RT-PCR

Figure 1. MADS-box type I transcription factors share similar expression profiles during seed development. (a) Maximum likelihood phylogeny of alpha and gamma MADS-box type I genes in *Arabidopsis thaliana*. The tree was inferred using the GTRGAMMA model on 41 genes with 532 unambiguously aligned nucleotides. Scale bar represents the mean number of nucleotide substitutions per site. Only bootstrap values above 65% are shown. (b) Gene expression profiles of alpha and gamma MADS-box type I genes were ordered in five groups according to the branching pattern. Transcript quantification and differential expression analysis was performed with RSEM and visualized using R. Gene expression profiles for stages ranging from one to 12 days after pollination (DAP) are relative to four DAP using a base-2 logarithmic scale (logFC). Two biological replicates with three technical replicates were analyzed. Note that genes within groups show similar gene expression profiles, with a common maximum reached between three and six DAP.



products from hybrid crosses were digested with SNP-specific restriction endonucleases (Table S1) and fragments analyzed on a Bioanalyzer 2000 as well as by Sanger sequencing (Figures 2a and S2). The M_γ genes *AGL36* and *AGL90* and the M_α *AGL28* were imprinted and maternally expressed. The M_γ *AGL35* was biparentally expressed as previously reported (Zhang *et al.*, 2018). *AGL36* (Shirzadi *et al.*, 2011; Wolff *et al.*, 2011; Zhang *et al.*, 2018) and *AGL90* (Zhang *et al.*, 2018) has previously been shown to be imprinted. Here, we show that *AGL90* is maternally biased in its expression but the paternal allele show accession dependent imprinting and is not completely silenced from Col-0 pollen donors (Figures 2a and S2). A similar lack of silencing of the *AGL90* paternal Col-0 allele was recently also reported (Hornslien *et al.*, 2019). In contrast, *AGL28* was previously reported not to be imprinted (Zhang *et al.*, 2018), or to display accession dependent imprinting in hybrids (Wolff *et al.*, 2011).

To address the regulation of imprinted genes we contrasted 4 DAP parental seed expression from crosses

between *Tsu-1* and *Col-0* accessions versus *Tsu-1* and a hemizygous *met1-7+/-* in a *Col-0* background to determine if *MET1* is involved in maintaining silencing of the paternal copy of *AGL28*, *AGL36* and *AGL90* (Figure 2b, Table S1, Figure S2). The paternal copy of *AGL28* was shown to be expressed using *met1-7+/-* as pollen donor, suggesting that *MET1* is required for silencing of the paternal *AGL28* allele. In our experimental settings, however, lack of *MET1* did not reactivate the paternal copy of *AGL36* and *AGL90*, whereas the *AGL35* biparental control remains unchanged (Figure 2b). In contrast with these findings, the paternal allele of *AGL36* was previously shown to be reactivated in crosses with homozygous and hemizygous *MET1* mutant pollen using the *met1-4+/-* allele (Saze *et al.*, 2003; Shirzadi *et al.*, 2011). Lack of *MET1*, both in homozygous and heterozygous mutants, leads to DNA hypomethylation and eventually the accumulation of epimutations; we attributed the previously observed paternal expression to such effects in the *met1-4+/-* background. The *met1-7+/-* allele used in our study was kept hemizygous through repeated

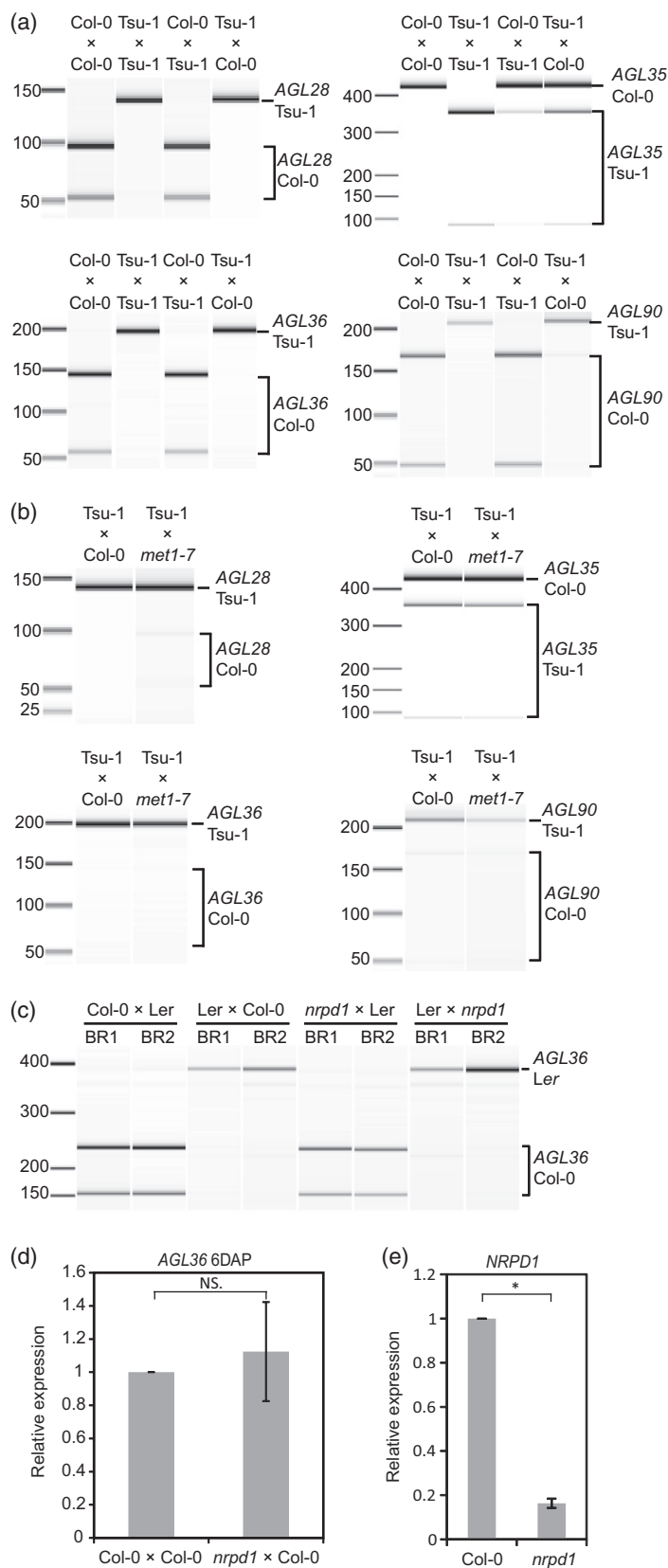


Figure 2. Imprinting and epigenetic regulation of *AGL28*, *AGL35*, *AGL36* and *AGL90*. (a) Imprinting analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90* using accession-specific restriction digest on single nucleotide polymorphisms (SNPs) in reciprocal crosses between accessions *Tsu-1* and *Col-0*. For each panel, the accession-specific digestion pattern is indicated. Seeds were harvested for analysis 4 days after pollination (DAP). Bioanalyzer images of one of three biological replicates is shown. (b) SNP analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90* in crosses with *met1-7+/-* (*Col-0* background) pollen. Only *AGL28* display paternal activation in the *met1-7+/-* mutant. Wild type (WT) crosses are duplicated from (a) for visualization. Crosses were harvested as (a). Bioanalyzer images of one of three biological replicates is shown. (c) SNP analysis of *AGL36* in reciprocal crosses between WT (*Col-0*) or *nrdp1* (*Col-0*) and WT (*Ler*). Crosses were harvested six DAP. The *AGL36* imprinting pattern is not changed in the *nrdp1* crosses compared with WT crosses. (d) Real-time PCR analysis of *AGL36* expression in three biological replicates the *nrdp1* background at six DAP. The relative expression difference is not significant (NS, *t*-test; $P = 0.389$). (e) Real-time PCR verifying significant knock-down (*t*-test; $P = 0.015$, indicated by asterisk) of *NRPD1* in three biological replicates of *nrdp1* homozygous background. Error bar indicates standard deviation (SD).

outcrosses and therefore more likely devoid of such effects (Hornslien *et al.*, 2019).

According to a report investigating the role of small interfering (si) RNA and RdDM in interploidy crosses, several MADS-box type I genes, including *AGL36*, were deregulated in diploid crosses with NUCLEAR RNA POLYMERASE D1 (*NRPD1*) mutant mothers, deficient in the largest subunit of RNA polymerase IV, a key component of canonical RdDM (Lu *et al.*, 2012). Using nuclear rna polymerase d1 (*nrdp1*) mutant mothers, *AGL36* was upregulated more than 20-fold (Lu *et al.*, 2012). We have previously shown that imprinting of *AGL36* do not require paternal DOMAINS REARRANGED METHYLTRANSFERASE 2 (*DRM2*) or ARGONAUTE4 (*AGO4*), both part of the RdDM pathway (Shirzadi *et al.*, 2011). The data from Lu *et al.* (2012) suggested that an RdDM-dependent mechanism maintained the expression level of *AGL36* or is active in maintaining the silencing of the paternal *AGL36* allele after fertilization. To test the latter hypothesis, we analyzed parental expression from 6 DAP seeds, using *nrdp1* both as a maternal and paternal contributor in crosses to wild type (Figures 2c and S2). The *AGL36* imprinting pattern was not affected in any cross direction, suggesting that reactivation of the paternal allele is not causing elevated levels of *AGL36*. In contrast with the previous report (Lu *et al.*, 2012), we could also not detect any significant upregulation of *AGL36* by real-time PCR (Figure 2d) in crosses using a homozygous *nrdp1* knock-out allele (Figure 2e) as maternal cross partner. We concluded that neither the MET1 nor the PolIV RdDM pathway is sufficient to silence the paternal allele of *AGL36*.

Next, we analyzed the effect of PRC2 on MADS-box type I genes. The endosperm cellularization defect observed in Arabidopsis interspecies hybrid seeds is highly reminiscent to the failure of endosperm cellularization phenotype observed in mutants of *FIS-PRC2* (Lafon-Placette *et al.*, 2017). We therefore compared the RNA-seq relative expression of all MADS-box type I genes between a *FIS-PRC2* mutant and wild type at seed developmental stages ranging from one DAP to 12 DAP (Figures 3a and S3). Clustering of transcript profiles revealed four main patterns of regulation, ranging from highly regulated to no effect (Figure S3). Overall, the M β class as a group was significantly less regulated by the PRC2 *medea* (*mea*) mutation than the M α and M γ

classes, and also displayed the least variation (Figure 3b). This is in accordance with previous observations, that the M β class TF are mainly expressed at low levels or in female gametophytic stages (Bemer *et al.*, 2010).

In the deregulated classes of transcript profiles, consisting of mainly but not exclusively M α and M γ , the wild type profiles are generally characterized by increasing expression that decreases after a peak (Figure S3, left panels). In the *mea* cross, both this pattern and the peak are shifted toward higher expression levels and later developmental stages. In certain cases, decrease is not observed within the analyzed developmental time-frame (Figure S3, right panels). A distinct shift in transcript profiles could also be observed between the profile clusters in *mea*, with one class downregulated before 6 DAP, while a second class started at 9 DAP and the third class at 12 DAP (Figure 3a, three top clusters). M γ dominates the two former classes of transcript profiles together with M α whereas the latter constitutes of M α and M β genes. In a recent report, Zhang *et al.* (2018) analyzed MADS-box type I deregulation in a *swinger* (*swi*) *mea* double mutant. These authors identified two major expression clusters (C1 and C2) based on difference in temporal expression patterns both in the wild type and in the PRC2 double mutant. The latter cluster was distinguished by the an upregulation of the expression pattern in the mutant and could be further divided in two clusters (C2.1 and C2.2) based on the timing of downregulation in the wild type (Zhang *et al.*, 2018). The three clusters described in our study (Figure 3a, three top clusters) are well in line with the real-time-PCR-based study of Zhang *et al.* (2018). Eight out of 12 genes in the C2.1 cluster are also found in our top cluster, starting deregulation at the earliest stage (Figure 3a, top cluster), whereas three genes are found in our second cluster (Figure 3a, second top cluster) together with all *mea swi* upregulated genes identified in the C2.2 cluster. This also includes *AGL91*, *AGL49* and, importantly, *AGL34* that are upregulated in our study, whereas no upregulation was identified by Zhang *et al.* (2018). In contrast, we could not detect any upregulation for *AGL64*, as reported by the other study (Zhang *et al.*, 2018).

We conclude that the *AGL36* subclade, including *AGL34* and *AGL90*, as well as the *AGL36* interacting M α *AGL28* are commonly repressed by MEDEA from 4-6 DAP. The *AGL36*

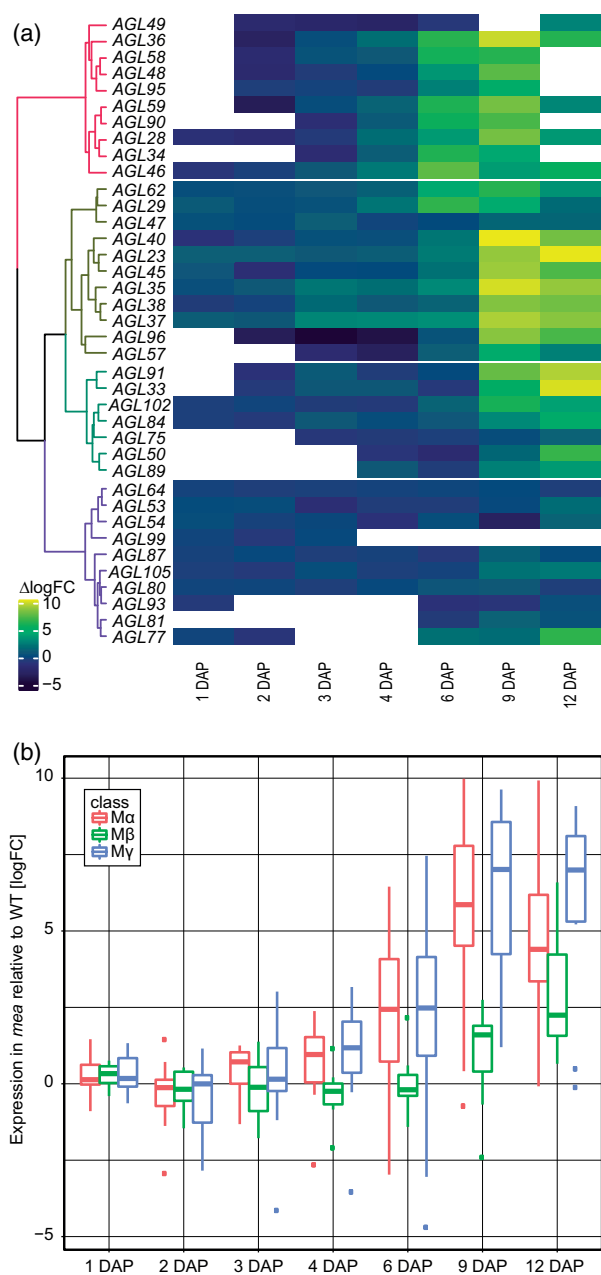


Figure 3. PRC2 dependent transcriptional repression of MADS-box type I transcription factors during seed development. (a) Heat-map clustering of gene expression profiles of MADS-box type I transcription factors (TFs) in the Polycomb Repressive Complex 2 (PRC2) *MEDEA* mutant seeds (*mea*) compared to wild type (WT). Expression profile-based clustering, transcript quantification and differential expression analysis was performed for three biological replicates using RSEM and visualized using R. Differential gene expression profiles for stages ranging from one to 12 days after pollination (DAP) are shown using a base-2 logarithmic scale (logFC). Note strong expression change in three out of four expression clusters. (b) Box plot showing class specific expression of MADS-box type I TFs at stages from 1 to 12 DAP. The MEA-dependent repression is shared by the alpha ($M\alpha$) and gamma ($M\gamma$) class TFs, while the beta ($M\beta$) class is weakly affected in the *mea* background. Relative expression changes are displayed in a base-2 logarithmic scale. Points indicate outliers.

and also the AGL90 interacting $M\alpha$ AGL62 are upregulated in *mea* at 9 DAP, in accordance with the role of AGL62 in endosperm cellularization (Kang *et al.*, 2008).

Conservation of Arabidopsis *AGL36-like* imprinting in hybrid seeds is species dependent

Having analyzed the expression and regulation of MADS-box type I genes in *A. thaliana*, we turned our focus to the expression and role of these genes in the genus Arabidopsis. MADS-box type I genes are often less conserved between model species. For instance, no orthologs of *AGL36* were identified in rice or maize (Masiero *et al.*, 2011). *AGL36-like* genes can be found when analyzing more closely related species such as in the genus Arabidopsis (Figures 4a and S4). Two genomic loci of *AGL36-like* genes were identified in *A. arenosa* by assembling online resources (see Experimental procedures). Both were verified in various individuals from two *A. arenosa* populations using PCR amplification (MJ09-1, MJ09-4) (Jørgensen *et al.*, 2011; Lafon-Placette *et al.*, 2017). The two genes differ in length (1050 bp versus 1008 bp). The shorter gene does not have a continuous open reading frame and most likely harbors an intron based on two open reading frames spaced by an 88-bp sequence. The 1050-bp locus, but not the 1008-bp *AGL36* locus, was confirmed to produce a transcript in 9 DAP seeds, corresponding to the globular-embryo seed stage. Online genome sequencing resources of *A. lyrata* subsp. *lyrata* suggest one *AGL36-like* gene (Figure 4a). In the subspecies *A. lyrata* subsp. *petrea*, two loci have been indicated (Yoshida and Kawabe, 2013), but by performing Sanger sequencing from the *A. lyrata* subsp. *petrea* population MJ09-11 (Jørgensen *et al.*, 2011) combined with online resources, we concluded that *A. lyrata* contains only one *AGL36-like* locus (Figures 4a and S4). Notably, although two *AGL36-like* loci are present in *A. arenosa* and *A. halleri*, our analysis indicates that the two duplication events creating the *AGL36* subclade (*AGL34*, *AGL36*, *AGL90*) do not exist outside *A. thaliana*.

To analyze imprinting of *A. arenosa* *AGL36-like* (*AaAGL36-like*), we screened natural populations (MJ09-4 and MJ09-1) for SNPs that could be used to distinguish the parental alleles. We identified one individual that had a SNP in *AaAGL36-like* (I, cf. Figure 4a) that also allowed SNP detection with restriction enzymes (Table S1c). Seed RNA was harvested from reciprocal crosses at 9 DAP followed by RT-PCR of the SNP-containing regions from *AaAGL36-like* (Table S2). The PCR products were digested with SNP-specific enzymes (Table S1c) and fragments analyzed (Figure 4b). Only maternal expression was found, suggesting that *A. arenosa* *AGL36-like* (*AaAGL36* I) is an imprinted maternally expressed gene.

Next, we analyzed *AGL36-like* imprinting in hybrids of *A. arenosa* and *A. lyrata*. Amplifying *AGL36-like* (Table S1c)

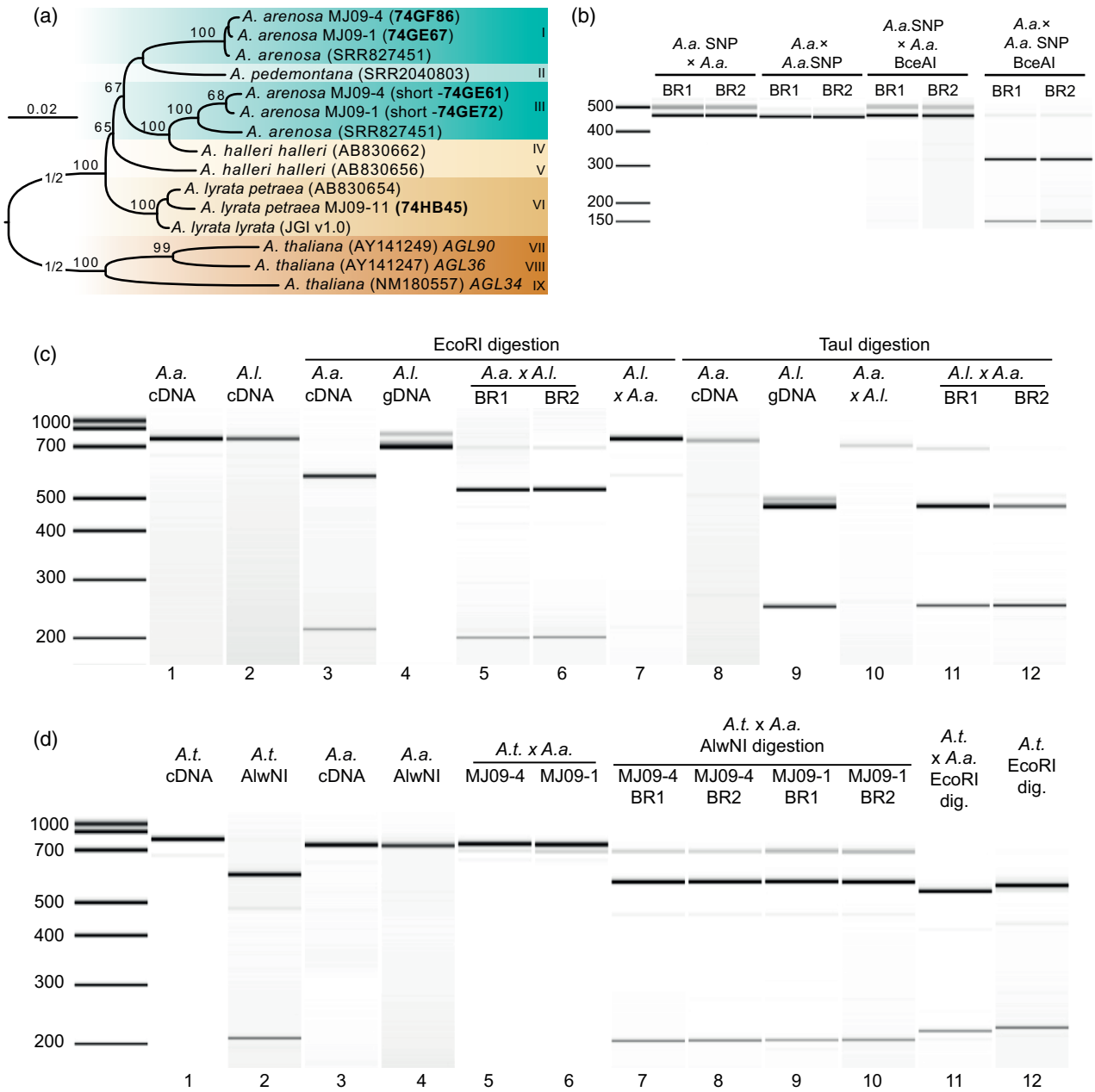


Figure 4. Conservation and imprinting of *AGL36*-like genes in Arabidopsis and in selected hybrid crosses. (a) Maximum likelihood phylogeny of *AGL36*-like genes in *A. arenosa*, *A. pedemontana*, *A. halleri*, *A. lyrata* and *A. thaliana*. The tree was inferred using the GTRGAMMA model on 15 sequences with 938 unambiguously aligned nucleotides. Because of space limitations branches marked ½ are shortened to half their original length. Only bootstrap values above 65% are shown. Scale bar represents the mean number of nucleotide substitutions per site. Arabidopsis species are indicated by colors. Roman numerals (right) indicate distinct genes in the respective species and is used as reference in (b–d). (b) SNP analysis of *A. arenosa* *AGL36* (I) in MJ09-4 background at 9 days after pollination (DAP). Left half, undigested; right half, digested with *Bce*AI. Cross plants in which *AaAGL36* is digested by *Bce*AI are indicated ‘A.a.’. Cross plants in which *AaAGL36* is not digested by *Bce*AI are indicated ‘A.a. SNP’. The shorter 475-bp fragment is *AaAGL36*. *AaAGL36* (I) is maternally expressed in *A. arenosa* as only the undigested *AGL36* fragment is present when A.a. SNP is the mother, whereas completely digested fragments result when A.a. is the mother. (c) Imprinting and maternal expression of *AGL36* is conserved in reciprocal crosses between *A. arenosa* and *A. lyrata* at 9 DAP. *AaAGL36* (I) and *AlAGL36* (VI) fragments are both 768 bp of length (lanes 1–2). *Eco*RI digests *AaAGL36* (I) only (lanes 3 and 4) and *Tau*I digests *AlAGL36* (VI) only (lanes 8 and 9). In hybrid crosses, only the maternal fragments are digested, and no or very weak undigested fragments are left (lanes 5–7, 10–12). (d) Imprinting of *AGL36* is lifted in crosses between *A. thaliana* and *A. arenosa*. The undigested fragments of *AtAGL36* (VIII) and *AaAGL36* (I) are 819 and 768 bp, respectively (lanes 1 and 3). An *Alw*NI restriction site is only present in *AtAGL36* (VIII) (lanes 2 and 4). In hybrid crosses, both fragments are visible (lanes 5 and 6) and *Alw*NI restriction digest only the maternal *AtAGL36* fragment. Note that the paternal *AaAGL36* (I) allele is expressed (lanes 7–10). *Eco*RI digests both *AaAGL36* and *AtAGL36* leaving only digested fragments (lanes 11 and 12). BR, biological replicate; A.t., *A. thaliana*; A.a., *A. arenosa*; A.l., *A. lyrata*. All crosses shown are in the order female × male. Two biological replicates represent results from three biological replicates tested. DAP stages used are four for A.t., seven for A.t. × A.a. and 9 DAP for A.a., A.l. and the A.a.xA.l. reciprocal cross.

from cDNA of reciprocal crosses of *A. arenosa* and *A. lyrata* resulted in one fragment because *AGL36-like* (I and VI, respectively in Figure 4a) from both species are the same length (Figure 4c). When *A. arenosa* is crossed as mother to *A. lyrata*, *AaAGL36-like* is successfully digested by *EcoRI* (Figure 4c, Table S1c). Using *A. lyrata* as a mother crossed to *A. arenosa*, no *EcoRI* digestion fragments occur and thus only expression of *AtAGL36-like* was observed (Figure 4c). As a reciprocal control, we also used *A. lyrata*-specific digestion by *TauI*, giving only digestion fragments in crosses with *A. lyrata* mothers (Figure 4c). To verify the identity of the amplified fragments and the maternal expression pattern, the undigested PCR products were Sanger sequenced and identified as only maternally contributed (Figure S5). In conclusion, only the maternal allele is expressed in reciprocal interspecies crosses between *A. arenosa* and *A. lyrata* indicating that imprinting of *AGL36-like* is preserved in *A. arenosa* × *A. lyrata* hybrid seeds.

Finally, we investigated *AGL36/AGL36-like* imprinting in *A. thaliana* × *A. arenosa* hybrid seeds, using *A. arenosa* as the paternal cross partner. Upon amplification of cDNA, both maternally expressed *AtAGL36* and paternally expressed *AaAGL36-like* fragments could be identified (Figure 4d, lanes 5–6, see figure legend). Restriction with *AlwNI* digest the maternally contributed cDNA (Figure 4d and Table S1c). The paternal *AaAGL36-like* fragments remained undigested in hybrid crosses employing two independent *A. arenosa* populations (Figure 4d, lanes 7–10). The paternal cDNA was verified by Sanger sequencing and comparison with the *A. arenosa* control (Figure 4d). We thus conclude that in hybrid *A. thaliana* × *A. arenosa* seeds, and in contrast with hybrid *A. lyrata* × *A. arenosa* seeds, the silenced paternal *A. arenosa* allele is reactivated, demonstrating differential action by the maternal species in the hybrid. Furthermore, this finding demonstrates that the paternal alleles of maternally expressed imprinted genes are deregulated in hybrid crosses, and not limited to paternally expressed genes as described in previous studies (Josefsson *et al.*, 2006; Burkart-Waco *et al.*, 2015).

Genetic and environmental factors influence post-zygotic hybrid barriers

Deregulation of MADS-box type I TF genes has been implicated in setting up the post-zygotic hybridization barrier in incompatible hybrid *A. thaliana* × *A. arenosa* (Strecno1) seeds, and *A. thaliana* mutation of some of these genes could partially rescue the severe late seed phenotype in the same hybrid cross from approximately 1 to 10% (Walia *et al.*, 2009). In order to systematically examine if the MADS-box type I TF genes analyzed in this work influence the strength of the hybrid barrier we noted that previous analyses to investigate hybrid barriers in Arabidopsis have been performed under slightly different temperature regimes (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Burkart-

Waco *et al.*, 2012; Lafon-Placette *et al.*, 2017) and in line with this, the major hypotheses to explain hybrid barriers are centered on genetic factors, not taking environmental variation into account. In rice, however, it has been demonstrated that temperature affects cellularization of the endosperm (Folsom *et al.*, 2014) and that type I MADS-box TF genes are deregulated during moderate heat stress (Chen *et al.*, 2016). To rule out a temperature effect in our experiments, we therefore repeated crosses first performed by Walia *et al.* (2009) with Strecno1 and Col-0 using the original temperature regime (22°C) and at slightly lower temperatures (18°C) and also included the *A. arenosa* accession used in this study, MJ09-4. Surprisingly, both the difference in temperature and genetic variation between *A. arenosa* populations had a major effect on the strength of the post-zygotic hybrid barrier (Figure 5a).

To quantify this observation, we first investigated seed survival in the same crosses and temperature conditions. Using *A. arenosa* accessions MJ09-4 and Strecno1 (SN1) in crosses to *A. thaliana* Col-0 at both 18°C and 22°C, a substantial increase in the survival of hybrid seeds at 18°C for both accessions was observed (Figure 5b) with 18°C MJ09-4 replicates showing up to 60% live seeds while at the same time also obtaining the same results as Walia *et al.*, 2009 when crossing Col-0 with Strecno1 at 22°C (live seed count 1%, $N = 162$) (Figure 5b).

In hybrid seed germination experiments, the temperature dependency of the strength of the hybrid barrier became even more evident for the accessions MJ09-4 and SN1 (Figure 5c; $P < 0.001$). Interestingly, when comparing two other *A. arenosa* accessions, MJ09-1 and SN2, in crosses to Col-0, these were found to be insensitive to the temperature change tested here although they display a higher seed survival rate than SN1 crossed to Col-0 (Figure 5c). These accessions may still be affected at larger differences in temperature due to genetic variation and different adaptation. Even though the variation between replicates is high, especially in 18°C crosses, a clear bypass effect of low temperature on the post-zygotic barrier is observed. Furthermore, germination of both 18°C and 22°C crosses demonstrated an increased germination rate of hybrid seeds involving the *A. arenosa* MJ09-4 population as a paternal cross partner compared with crosses with *A. arenosa* SN1, SN2 or MJ09-1 (Figure 5c). Control interspecies crosses in *A. thaliana* accessions and *A. arenosa* MJ09-4 at 18°C and 22°C displayed no significant difference in germination between the temperatures (Figure S6). Furthermore, tetraploid *A. thaliana* mothers have been shown to alleviate the hybridization barrier and to exclude this scenario we verified the diploidy of *A. thaliana* Col-0 accessions, the *A. arenosa* MJ09-4 population and the Strecno lines (SN1 and SN2) using flow cytometry (Figure S7).

Our findings indicated that genetic variation between *A. arenosa* populations also influences the success rate of hybridization, as previously demonstrated for different *A.*

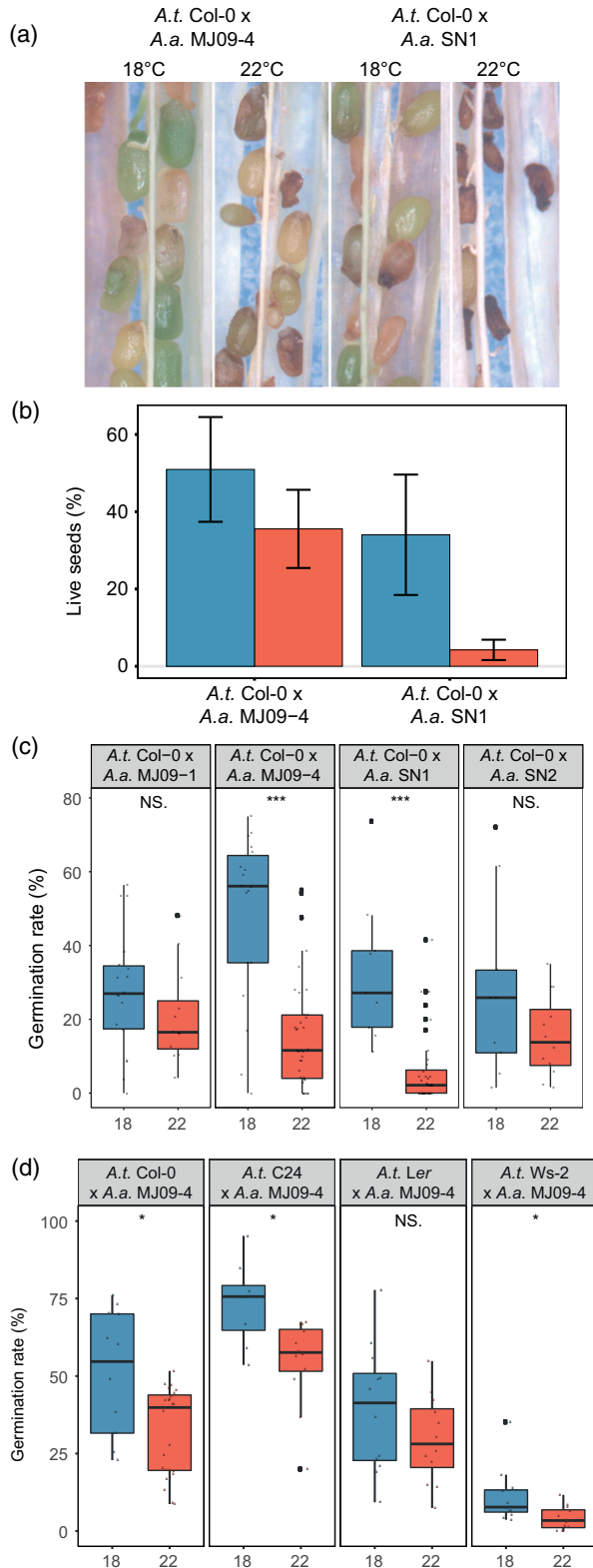


Figure 5. Temperature has a significant effect on the hybrid barrier between *A. thaliana* and *A. arenosa*. (a) Micrographs of siliques with *A. thaliana* crossed to *A. arenosa* hybrid F1 seeds grown at 18°C and 22°C 20 days after pollination (DAP). The crosses were made using two different *A. arenosa* lines, MJ09-4 and Strečno1 (SN1). Live seeds are green, collapsed seeds are brown or pale green. (b) Graph showing percentage live *A. thaliana* × *A. arenosa* hybrid F1 seeds from crosses in (a). Three biological replicates were tested for each temperature for both *A. arenosa* lines MJ09-4 and SN1 (*N* = 174, 163, 175, 162, respectively). (c) Germination rate of *A. thaliana* × *A. arenosa* hybrid F1 seeds. Four *A. arenosa* accessions were crossed to *A. thaliana* at 18°C and 22°C, MJ09-1, MJ09-4, SN1 and SN2 (18°C: MJ09-1 *N* = 18 BR (931 seeds), MJ09-4 *N* = 18 BR (986 seeds), SN1 *N* = 9 BR (524 seeds) and SN2 *N* = 9 BR (475 seeds). 22°C: MJ09-1 *N* = 12 BR (612 seeds), MJ09-4 *N* = 36 BR (1482 seeds), SN1 *N* = 36 BR (1544 seeds), SN2 = 12 BR (673 seeds)). (d) Germination rate of *A. thaliana* Col-0, Ler, C24 and Ws-2 crossed to *A. arenosa* MJ09-4 at 18°C and 22°C (18°C: Col-0 *N* = 12 BR (572 seeds), C24 *N* = 8 BR (451 seeds), Ler *N* = 12 BR (772 seeds), Ws-2 *N* = 12 BR (622 seeds). 22°C: Col-0 *N* = 24 BR (1212 seeds), C24 *N* = 12 BR (462 seeds), Ler *N* = 12 BR (751 seeds), Ws-2 *N* = 12 BR (727 seeds)). *A.t.*, *A. thaliana*; *A.a.*, *A. arenosa*. Blue color: 18°C, red color: 22°C. Outliers are plotted as large points. Dots indicate single BR. BR, biological replicate. Significance is indicated for the comparison of lines at 18°C and 22°C (Wilcoxon rank-sum test: NS: *P* > 0.05; **P* ≤ 0.05; ***P* ≤ 0.01; ****P* ≤ 0.001). Error bar indicates standard deviation (SD).

accession used in the hybrid cross. Burkart-Waco *et al.* (2012) crossed 56 accessions using the Strečno1 line at 22°C to investigate the effect of the genetic variation on the hybrid barrier and could demonstrate a weaker barrier when using C24, producing 17% normal seeds, while using Ler-1 and Ws-2 resulted in 5.2% and 3.5% normal seeds, respectively. Using Col-0 they obtained 1.7% normal seeds. Here, we demonstrate that crossing Col-0, Ler-1, C24 and Ws-2 accessions to the *A. arenosa* accession MJ09-4 give the effect of elevated seed survival for all accessions except Ws-2 (Burkart-Waco *et al.*, 2012) when comparing to the previous report using SN1 at 22°C (Figure 5d). In addition, the accessions Col-0, C24 and Ws-2 have a significant increase in seed survival when decreasing the temperature to 18°C (Figure 5d; *P* < 0.05). Ler-1 appeared to be insensitive to the temperature change, similar to the observations using *A. arenosa* accessions MJ09-1 and SN2 in combination with Col-0 (Figure 5c).

Previously published data reported that the embryo does not make the transition to the heart stage in crosses between diploid *A. thaliana* Col-0 and *A. arenosa* Strečno1 at 22°C (Burkart-Waco *et al.*, 2013), which we also could confirm for Strečno1 under our laboratory conditions at 22°C (Figure 5). Our analysis at 18°C, however, demonstrated that most seeds develop past this point (Figure 6a–i). There was a clear correlation between the severity of the hybrid barrier in *A. arenosa* accessions and the timing of endosperm cellularization in hybrid seeds (Figure 6a–i) suggesting that endosperm cellularization is the major mechanism for setting up the barrier.

The role of MADS type I loci in the *A. thaliana* × *A. arenosa* hybrid barrier

In order to investigate the specific effect of selected MADS type I loci in establishing or bridging the *A. thaliana* ×

thaliana genotypes (Burkart-Waco *et al.*, 2012). We therefore further investigated the temperature dependency of the hybrid barrier by varying the maternal *A. thaliana*

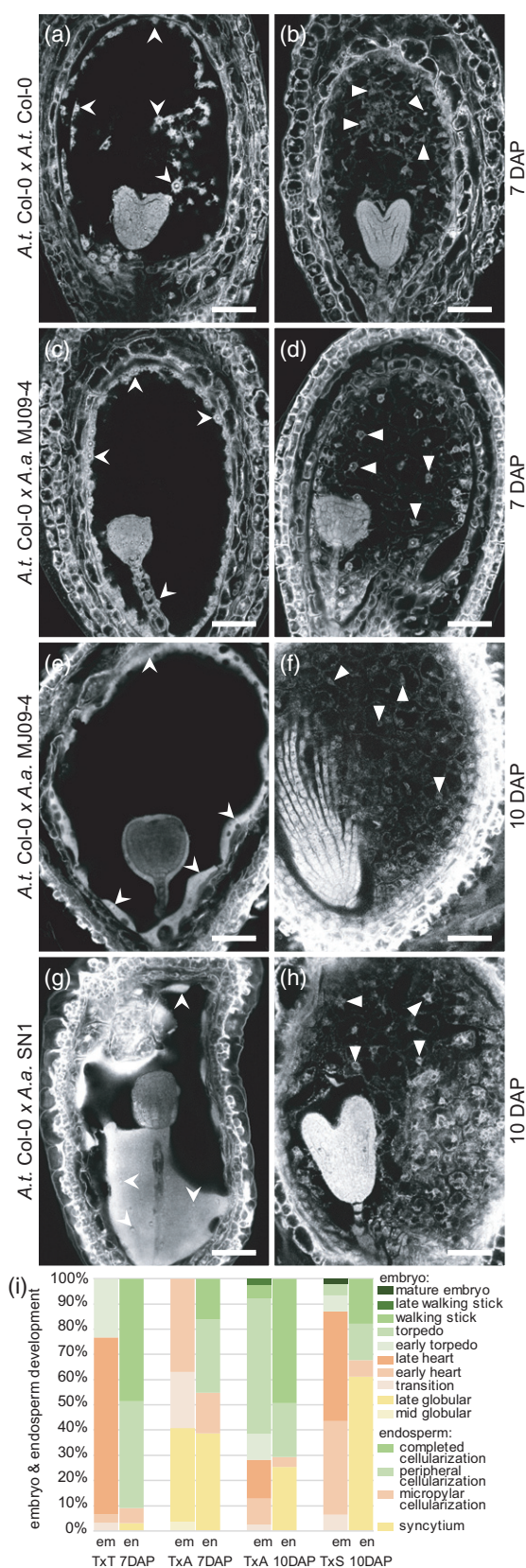


Figure 6. Variation in endosperm cellularization between *A. thaliana* and *A. arenosa* hybrids. (a–h) Confocal scanning laser micrographs of endosperm cellularization in hybrid seeds visualized by Feulgen staining. For all crosses, both non-cellularized and cellularized endosperm is observed and micrographs representative for each class are presented in the left and right panels respectively. Open arrowheads point to syncytial endosperm nuclei while closed arrowheads point to cellularized endosperm nuclei. Scale bar = 50 μ m. (a and b) *A. thaliana* control 7 days after pollination (DAP) typically at the embryo late heart stage in which most seeds display complete endosperm cellularization (b). (c, d) *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds at 7 DAP. Embryo development is slower compared with *A. thaliana* controls. Both non-cellularized (c) and cellularized endosperm (d) was frequently observed. (e, f) *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds at 10 DAP. Only a few seeds fail to cellularize (e) and most seeds exhibit completed endosperm cellularization (f). (g, h) *A. thaliana* \times *A. arenosa* SN1 hybrid seeds at 10 DAP. A higher fraction of seeds display syncytial stage endosperm (g) compared with *A. arenosa* MJ09 hybrid seeds (e, f), but some have completed endosperm differentiation (h). (i) Quantification of the described embryo and endosperm stages. All crosses are indicated as female \times male. T \times T, *A. thaliana* seeds, $N = 34$; T \times A, *A. thaliana* \times *A. arenosa* MJ09 hybrid seeds, $N = 81$; T \times S, *A. thaliana* \times *A. arenosa* SN1 hybrid seeds, $N = 98$; em, embryo stages; en, endosperm stages.

arenosa hybrid barrier, we analyzed insertional mutant alleles of the selected candidate genes (Figure S8a and Table S3). Homozygous mutants could be obtained for all investigated loci except as previously described for *AGL62* (Kang *et al.*, 2008), suggesting no vital requirement in seed development. Significantly reduced transcript levels were demonstrated in all lines with the exception of *AGL34* in which transcript levels were significantly elevated (Figure S8b). Segregation analysis could not detect reduced transmission of the mutant alleles suggestive of a requirement in male or female gametophytes or a recessive effect in embryo or endosperm (Figure S8c and Table S4). We also inspected seed size, seed germination and flowering time. For this analysis, *agl28-1* was omitted due to mixed Ws-2 Col-0 accession background (Yoo *et al.*, 2006). Only minor differences were observed in seed size and flowering time (Figure S9a,b) and no difference in germination of mutant seeds was observed (97–100%, $N = 200$). Finally, seed developmental phenotypes in single and higher order mutants were investigated, scoring live, aborted and unfertilized seeds (Figure S9c). Notably, a heterozygous *agl62-1* mutation in a double homozygous *agl28-1 agl36-1* background did not differ from single *agl62* mutants. We concluded that a thorough analysis of seed development in single, double and triple mutants of *AGL34*, *AGL35*, *AGL36* and *AGL90* including their interaction partners *AGL28* and *AGL62* did not result in any obvious seed developmental phenotypes (Figures S8 and S9).

In the case of *agl28-1* a mixed Ws-2 Col-0 accession background did not allow a direct comparison of hybrid seed barrier strength effects, due to the strong effect of the Ws-2 accession (Figure 5d, right panel). Indeed, single *agl28-1* mutants as well as double or triple mutant combinations with *agl36*, *agl62* and *agl90* crossed with *A. arenosa* all produced significantly lower seed germination

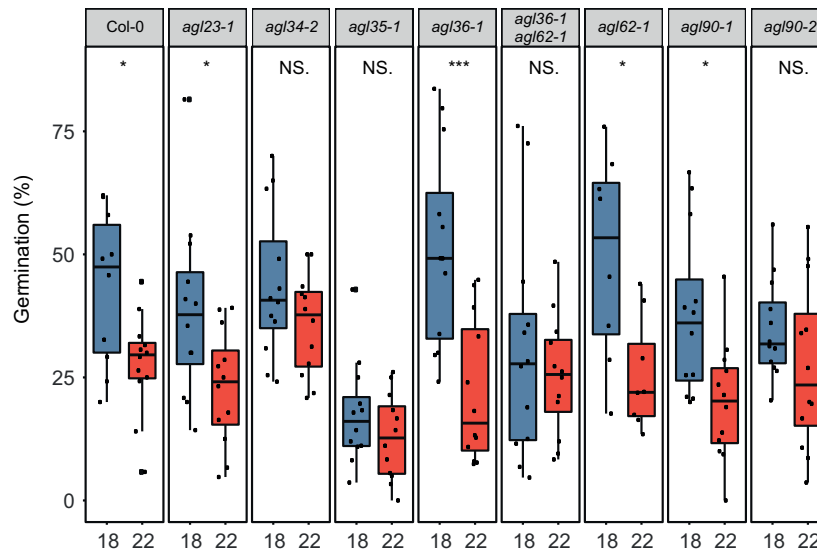


Figure 7. Genetic and environmental parameters influence the F1 hybrid barrier. (a) Germination rate of seeds from *A. arenosa* MJ09-4 crossed as pollen to the *A. thaliana* (Col-0), single mutants *agl23-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl62-1*, *agl90-1*, and *agl90-2* and the double mutant *agl36-1 agl62-1* at 18°C and 22°C. Box plot contains scattered dots representing germination rates observed per silique. Outliers are plotted as large points. Significance is indicated for the comparison of the mutant lines between 18°C and 22°C (Kruskal–Wallis test: $P < 2.2 \times 10^{-16}$; Wilcoxon rank-sum test: NS: $P > 0.05$; * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$). 18°C: Col-0 $N = 10$ BR (536 seeds), *agl23-1* $N = 12$ BR (294 seeds), *agl34-2* $N = 12$ BR (707 seeds), *agl35-1* $N = 12$ BR (641 seeds), *agl36-1* $N = 12$ BR (704 seeds), *agl36-1 agl62-1* $N = 12$ BR (532 seeds), *agl62-1* $N = 8$ BR (442 seeds), *agl90-1* $N = 12$ BR (568 seeds), *agl90-2* $N = 12$ BR (753 seeds). 22°C: Col-0 $N = 12$ BR (578 seeds), *agl23-1* $N = 12$ BR (407 seeds), *agl34-2* $N = 12$ BR (610 seeds), *agl35-1* $N = 12$ BR (571 seeds), *agl36-1* $N = 12$ BR (635 seeds), *agl36-1 agl62-1* $N = 12$ BR (498 seeds), *agl62-1* $N = 8$ BR (403 seeds), *agl90-1* $N = 12$ BR (514 seeds), *agl90-2* $N = 12$ BR (635 seeds). BR, biological replicate.

rates in nine out of 10 cross combinations (Figure S10a). In a direct comparison using *Ws-2* wild type as the control, compared with Col-0 and *agl28-1* when crossed with *A. arenosa* MJ09-4, revealed no difference in seed germination rate between *Ws-2* and *agl28-1* (Figure S10b, left). In order to further investigate the role of *agl28-1* we generated an *AGL28* genomic rescue construct that was transformed into the homozygous *agl28-1* mutant background. Six homozygous T2 lines were crossed to *A. arenosa* MJ09-4 and scored for germination (Figure S10b, right). None of the rescue lines was significantly different from *agl28-1* or the *Ws-2* control, suggesting that the observed reduced germination is caused by the *Ws-2* background alone. As the *Ws-2* background effect co-segregated with the *agl28* mutation through repeated introgression to Col-0, we hypothesized a major part of the genetic variation causing the strong *Ws-2* *A. arenosa* hybrid barrier to be linked to the *AGL28* locus. In order to genetically map the effect, we backcrossed a Col-0 introgressed hemizygous *agl28-1* to Col-0 and genotyped the progeny for the presence of *agl28-1*, and crossed the two types of plants resulting with *A. arenosa* MJ09-4 pollen donors (Figure S10c, see legend for detail). Plants wild type for the *AGL28* locus had a high probability to be Col-0 in the *AGL28* region, and did also regain Col-0 germination rates in crosses with *A. arenosa* (Figure S10c; NS, not significantly different). In addition, plants hemizygous for *agl28-1* having a high probability to be *Ws-2* Col-0 heterozygous in the *AGL28*

region, displayed intermediate germination levels and were still significantly different from Col-0 (Figure S10c; $P < 0.05$). This strongly suggested that the strong *Ws-2* effect on the hybrid barrier was linked to the *AGL28* region on top of chromosome 1.

Finally, various single and double *A. thaliana* MADS-box type I mutants were crossed with the *A. arenosa* MJ09-4 population at the two temperatures established (18°C and 22°C). First, we wanted to investigate the influence of the mutated loci on the strength of the hybrid barrier, as measured by germination rate. Secondly, we wanted to explore if the identified temperature effect on the strength of the hybrid barrier was modulated by the mutated loci, as observed using in both *A. thaliana* and *A. arenosa* accessions (Figure 5c, d).

Hybrid seed phenotypes were inspected for some representative crosses at 18°C (Figure S11a–c). We observed the seed classes normal non-collapsed seeds, collapsed brown seeds and viviparous seeds (Figure S11a). The frequency of non-collapsed seeds in a silique and seed germination frequency were well correlated (Figure S11b). Seed size measurements also displayed variation but no strong effect of the mutant mother background (Figure S11c). The ploidies of *A. thaliana* × *A. arenosa* hybrids with both Col-0 and *agl36-1 agl90-2* as mother were verified using flow cytometry (Figure S7).

In general, none of the single or higher order *A. thaliana* mutants had a significant effect to alleviate the hybrid

barrier when crossed to the *A. arenosa* MJ09-4 accession (Figure 7). This is conflicting with previous data reported for *agl62* and *agl90* crossed with an *A. arenosa* Strecno1 population at 22°C (Walia *et al.*, 2009). In the Walia study, mutants of *agl62* and *agl90* used as mothers in the cross increased the germination frequencies from 1% in crosses with Col-0 to 10% in crosses with *agl62* and *agl90*. In our experiment, the average germination frequencies of *agl62-1* at 22°C, and the average germination frequencies of *agl34-2* at 22°C were indeed slightly higher than the Col-0 control crossed to *A. arenosa*, but the difference is not significant. We could therefore not reproduce the findings of Walia *et al.* (2009) in crosses using the *A. arenosa* MJ09-4 accession. The reason for this discrepancy may be genetic differences in the *A. arenosa* accessions used (MJ09-4 in this study versus Strecno1 in Walia *et al.*, 2009). As demonstrated here, different *A. arenosa* accessions can modulate the strength of the hybrid barrier (Figure 5c).

We did however see a significant effect of the single mutant *agl35-1* to aggravate the hybrid barrier when crossed to the *A. arenosa* MJ09-4 accession (Figure 7). Significant reduction was observed at both temperatures tested. When the homozygous *agl35-1* mutant was crossed to *A. arenosa* at 18°C, a significant reduction from average values of close to 50% germination to less than 20% was observed (Figure 7; $P < 0.001$), and in the 22°C experiment we found a reduction from more than 25% to close to 15% (Figure 7; $P < 0.01$). These findings suggest that AGL35 may play a role to relieve and bypass the hybrid barrier, or that lack of AGL35 disrupts or lowers the threshold for disruption of endosperm cellularization.

It is noteworthy that the effect of lower temperature to alleviate the hybrid barrier is also bypassed by mutation of AGL35. We analyzed if any of the loci investigated played a role in establishing the observed temperature effect on the hybrid barrier strength (Figure 7). The Col-0 control and most mutant crosses, including *agl23*, *agl36*, *agl62* and *agl90*, displayed significant differences in germination frequencies between the two temperatures examined (Figure 7). Crosses with the single mutants *agl34* and *agl35* and the *agl36 agl62* double mutant, however, displayed no significant differences in its germination rate between 18° and 22°C (Figure 7), similar to the previous observation in the *Ler-1* accession cross (Figure 5d). The temperature immunity seen in *agl35* appears to be the most prominent due to low variation in the observations, and may suggest that the AGL35 links the hybrid block to the temperature effect.

DISCUSSION

The role and regulation of imprinted genes

In this report, we have systematically analyzed the function and conservation of imprinting a subset of MADS-box type

I TFs in hybrid crosses in the context of biological and environmental variance. In our expression analysis, we observed that $M\alpha$ and $M\gamma$ class type I TFs are highly represented in the transcriptome of the developing seed, and that both classes contain correlated transcript profiles that peak around the onset of endosperm cellularization. The expression peak occurs at a crucial developmental time point when the endosperm switches from nutrient sink to source for the developing embryo (Lafon-Placette and Köhler, 2016), which is suggestive of a function in this process. Defects in endosperm cellularization are also the mechanistic basis for a post-zygotic reproductive barrier between *Arabidopsis* species and thus support a role for these genes in a hybrid scenario. The $M\alpha$ and $M\gamma$ class coregulation is consistent with the notion that $M\alpha$ forms dimers with the $M\gamma$ -type MADS-box TF class (de Folter *et al.*, 2005). Furthermore, since imprinted genes interact with biparentally expressed genes this would favor the dosage hypothesis for the selection of imprinted genes (Dilkes and Comai, 2004). According to this hypothesis imprinting is a means to optimize the expression level of a gene, opposed by the parental conflict theory, in which parental conflict over resources leads to the selection of genes that promotes or restricts resource allocation (Haig and Westoby, 1991).

It has been postulated that maternally expressed imprinted genes are regulated by the release of DNA methylation in the central cell by central cell specific action of the DNA glycosylase DEMETER (DME). We could, however, demonstrate that the lack of DNA methylation maintenance in the pollen germline does not activate all silenced paternal alleles. There is a discrepancy between the lack of activation of paternal AGL36 observed in this work and previously published data (Shirzadi *et al.*, 2011). However, the two studies were done using different accessions and also different *met1* mutant alleles. Accession-specific effects are less likely as we see paternal reactivation of AGL28 using the same accessions. However, as the history of zygosity is not known for the *met1* allele used in the previous study, accumulated hypomethylation may explain the inconsistency. Nevertheless, even though the paternal allele of AGL28 reactivated in our study, paternal silencing of AGL90 and AGL36 was not lifted. In this study, *met1-7* was held as heterozygous, thus avoiding parental demethylation. This suggests that maternally expressed imprinted genes may be regulated by other mechanisms or have different regulatory requirements.

Evolution of silencing of the paternal allele of maternally expressed imprinted genes has been associated with global methylation patterns generated toward suppressing transposons (Kim and Zilberman, 2014; Anderson and Springer, 2018). *De novo* methylation of transposons is mainly performed by the action of RdDM, and in the lack of a mechanistic scenario for the imprinting of maternally

expressed genes this pathway is an obvious candidate. Several $M\alpha$ and $M\gamma$ class type I TFs including *AGL36* have also been suggested to be upregulated in the lack of RdDM (Lu *et al.*, 2012). For *AGL36* we could, however, not observe any change in paternal silencing in reciprocal crosses with a PolIV mutant (*nprp1*) that is blocking canonical RdDM. Furthermore, we were also unable to verify the previously reported upregulation of *AGL36* in the same crosses and time points (Lu *et al.*, 2012). To determine the global effect, however, a systematic elucidation of the role of RdDM in the regulation of maternally expressed imprinted genes is required.

Although it does not regulate imprinting of maternally expressed genes, we showed that repression by PRC2 is specifically targeted toward $M\alpha$ and $M\gamma$ class type I TFs and acts to repress gene activity concurrent with and at post-cellularization stages. In the absence of MEA, only the maternal allele of *AGL36* is upregulated (Shirzadi *et al.*, 2011), indicating that the mode of regulation does not interfere with the actual imprinting mechanism. We hypothesize that DNA methylation of the paternal allele protects from PRC2 repression, as suggested for the paternally expressed imprinted gene *PHE1* (*AGL37*) (Köhler *et al.*, 2003; Makarevich *et al.*, 2008; Villar *et al.*, 2009). In such a scenario the paternal allele of paternally expressed genes should be targeted by PRC2 at cellularization, but the mechanism here remains to be explored. We observed, however, that $M\alpha$ and $M\gamma$ class type I TFs are regulated by PRC2 in distinct clusters from 6 to 12 DAP and that maternally and paternally expressed genes are co-regulated in the same cluster. While the meaning of the observed gradual repression by FIS-PRC2 in the context of $M\alpha$ and $M\gamma$ role in seed development remains open to speculation, it seems clear that FIS-PRC2 acts through specific sets of genes at successive stages in seed development, rather than exerting a global effect in the seed concurrent with endosperm cellularization.

Species-dependent deregulation of imprinting

We identified orthologs of *AGL36* in *A. arenosa* and *A. lyrata* and studied imprinting in *A. arenosa* and hybrids of *A. arenosa*, *A. lyrata* and *A. thaliana*. *AGL36* is imprinted in *A. thaliana* and maternally expressed, and here we prove that one of the two *AGL36-like* genes in *A. arenosa* is also exclusively maternally expressed, demonstrating conservation of imprinting. The maternal allele of the paternally expressed imprinted genes have previously been shown to be reactivated in the hybrid of *A. thaliana* and *A. arenosa* (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Burkart-Waco *et al.*, 2015). We show that, in hybrid *A. thaliana* \times *A. arenosa* seeds, the paternal *AGL36-like* allele is reactivated, suggesting that deregulation in hybrid crosses is not limited to paternally expressed genes as described in previous studies (Walia *et al.*, 2009). This loss of regulation is in

strong contrast with the *A. lyrata* \times *A. arenosa* hybrid cross in which the imprinting of *AGL36-like* is maintained.

The variation or perturbation of the pattern of imprinting might play a role in the endosperm hybridization barrier between species (Florez-Rueda *et al.*, 2016), and in the case described here, different mechanisms may act in the same species depending on the cross partner. Previous imprinting analyses involved crosses between inbreeders and outbreeders, fueling speculation that the mechanisms regulating imprinted genes may differ depending on the mating system (Josefsson *et al.*, 2006; Burkart-Waco *et al.*, 2012; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016). However, intraspecific variation of imprinting within *A. thaliana* has previously been demonstrated (Waters *et al.*, 2013; Pignatta *et al.*, 2014). Consistent with the notion that imprinting can vary independently of mating systems, reactivation of the normally imprinted paternal *A. arenosa* *AGL36* allele depended on the maternal species: it was observed only in crosses to *A. thaliana* but not to *A. lyrata*.

Temperature- and population-dependent quantitative barrier phenotype

We have identified an important role of temperature in the establishment of endosperm-based post-zygotic hybrid barriers. This opens for speculation and a multifaceted scenario emerges identifying several parameters; both intrinsic variation in genetic pathways in both parents and external abiotic factors such as temperature may act in concert to mediate the generation of post-zygotic species barriers.

The discovery that lowering the temperature by only 4°C from 22°C to 18°C during the fertilization and development of the *A. thaliana* \times *A. arenosa* hybrid seed increases survival was surprising. Nevertheless, incompatibility between diverged individuals can be sensitive to temperature as, for example, in seedling hybrid necrosis (Bomblied and Weigel, 2007). In this case, however, appropriate temperatures can ameliorate acute incompatibilities manifested only during early seed development and, therefore, may play a significant role in reducing interspecific hybridization barriers. High environmental sensitivity may also explain inconsistencies with previous studies noted in Results.

Crossing *A. thaliana* wild type Col-0 and mutants to different *A. arenosa* accessions demonstrated variation depending on the pollen donor and no significant increase in seed viability could be observed by using the mutants of the MADS-box type I TFs as mothers. The variability caused by natural variation has been demonstrated in *A. thaliana* by using different accessions. Burkart-Waco *et al.* (2012) used the diploid *A. arenosa* accession Strecno-1 and crossed it as the pollen donor to 56 *A. thaliana* accessions and scored seed viability ranging from 30% normal seeds to close to 2%. Burkart-Waco *et al.* (2013) compared seed

development of *A. thaliana* accessions Col-0 and C24 crossed to *A. arenosa*, which respectively produced 0 to 1% and ~17% live hybrid seeds. Hybrid embryos at all stages beyond 2 DAP were developmentally delayed and similar between Col-0 and C24, with the exception of a few C24 × *A. arenosa* hybrid embryos displaying developmental progression to heart stage by 6 DAP, whereas no Col-0 embryos made this transition. This is in clear contrast with our findings where most seed survive this stage. Using different *A. thaliana* accessions in crosses to the *A. arenosa* accession MJ09-4, we obtained highest seed germination rates in crosses with C24. However, the cross combination with the Ws-2 accession produced similarly low germination rates as the *A. arenosa* accession Strecno-1 crossed with Col-0. This indicates that rather than a specific accession having specific effects, it is the combination of accessions that determines the strength of the hybrid barrier. Although we could not observe elevated seed viability when single or multiple MADS-box type I mutants were used in crosses to *A. arenosa*, deficiency of *AGL35* resulted in significant reduction of the germination frequency. Notably, mutants of *AGL35* crossed to *A. arenosa* display two effects: first, it increase lethality; second, it decrease or eliminate the temperature effect, suggesting a critical role for this gene in mediating the strength of the hybrid block and the temperature effect.

Weakening of hybridization barriers at lower temperatures might increase fitness of a self-incompatible colonizer by broadening mate choice: few viable seeds are better than no seed. This mechanism could, for instance, have been important for recolonization after the Pleistocene glaciations, a period that was characterized by secondary contact and high amount of hybridization with or without genome duplication (Stebbins, 1984; Brochmann *et al.*, 2004). Although *A. thaliana* is a self-crosser, the mechanism might be ancestral and predate *A. thaliana*. Its occurrence should be investigated by hybridizing more species under varying temperature. In northern Europe and the Fennoscandian region, *A. thaliana* with unreduced gametes has most likely hybridized with pollen from tetraploid *A. arenosa* to create the allotetraploid *A. suecica* on multiple hybridization events (Novikova *et al.*, 2017). Making a synthetic tetraploid *A. thaliana* and crossing it to *A. arenosa*, can make viable, although unstable, hybrids (Comai *et al.*, 2000). Environmental stress such as heat or cold may increase the rate of unreduced gamete formation (De Storme and Mason, 2014) and facilitate hybridization between diploids and tetraploids (Vallejo-Marin and Hiscock, 2016). Such instances of genome doubling, however, did not occur in our experiments as flow cytometry of sampled hybrids indicated genome content consistent with reduced gametes of the diploid parents *A. thaliana* and *A. arenosa*. Formation of 2N gametes may be under different regulation in Arabidopsis. The temperature-sensitive mechanism affecting the endosperm-

based barrier and its dependency on AGL function remain an open area of investigation.

In rice, it has been shown that temperature affects cellularization of the endosperm (Folsom *et al.*, 2014) and that type I MADS-box TF genes are deregulated during moderate heat stress (Chen *et al.*, 2016). The temperature stress tested were much higher than tested here for Arabidopsis, but the different species have different temperature adaptations in general. The rice MADS-box TF OsMADS87 is a heat-sensitive imprinted gene which is associated with syncytial stage endosperm and regulates rice seed size (Chen *et al.*, 2016). OsMADS87 is a putative ortholog of Arabidopsis *PHE1*. Mutants of OsMADS87 have accelerated endosperm cellularization and lower sensitivity to a moderate heat stress in terms of seed size (Folsom *et al.*, 2014; Chen *et al.*, 2016). Our results indicated that some of the *Arabidopsis thaliana* MADS-box mutants hybridized to *A. arenosa*, display a lowered heat sensitivity that may indicate a temperature-sensing role for the MADS-box genes in the endosperm of *A. thaliana* as well. Considering that many of the type I MADS-box genes are regulated by the PRC2 complex and that the finding that OsFIE1 is imprinted and temperature sensitive in rice seeds (Folsom *et al.*, 2014), this proposes an epigenetic regulation during hybrid seed development which is altered during environmental perturbations (Folsom *et al.*, 2014). Given the importance of overcoming post-zygotic isolation for the early stages of neo-hybridization (Vallejo-Marin and Hiscock, 2016), this temperature effect can be a useful tool for investigating the endosperm-based post-zygotic barrier and early speciation.

EXPERIMENTAL PROCEDURES

Plant material and cultivation

A. thaliana accessions and mutant plant lines were obtained from the Nottingham Arabidopsis Stock Center (NASC) unless specified otherwise. For details on MADS-box type I mutant lines see Table S1. The *met1-7* and *nrdp1* accession numbers are SALK_076522 and SALK_083051, respectively. The *mea/fis1* mutant was kindly supplied and described in Chaudhury *et al.* (1997). The *A. arenosa* populations MJ09-1 and MJ09-4 and the *A. lyrata* MJ09-11 descended from natural populations in central Europe as described, respectively, by Jørgensen *et al.* (2011) and Lafon-Placette *et al.* (2017). *A. arenosa* populations Strecno1 (SN1) and Strecno2 (SN2) were kindly supplied by Kirsten Bomblies (Hollister *et al.*, 2012). Seeds were surface sterilized either by washing steps with 70% ethanol, 20% bleach and wash solution (0.001% Tween20) or by over-night chlorine gas sterilization (Lindsey *et al.*, 2017), sown out on 0.5 Murashige and Skoog (MS) plates (Murashige and Skoog, 1962) supplemented with 2% sucrose and appropriate antibiotics for selection of mutant lines. The seeds were then stratified over-night (*A. thaliana*) or 1 to 3 weeks (*A. arenosa* and *A. lyrata*) at 4°C before transferring to growth chambers with either 18°C or 22°C under long-day conditions (16 h light, 160 µmol/m²/sec, relative humidity 60–65%). *A. thaliana* *A. arenosa* F1 hybrid seeds were stratified at 4°C for 4–6 days before being placed in growth chambers for scoring of germination as seedling root protruding from the seed regardless

of survival at later stages. Germinated seedlings were transferred to soil and grown under long-day conditions at 18°C or 22°C. *A. arenosa* and *A. lyrata* plants were vernalized at 8°C under short-day conditions (10 h light) for 4–5 weeks to stimulate flowering. To avoid self-pollination, controlled crosses were performed by emasculating unopened flower buds followed by hand pollination after 2 days. Developing or mature seeds were harvested for designated purposes at defined time points. See Table S5 for an overview of interspecies crosses. For flowering time analysis, flowering time was scored as day after stratification and the average number of leaves at stem emergence from the rosette.

Tissue handling, DNA and RNA extraction and cDNA synthesis

Tissue was harvested directly in liquid nitrogen and DNA was isolated using E.Z.N.A. Plant DNA kit (Omega) according to manufacturer's instructions. For total RNA isolation, seeds were hand dissected from siliques directly into pre-chilled tubes with MagNA Lyser Green Beads (Roche) and ground in lysis buffer (Sigma Plant total RNA kit) using a MagNA Lyser Instrument (Roche). Isolated RNA was treated with DNase I (Sigma) and cDNA synthesized with oligo(dT) and Superscript III reverse transcriptase (Invitrogen). Samples were cleaned using a QIAquick PCR purification kit (Qiagen). DNA or RNA concentration was measured using a NanoDrop1000 spectrophotometer or RNA was measured with a Qubit 3 fluorometer (ThermoFisher) using the Qubit RNA BR Assay kit (Invitrogen). All kits were used according to the manufacturers' instructions.

RNA Sequencing and sequence analysis

Total RNA was isolated from dissected seeds at 1, 2, 3, 4, 6, 9 or 12 DAP from *Ler* crossed to *Col-0* and from *meal/fis1* crossed to *Col-0* in two biological replicates as described previously (Shirzadi *et al.*, 2011). RNA samples were DNase treated before quality checked using an Agilent 2100 Bioanalyzer. Total RNA was prepared to a strand-specific TruSeq™ RNA-seq library and all 28 samples sequenced over three lanes on an Illumina HiSeq 4000, 150 bp paired end reads. Differential expression analysis was performed with RSEM (Li and Dewey, 2011) using the edgeR software package (McCarthy *et al.*, 2012). The expression profiles were analyzed and visualized using the Tidyverse 1.2.1, ComplexHeatmap 1.17.1 (Gu *et al.*, 2016), Dendextend 1.7.0 (Galili, 2015), Viridis 0.5.0 packages in R version 3.4.3.

Molecular cloning and genotyping

All T-DNA mutant lines were genotyped using specific primers (Tables S1 and S2). Due to high sequence similarities between MADS-box genes and AGL36-like genes in different species, primers were optimized to ensure specific amplification and fragments were sequenced for confirmation of identity. The sequencing and characterization of AGL36-like genes from MJ09 *A. arenosa* and *A. lyrata* lines were performed by PCR amplification with designated primers (Table S2) using KOD Hot Start DNA polymerase (Sigma) according to the manufacturer's instructions with 1.5 mM of MgSO₄ and PCR program with 95°C denaturation, 55°C annealing, and 70°C extension for 35 cycles. The amplified fragment was subsequently cloned into a TOPO Blunt pCR Zero vector (Invitrogen) according to the manufacturers' instructions. The AGL28 genomic complementation construct was created by nested PCR using primers described in Table S2 containing *att* sites for Gateway cloning (Invitrogen) according to the manufacturer's instructions. The genomic AGL28 fragment was 3531 bp

including 2000 bp upstream of the start codon and 500 bp downstream of the stop codon and was cloned into the destination vector pMDC99. The construct was transformed into *Agrobacterium tumefaciens* strain GV3101 pMP90RK which was used to transform *agl28-1* mutant using the floral dip method (Clough and Bent, 1998) and transformants were selected for by the appropriate resistance encoded in the inserted T-DNA, hygromycin. All sequences from *A. arenosa* and *A. lyrata* generated for this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/>) with accession numbers MN380433 to MN380437.

Phylogenetic analysis

All alpha and gamma MADS-box genes were extracted from the *A. thaliana* genome (TAIR10 at <https://www.arabidopsis.org/>). The genomic sequence was translated into amino acids with the AUGUSTUS gene prediction program (Stanke and Morgenstern, 2005) using the *A. thaliana* gene model. Nucleotides from the coding regions were aligned based on the protein sequence with PAL2NAL (Suyama *et al.*, 2006). Positions with more than 80% gaps and ambiguously aligned positions were removed from the alignment. A phylogenetic tree was inferred from the resulting alignment using the GTRGAMMA model and the automatic bootstrapping criteria MRE (option -l autoMRE) in RAxML v8.0.26 (Pattengale *et al.*, 2011; Stamatakis, 2014). The genes AGL34, AGL36 and AGL90 from *A. thaliana* were used as queries in BLAST against the non-redundant nucleotide database at NCBI (blast.ncbi.nlm.nih.gov/) to find all homologous genes in the Arabidopsis genus. In addition, all available Arabidopsis Illumina whole genome sequence data from Sequence Read Archive (SRA) was employed in the phylogenetic analysis. These libraries were assembled with rna-spades (Bankevich *et al.*, 2012) and AGL36 related genes were identified with BLAST. All AGL36 related genes were aligned and phylogenetic trees were inferred as for the MADS-box genes. In the final analysis only one copy of the gene was kept for each species.

Real-time quantitative PCR

Real-time PCR was performed on a LightCycler 96 instrument using FastStart Essential DNA Probes Master protocol (Roche) and FastStart Essential DNA Green Master protocol (Roche) using designated primers (Table S2). Relative expressions were calculated according to Pfaffl (2001) and are average values of at least two biological replicates. In reactions with low relative expression an E-value of 2.0 was used. All PCR products were sequenced to verify identity of the product amplified.

Single nucleotide polymorphism analysis

Seed tissue was sampled at 4 DAP from reciprocal crosses from *A. thaliana*, at 7 DAP from crosses between *A. thaliana* and *A. arenosa*, and at 9 DAP from reciprocal crosses of *A. arenosa* and *A. lyrata*. cDNA was amplified by PCR using designated primers (Table S2) and digested using SNP-specific enzymes (Table S4) analyzed on an Agilent 2100 Bioanalyzer using the DNA-1000-LabOnChip system (Agilent Technologies). Images for figures were assembled using Illustrator software (Adobe).

Dry seed phenotyping using ImageJ

For *A. thaliana* *A. arenosa* hybrid crosses, mature seeds were harvested one silique at a time before imaging using a Nikon D90 and analysis using ImageJ to determine seed size (mm²). Viviparous seeds were excluded from the analysis. MADS-box type I mutants were grown and harvested at the same time and

conditions. Seed size was measured and analyzed using ImageJ to determine the average seed size (mm²) per plant. The significance differences between plant lines were tested using the Kruskal–Wallis test: $P = 0.0047$ and in pairwise comparison with wild type (Wilcoxon rank-sum test).

Microscopy and Feulgen staining of seeds

Tissue was harvested from *A. thaliana* and *A. thaliana* × *A. arenosa* at 7 DAP and 10 DAP (seeds from three siliques per biological replicate). The seeds were stained with Schiff's reagent (Sigma-Aldrich S5133) following fixation and embedding in LR White (London Resin) as described by Braselton *et al.* (1996). An Olympus FluoView 1000 Confocal laser scanning microscope (BX61WI) with an excitation of 488 nm and emission from 500 to 600 nm was used for imaging.

Ploidy measurements of adult hybrids as well as crossing parents

Ploidy was measured for a selection of the plants with two of the *A. arenosa* parents (from MJ09-4 population), two Col-0 plant individuals, 14 *A. thaliana* Col-0 × *A. arenosa* F1 hybrids and 12 *A. thaliana agl36-1 agl90-2* × *A. arenosa* F1 hybrids. Also, the *A. arenosa* lines Strecno1 and Strecno2 were included to confirm ploidy. One rosette leaf and one inflorescence were analyzed for all samples except for *agl36-1 agl90-2* × *A. arenosa* F1 hybrids in which only rosette leaves were analyzed. The ploidy was assessed by establishing the genome content by estimating the relative fluorescence intensities by flow cytometry (FCM) and the two-step methodology according to Dolezel *et al.* (2007). The reference standards for the raw cytometric analysis were *Solanum pseudocapsicum* for the *A. thaliana* and *A. arenosa* hybrid comparison with *A. thaliana*, and *Carex acutiformis* for the *A. arenosa* Strecno1 ($N = 10$) and Strecno2 ($N = 12$). The samples and the internal reference were chopped with a razor blade in 0.5 ml ice-cold Otto I buffer (0.1 M citric acid, 0.5% Tween 20). This was then filtered through a nylon mesh (loop size 0.42 μm), incubated at room temperature for 5 min before being stained with 1 ml of Otto II buffer (0.4 M Na₂HPO₄ · 12 H₂O) supplemented with AT-selective fluorescent dye DAPI (4',6-diamino-2-phenylindol) and 2-mercaptoethanol in final concentrations of 4 μg/ml and 2 μl/ml, respectively. After about 5 min of incubation at room temperature, the relative fluorescence intensity for a minimum of 3000 nuclei was recorded using a Partec Space flow cytometer (Partec GmbH, Münster, Germany) equipped with an UV-LED chip (365 nm). The FCM results are the fluorescence intensities relative to unit fluorescence intensity of the internal reference standard.

ACCESSION NUMBERS

All sequences generated in this study have been deposited in the National Center for Biotechnology Information SRA (<https://www.ncbi.nlm.nih.gov/sra/>) with project number PRJNA562212.

ACKNOWLEDGMENTS

We thank Kirsten Bomblies for supplying *A. arenosa* Strecno1 and Strecno2 populations. Jason Miller and Harinder Singh are acknowledged for RNA-seq analysis. This work was supported by Fri prosjektstøtte grants 214052 and 262247 from the Norwegian Research Council to PEG and AKB.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

KNB and PEG conceived the project and wrote the article with contributions of all the authors. KNB designed the experiments. KNB, JB, KSH and IMJ performed experiments. KNB, JB and KSH analyzed the data. KSH, IMJ, AKK, YSE, MK and RS provided technical assistance. KNB, JB and KSH made figures. JB, AKB and LC contributed to project design and discussion. PEG supervised the project and agrees to serve as the author responsible for contact and ensures communication.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phylogeny and expression of MADS-box type I transcription factors during seed development.

Figure S2. Imprinting analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90*.

Figure S3. Clustering of MADS-box type I genes based on expression pattern in wild type and *mea* mutant seeds.

Figure S4. *AGL36*-related genes from all available Arabidopsis genomes in Sequence Read Archive, GenBank and Phytozome.

Figure S5. Imprinting analysis of *AGL36-like* in the reciprocal cross of *A. arenosa* and *A. lyrata*.

Figure S6. Germination rate in self crosses of *A. thaliana* accessions and *A. arenosa* MJ09-4 is not affected by temperature.

Figure S7. Ploidy measurement of Arabidopsis populations and hybrids.

Figure S8. Genetic analysis of selected MADS-box type I transcription factors.

Figure S9. Phenotypic characterization of MADS-box transcription factor mutants.

Figure S10. Characterization of a genetic background effect in mixed *A. thaliana* accessions crossed to *A. arenosa*.

Figure S11. F1 hybrid seed phenotypes from *A. arenosa* crosses to *A. thaliana* Col-0, *agl36-1*, *agl90-1*, *agl90-1* and *agl36-1 agl90-2* at 18°C.

Table S1. Schematic overview of restriction digest set-up

Table S2. Oligonucleotide name, sequence and description

Table S3. Characterization of mutants used in this study

Table S4. Segregation of the mutant alleles *agl28-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1* and *agl90-2* in self crosses and in reciprocal crosses to wild type (Col-0)

Table S5. Crossing scheme for various *A. thaliana* to *A. arenosa* experiments

OPEN RESEARCH BADGE



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA562212>. All sequences generated in this study have been deposited in the National Center for Biotechnology Information Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) with project number PRJNA562212.

REFERENCES

- Airoldi, C.A. and Davies, B. (2012) Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genom.* **39**(4), 157–165.
- Anderson, S.N. and Springer, N.M. (2018) Potential roles for transposable elements in creating imprinted expression. *Curr. Opin. Genet. Dev.* **49**, 8–14.
- Bankevich, A., Nurk, S., Antipov, D. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**(5), 455–477.
- Belmonte, M.F., Kirkbride, R.C., Stone, S.L. et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc. Natl Acad. Sci. USA*, **110**, E435–E444.
- Bemer, M., Wolters-Arts, M., Grossniklaus, U. and Angenent, G. (2008) The MADS domain protein DIANA acts together with AGAMOUS-LIKE80 to specify the central cell in Arabidopsis ovules. *Plant Cell*, **20**(8), 2088–2101.
- Bemer, M., Heijmans, K., Airoldi, C., Davies, B. and Angenent, G.C. (2010) An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiol.* **154**(1), 287–300.
- Berger, F., Grini, P.E. and Schnittger, A. (2006) Endosperm: an integrator of seed growth and development. *Curr. Opin. Plant Biol.* **9**, 664–670.
- Bombliès, K. and Weigel, D. (2007) Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat. Rev. Genet.* **8**(5), 382–393.
- Braselton, J.P., Wilkinson, M.J. and Clulow, S.A. (1996) Feulgen staining of intact plant tissues for confocal microscopy. *Biotech. Histochem.* **71**(2), 84–87.
- Brekke, T.D., Henry, L.A. and Good, J.M. (2016) Genomic imprinting, disrupted placental expression, and speciation. *Evolution*, **70**(12), 2690–2703. <https://doi.org/10.1111/evo.13085>.
- Brochmann, C., Brysling, A.K., Alsos, I.G., Borgen, L., Grundt, H.H., Scheen, A.C. and Elven, R. (2004) Polyploidy in arctic plants. *Biol. J. Lin. Soc.* **82**(4), 521–536.
- Burkart-Waco, D., Josefsson, C., Dilkes, B., Kozloff, N., Torjek, O., Meyer, R., Altmann, T. and Comai, L. (2012) Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiol.* **158**(2), 801–812.
- Burkart-Waco, D., Ngo, K., Dilkes, B., Josefsson, C. and Comai, L. (2013) Early disruption of maternal-zygotic interaction and activation of defense-like responses in Arabidopsis interspecific crosses. *Plant Cell*, **25**(6), 2037–2055.
- Burkart-Waco, D., Ngo, K., Lieberman, M. and Comai, L. (2015) Perturbation of parentally biased gene expression during interspecific hybridization. *PLoS ONE*, **10**(2), e0117293.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S. and Peacock, W.J. (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **94**(8), 4223–4228.
- Chen, C., Begcy, K., Liu, K., Folsom, J.J., Wang, Z., Zhang, C. and Walia, H. (2016) Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* **171**(1), 606–622. <https://doi.org/10.1104/pp.15.01992>.
- Chen, C., Li, T., Zhu, S. et al. (2018) Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiol.* **177**(4), 1754–1771.
- Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.
- Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. and Colombo, L. (2008) AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *Plant J.* **54**(6), 1037–1048.
- Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y. and Byers, B. (2000) Phenotypic instability and rapid gene silencing in newly formed Arabidopsis Allotetraploids. *Plant Cell*, **12**(9), 1551–1568.
- De Storme, N. and Mason, A. (2014) Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* **1**, 10–33.
- Dilkes, B.P. and Comai, L. (2004) A differential dosage hypothesis for parental effects in seed development. *Plant Cell*, **16**(12), 3174–3180.
- Dolezel, J., Greilhuber, J. and Suda, J. (2007) Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* **2**(9), 2233–2244.
- Florez-Rueda, A.M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U. and Stadler, T. (2016) Genomic imprinting in the endosperm is systematically perturbed in abortive hybrid tomato seeds. *Mol. Biol. Evol.* **33**(11), 2935–2946.
- Folsom, J.J., Begcy, K., Hao, X., Wang, D. and Walia, H. (2014) Rice fertilization-independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* **165**(1), 238–248. <https://doi.org/10.1104/pp.113.232413>.
- de Folter, S., Immink, R.G.H., Kieffer, M. et al. (2005) Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell*, **17**(5), 1424–1433.
- Galili, T. (2015) dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, **31**(22), 3718–3720.
- Gehring, M. and Satyaki, P.R. (2017) Endosperm and imprinting, inextricably linked. *Plant Physiol.* **173**(1), 143–154.
- Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**(18), 2847–2849.
- Haig, D. and Westoby, M. (1989) Parent-specific gene-expression and the triploid endosperm. *Am. Nat.* **134**(1), 147–155.
- Haig, D. and Westoby, M. (1991) Genomic imprinting in endosperm: its effect on seed development in crosses between species, and between different ploidy levels of the same species, and its implications for the evolution of Apomixis. *Phil. Trans.: Biol. Sci.* **333**(1266), 1–13.
- Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. (2016) Rapid evolution of genomic imprinting in two species of the Brassicaceae. *Plant Cell*, **28**(8), 1815–1827.
- Hollister, J.D., Arnold, B.J., Svedin, E., Xue, K.S., Dilkes, B.P. and Bombliès, K. (2012) Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet.* **8**, e1003093.
- Hornslien, K.S., Miller, J.R. and Grini, P.E. (2019) Regulation of parent-of-origin allelic expression in the endosperm. *Plant Physiol.* **180**(3), 1498–1519. <https://doi.org/10.1104/pp.19.00320>.
- Jørgensen, M.H., Ehrich, D., Schmickl, R., Koch, M.A. and Brysling, A.K. (2011) Interspecific and interplotal gene flow in Central European Arabidopsis (Brassicaceae). *BMC Evol. Biol.* **11**(1), 346.
- Josefsson, C., Dilkes, B. and Comai, L. (2006) Parent-dependent loss of gene silencing during interspecific hybridization. *Curr. Biol.* **16**(13), 1322–1328.
- 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. *Plant Cell*, **20**(3), 635–647.
- Kim, M.Y. and Zilberman, D. (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* **19**(5), 320–326.
- Klosinska, M., Picard, C.L. and Gehring, M. (2016) Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. *Nat. Plants*, **2**, 16145.
- Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. and Grossniklaus, U. (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* **17**(12), 1540–1553.
- 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. *Nat. Genet.* **37**(1), 28–30.
- Lafon-Placette, C., Johannessen, I.M., Hornslien, K.S. et al. (2017) Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc. Natl Acad. Sci. USA*, **114**(6), E1027–E1035.
- Lafon-Placette, C. and Köhler, C. (2016) Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Mol. Ecol.* **25**(11), 2620–2629. <https://doi.org/10.1111/mec.13552>.
- Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
- Lindsey, B.E. 3rd, Rivero, L., Calhoun, C.S., Grotewold, E. and Brkljacic, J. (2017) Standardized method for high-throughput sterilization of Arabidopsis seeds. *J. Vis. Exp.* (128), 56587. <https://doi.org/10.3791/56587>
- Lu, J., Zhang, C., Baulcombe, D.C. and Chen, Z.J. (2012) Maternal siRNAs as regulators of parental genome imbalance and gene expression in

- endosperm of Arabidopsis seeds. *Proc. Natl Acad. Sci. USA*, **109**(14), 5529–5534.
- Makarevich, G., Villar, C.B.R., Erilova, A. and Köhler, C.** (2008) Mechanism of PHERES1 imprinting in Arabidopsis. *J. Cell Sci.* **121**(6), 906–912.
- Masiero, S., Colombo, L., Grini, P.E., Schnittger, A. and Kater, M.M.** (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*, **23**(3), 865–872.
- McCarthy, D.J., Chen, Y. and Smyth, G.K.** (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**(10), 4288–4297.
- Murashige, T. and Skoog, F.** (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.
- Nam, J., Kim, J., Lee, S., An, G., Ma, H. and Nei, M.** (2004) Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proc. Natl Acad. Sci. USA*, **101**(7), 1910–1915.
- Novikova, P.Y., Tsuchimatsu, T., Simon, S. et al.** (2017) Genome sequencing reveals the origin of the allotetraploid *Arabidopsis suecica*. *Mol. Biol. Evol.* **34**(4), 957–968.
- Nowack, M.K., Ungru, A., Bjerkan, K.N., Grini, P.E. and Schnittger, A.** (2010) Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* **38**, 604–612.
- Parenticova, L., de Folter, S., Kieffer, M. et al.** (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*, **15**(7), 1538–1551.
- Pattengale, N.D., Aberer, A.J., Swenson, K.M., Stamatakis, A. and Moret, B.M.** (2011) Uncovering hidden phylogenetic consensus in large data sets. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**(4), 902–911.
- Pfaffl, M.W.** (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**(9), e45.
- Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W. and Gehring, M.** (2014) Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife*, **3**, e03198.
- Rebernik, C.A., Lafon-Placette, C., Hatorangan, M.R., Slotte, T. and Köhler, C.** (2015) Non-reciprocal interspecies hybridization barriers in the *Capsella* genus are established in the endosperm. *PLoS Genet.* **11**(6), e1005295.
- Rodrigues, J.A. and Zilberman, D.** (2015) Evolution and function of genomic imprinting in plants. *Genes Dev.* **29**(24), 2517–2531. <https://doi.org/10.1101/gad.269902.115>.
- Roth, M., Florez-Rueda, A.M. and Stadler, T.** (2019) Differences in effective ploidy drive genome-wide endosperm expression polarization and seed failure in wild tomato hybrids. *Genetics*, **212**(1), 141–152. <https://doi.org/10.1534/genetics.119.302056>.
- Saze, H., Mittelsten Scheid, O. and Paszkowski, J.** (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **34**, 65–69.
- Shirzadi, R., Andersen, E.D., Bjerkan, K.N. et al.** (2011) Genome-wide transcript profiling of endosperm without paternal contribution identifies parent-of-origin-dependent regulation of AGAMOUS-LIKE36. *PLoS Genet.* **7**(2), e1001303.
- Stamatakis, A.** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9), 1312–1313.
- Stanke, M. and Morgenstern, B.** (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467.
- Stebbins, G.** (1984) Polyploidy and the distribution of the arctic–alpine flora: new evidence and a new approach. *Bot. Helv.* **94**, 1–13.
- Steffen, J., Kang, I., Portereiko, M., Lloyd, A. and Drews, G.** (2008) AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. *Plant Physiol.* **148**(1), 259–268.
- Suyama, M., Torrents, D. and Bork, P.** (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612.
- Tonosaki, K., Sekine, D., Ohnishi, T., Ono, A., Furuumi, H., Kurata, N. and Kinoshita, T.** (2018) Overcoming the species hybridization barrier by ploidy manipulation in the genus *Oryza*. *Plant J.* **93**(3), 534–544.
- Vallejo-Marin, M. and Hiscock, S.J.** (2016) Hybridization and hybrid speciation under global change. *New Phytol.* **211**(4), 1170–1187.
- Villar, C.B., Erilova, A., Makarevich, G., Trosch, R. and Köhler, C.** (2009) Control of PHERES1 imprinting in Arabidopsis by direct tandem repeats. *Mol. Plant*, **2**(4), 654–660.
- Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J. and Comai, L.** (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* **19**(13), 1128–1132.
- Wang, L., Yuan, J., Ma, Y., Jiao, W., Ye, W., Yang, D.L., Yi, C. and Chen, Z.J.** (2018) Rice interploidy crosses disrupt epigenetic regulation, gene expression, and seed development. *Mol. Plant*, **11**(2), 300–314.
- Waters, A.J., Bilinski, P., Eichten, S.R., Vaughn, M.W., Ross-Ibarra, J., Gehring, M. and Springer, N.M.** (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl Acad. Sci. USA*, **110**, 19639–19644.
- Wolf, J.B., Oakey, R.J. and Feil, R.** (2014) Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity (Edinb)*, **113**, 167–175. <https://doi.org/10.1038/hdy.2014.11>.
- 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 Genet.* **7**, e1002126.
- Yoo, S.K., Lee, J.S. and Ahn, J.H.** (2006) Overexpression of AGAMOUS-LIKE 28 (AGL28) promotes flowering by upregulating expression of floral promoters within the autonomous pathway. *Biochem. Biophys. Res. Commun.* **348**, 929–936.
- Yoshida, T. and Kawabe, A.** (2013) Importance of gene duplication in the evolution of genomic imprinting revealed by molecular evolutionary analysis of the type I MADS-box gene family in Arabidopsis species. *PLoS ONE*, **8**, e73588.
- Zhang, S., Wang, D., Zhang, H., Skaggs, M.I., Lloyd, A., Ran, D., An, L., Schumaker, K.S., Drews, G.N. and Yadegari, R.** (2018) FERTILIZATION-INDEPENDENT SEED-polycomb repressive complex 2 plays a dual role in regulating type I MADS-box genes in early endosperm development. *Plant Physiol.* **177**, 285–299.

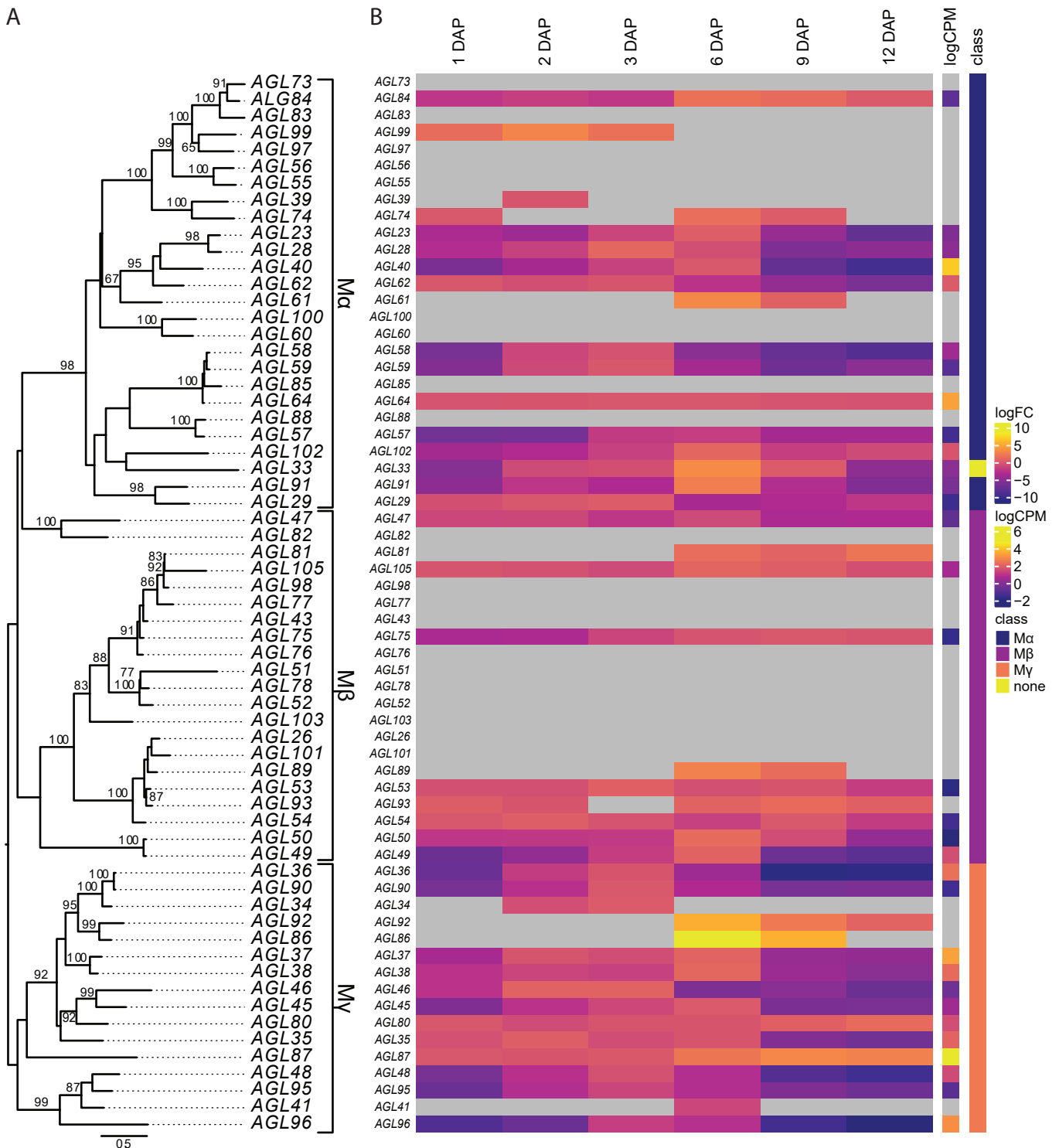


Figure S1: Phylogeny and expression of MADS-box type I transcription factors during seed development.

A) Maximum likelihood phylogeny of all 63 MADS-box type I transcription factors (TFs) in *A. thaliana*. The tree was inferred using the GTRGAMMA model in RAxML on 532 unambiguously aligned nucleotides. Scale bar represents the mean number of nucleotide substitutions per site. Only bootstrap values above 65% are shown. **B)** Gene expression profiles of MADS-box type I TFs in wild type during seed development, 1-12 days after pollination (DAP). Transcript quantification and differential expression was performed for two biological replicas with three technical replicas using RSEM and visualized using R. Gene expression profiles for stages ranging from one to 12 DAP are shown relative to four DAP using a base-2 logarithmic scale (logFC). Note that limited or no transcription of Mβ class genes was detected during seed development. Mα and Mγ MADS-box type I TFs peaked between three to six DAP. Mα=alpha class, Mβ=beta class, Mγ=gamma class.

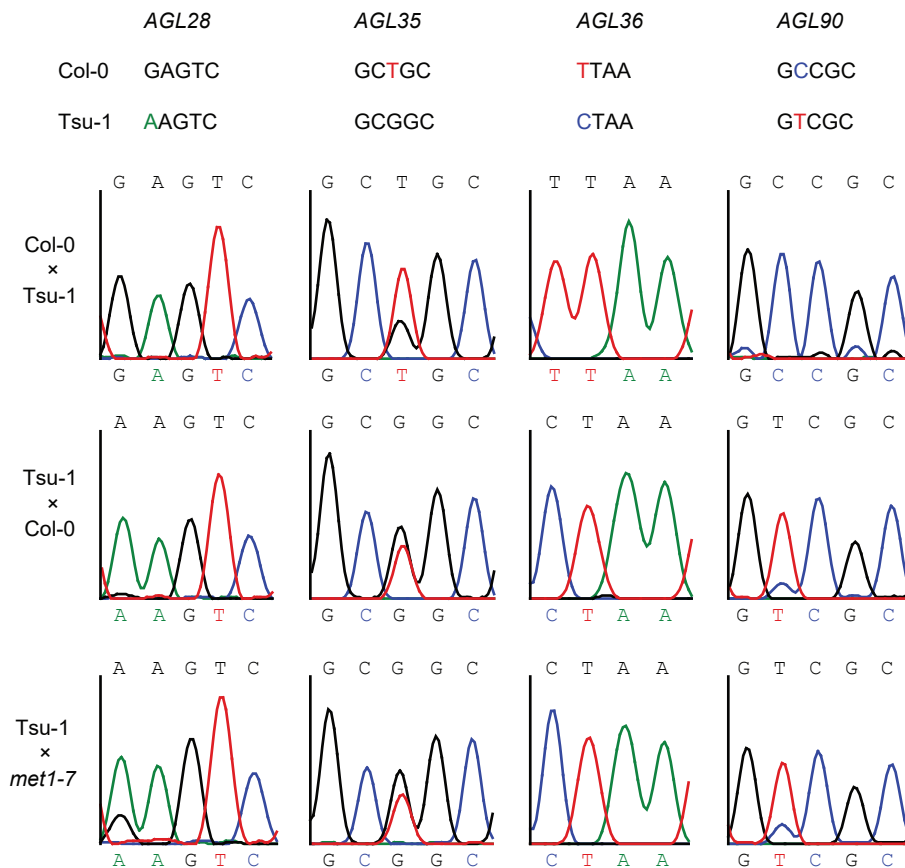


Figure S2: Imprinting analysis of *AGL28*, *AGL35*, *AGL36* and *AGL90*. Chromatograms show the sequence of the SNP containing region. PCR products from the amplification of cDNA from Col-0 x Tsu-1, Tsu-1 x Col-0 and Tsu-1 x *met1-7* (Col-0) were Sanger sequenced using the same forward primer as for PCR amplification. The base selection shown is the recognition site for the respective restriction enzymes used to detect maternal and paternal transcripts for each gene. The SNPs are indicated with same base color as used in the chromatograms. Note that the *AGL90* paternal signal in Tsu-1 x *met1-7* is indistinguishable from Tsu-1 x Col-0.

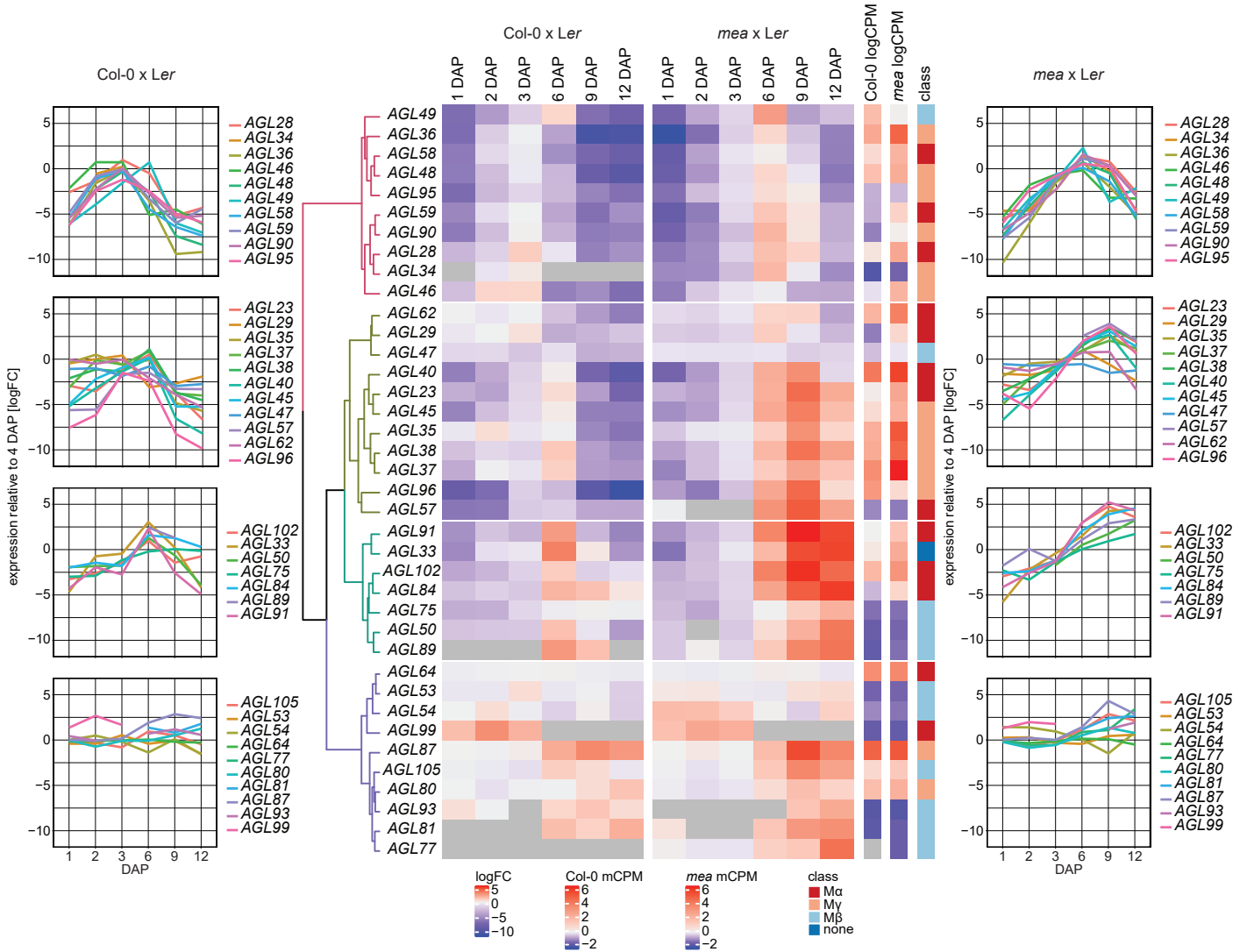


Figure S3: Clustering of MADS-box type I genes based on expression pattern in wild type and *mea* mutant seeds. Heat-map representation of the expression profile of all MADS-box type I genes detected through one to twelve days after pollination (DAP) during wild type (WT) and *mea* seed development. Expression patterns were clustered into four groups based on the mutational effect of *mea* on the MADS-box transcription factors. Left and right panels show the transcriptional profiles of the clusters in WT and *mea* mutant, respectively. Transcript quantification and differential expression analysis were performed for two biological replicas with three technical replicas using RSEM and visualized using R. Gene expression profiles for stages ranging from one to twelve DAP are relative to four DAP using a base-2 logarithmic scale (logFC). Ma =alpha class, Mb=beta class, My=gamma class.

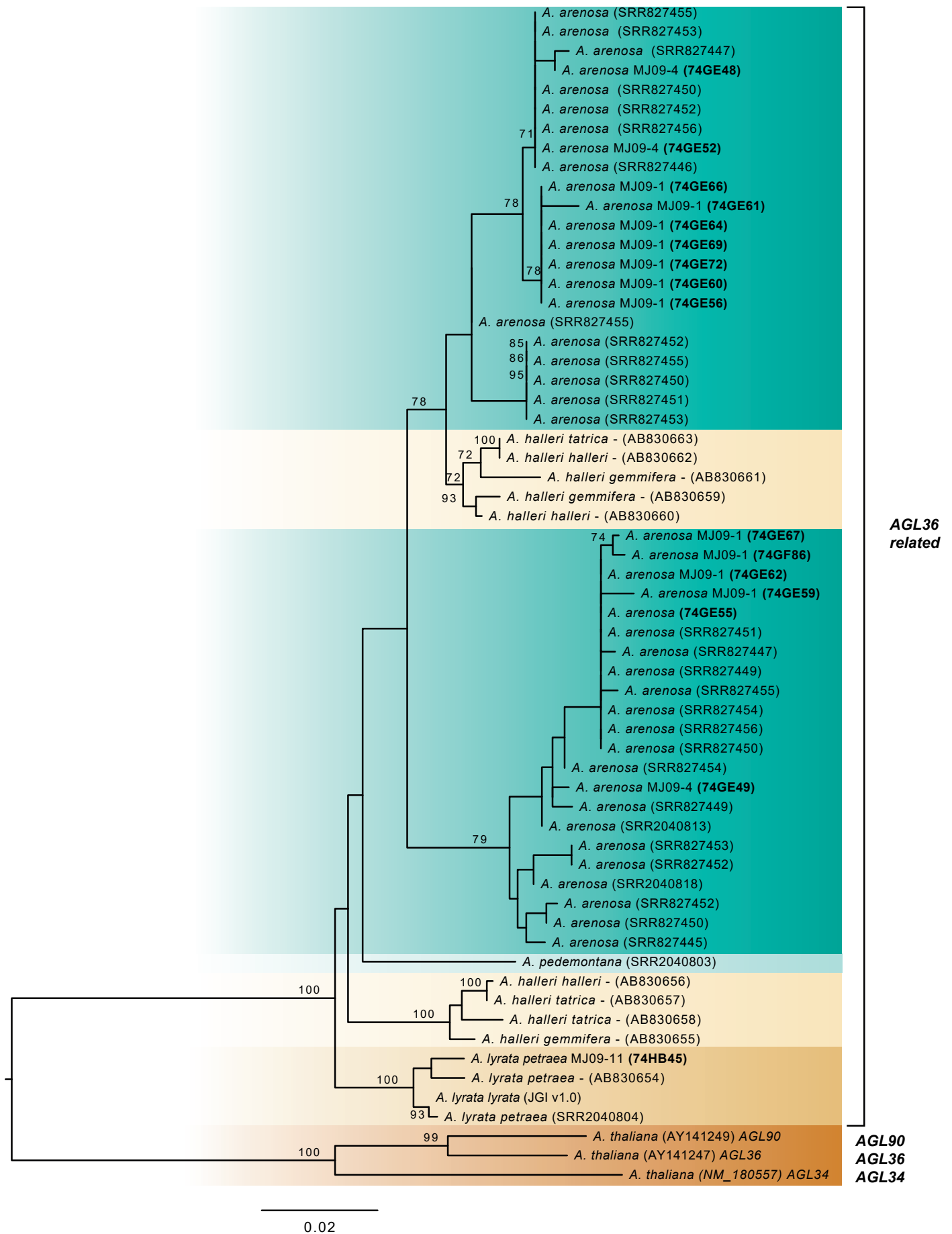


Figure S4: AGL36 related genes from all available *Arabidopsis* genomes in Sequence Read Archive, GenBank and Phytozome. The tree is inferred using the GTRGAMMA model in RAxML and the alignment consists of 978 unambiguously aligned nucleotides, and 61 sequences. Scale bar represents the mean number of nucleotide substitutions per site. Only bootstrap values above 65% are shown and *Arabidopsis* species are indicated by colors. SRR, Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>). JGI, Joint Genome Institute (<https://jgi.doe.gov/>). Sequence IDs in bold, are sequences generated from the populations used in this study.

AGL36

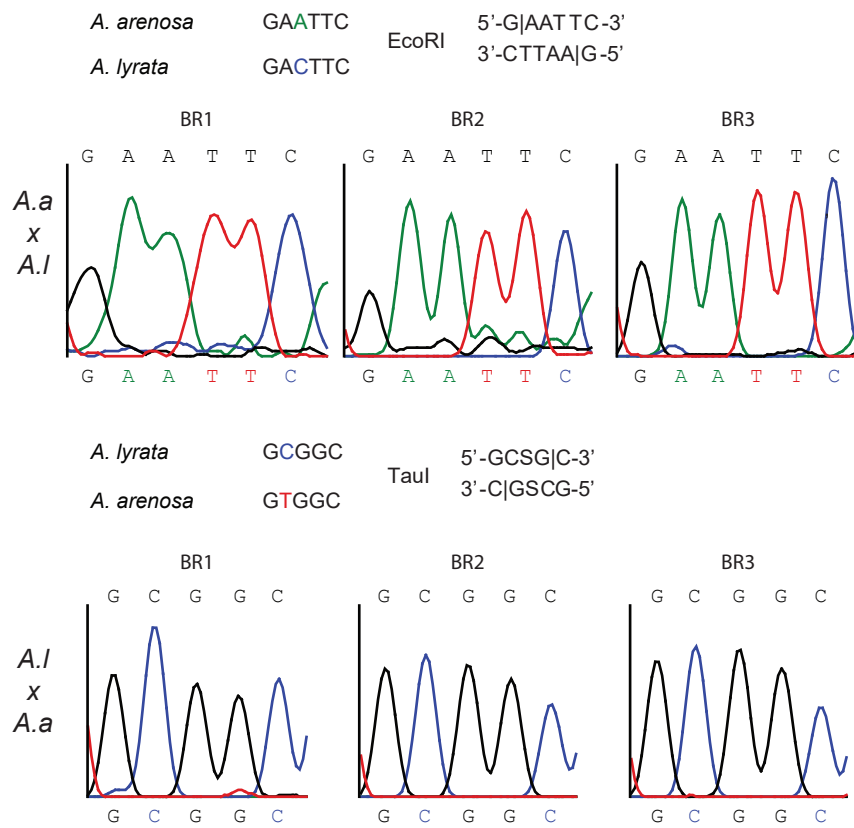


Figure S5: Imprinting analysis of *AGL36-like* in the reciprocal cross of *A. arenosa* and *A. lyrata*. Chromatograms show the sequence of the SNP containing region. PCR products from the amplification of cDNA from reciprocal crosses of *A. arenosa* and *A. lyrata* were Sanger sequenced using the same forward primer as for PCR amplification. The base selection shown is the recognition site for the respective restriction enzymes used to detect maternal and paternal transcripts for each gene. The base selections shown are EcoRI and TauI recognition sites used to detect maternal and paternal transcripts for *AGL36-like* in *A. arenosa* and *A. lyrata* reciprocal crosses.

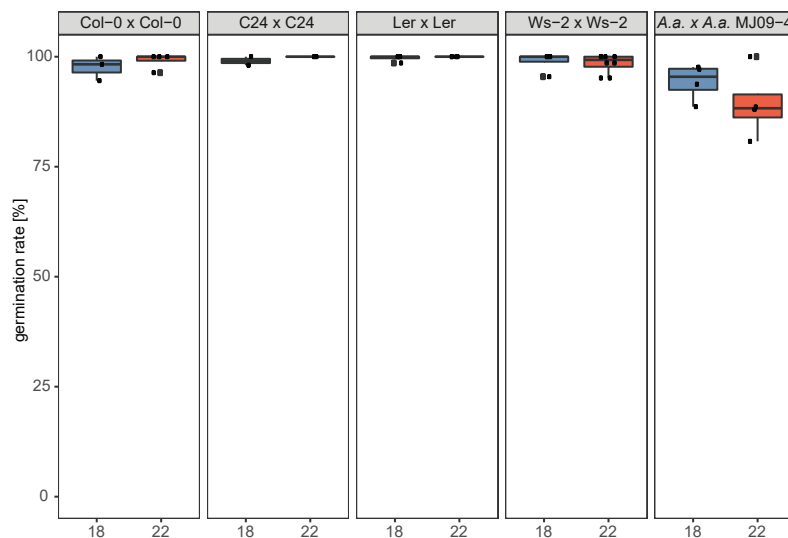


Figure S6: Germination rate in self crosses of *A. thaliana* accessions and *A. arenosa* MJ09-4 is not affected by temperature. Self crosses of Col-0, C24, Ler and Ws-2 and crosses of *A. arenosa* MJ09-4 within the same population at 18°C and 22°C show no significant species specific effect of temperature. (Wilcoxon rank-sum test: NS.: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). 18°C: Col-0 N= 3 BR (176 seeds), C24 N= 2 BR (107 seeds), Ler N= 4 BR (283 seeds), Ws-2 N= 4 BR (269 seeds), *A. a.* MJ09-4 x *A. a.* MJ09-4 N= 4 BR (169 seeds). 22°C: Col-0 N= 4 BR (237 seed), C24 N= 2 BR (92 seeds), Ler N= 4 BR (284 seeds), Ws-2 N= 8 BR (510 seeds), *A. a.* MJ09-4 x *A. a.* MJ09-4 N= 4 BR (100 seeds). Outliers are plotted as large points. BR, biological replica.

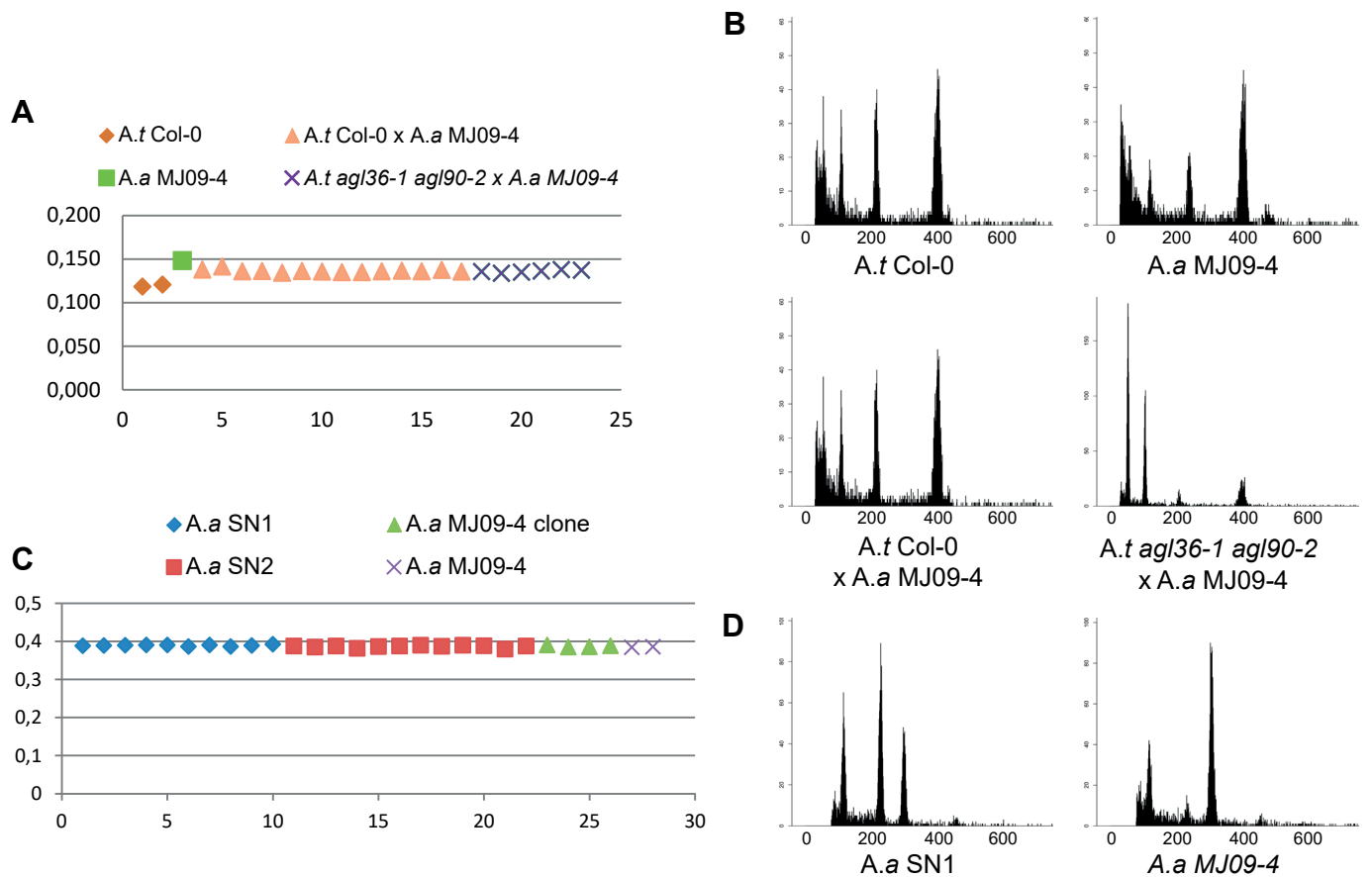


Figure S7: Ploidy measurement of *Arabidopsis* populations and hybrids. **A)** Genome content measurement for *A. thaliana* (Col-0), *A. arenosa* (MJ09-4) and hybrid crosses between Col-0 and MJ09-4 and the double mutant *agl36-1 agl90-2*. **B)** Raw cytometric outputs measured with *Solanum pseudocapsicum* as reference standard (channel 400 at the x-axis) for all *A. thaliana* and *A. thaliana* x *A. arenosa* hybrids. **C)** Genome content of *A. arenosa* Strečno lines SN1 and SN2 and *A. arenosa* MJ09-4 plants measured by flow cytometry show that all accessions are diploid. The plants termed MJ09-4 clone stem from aerial rosettes that have been cultivated on soil. **D)** Raw cytometric outputs measured with *Carex acutiformis* as reference standard (channel 300 at the x-axis) for *A. arenosa* lines SN1 and MJ09-4.

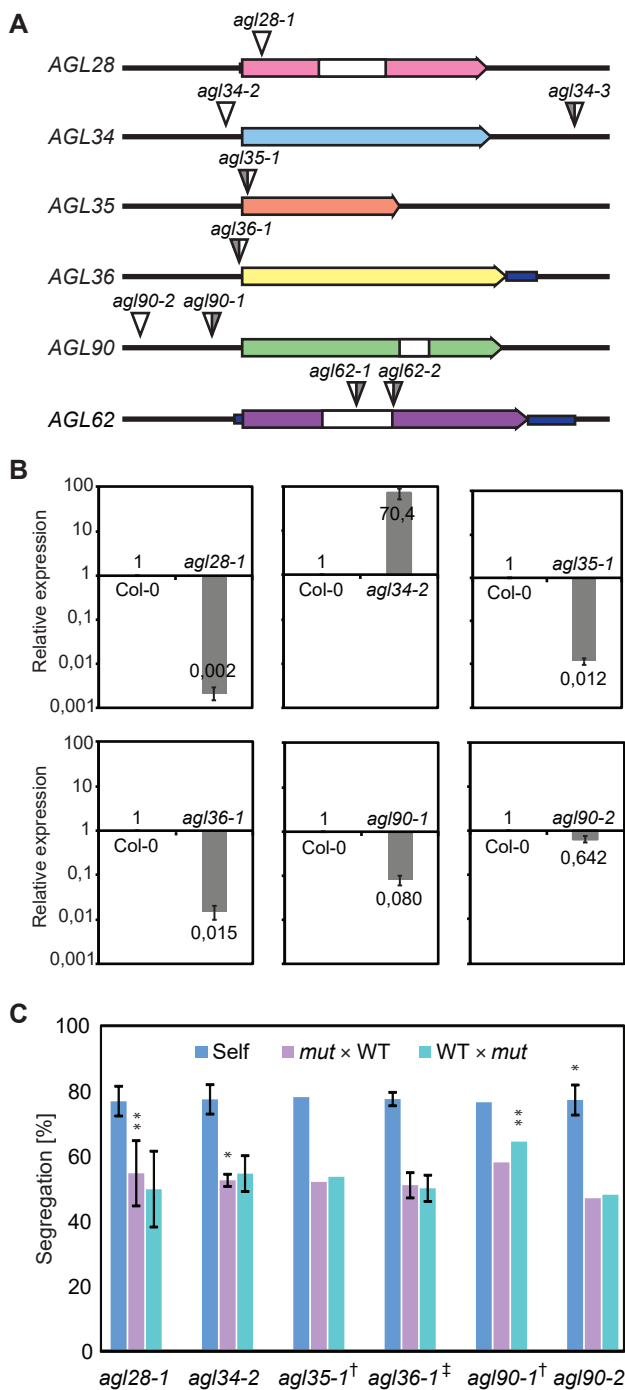


Figure S8: Genetic analysis of selected MADS-box type I transcription factors. A) T-DNA insertion overview for mutant lines of *AGL28*, *AGL34*, *AGL35*, *AGL36*, *AGL90* and *AGL62*. Gray shading of triangles indicates missing left or right T-DNA border information. Colored boxes indicate exons, white boxes indicate introns, dark blue boxes indicate untranslated regions (UTRs). **B)** Verification of transcript knock-down in *AGL28*, *AGL34*, *AGL35*, *AGL36* and *AGL90* mutants compared to wild type (Col-0) by Real-time PCR. Standard deviation between 3 biological replicates is indicated. **C)** Segregation analysis of the mutant alleles of *agl28-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1* and *agl90-2* in manual self crosses and in reciprocal crosses to wild type (Col-0). If gametophyte function or seed development is not affected by the mutation, the expected frequency of the mutant allele is 75% (self) and 50% (crosses to wild type). The observed values are not significantly different from the expected frequencies for all but four of the crosses. Asterisks denote p-values <0,05 (*) and <0,01 (**). Number of replica (N) for each cross varies from 50 to 2355 seeds (see Table S4). Dagger, verified by genotyping. Double-dagger, previously published by Shirzadi et al. (2011).

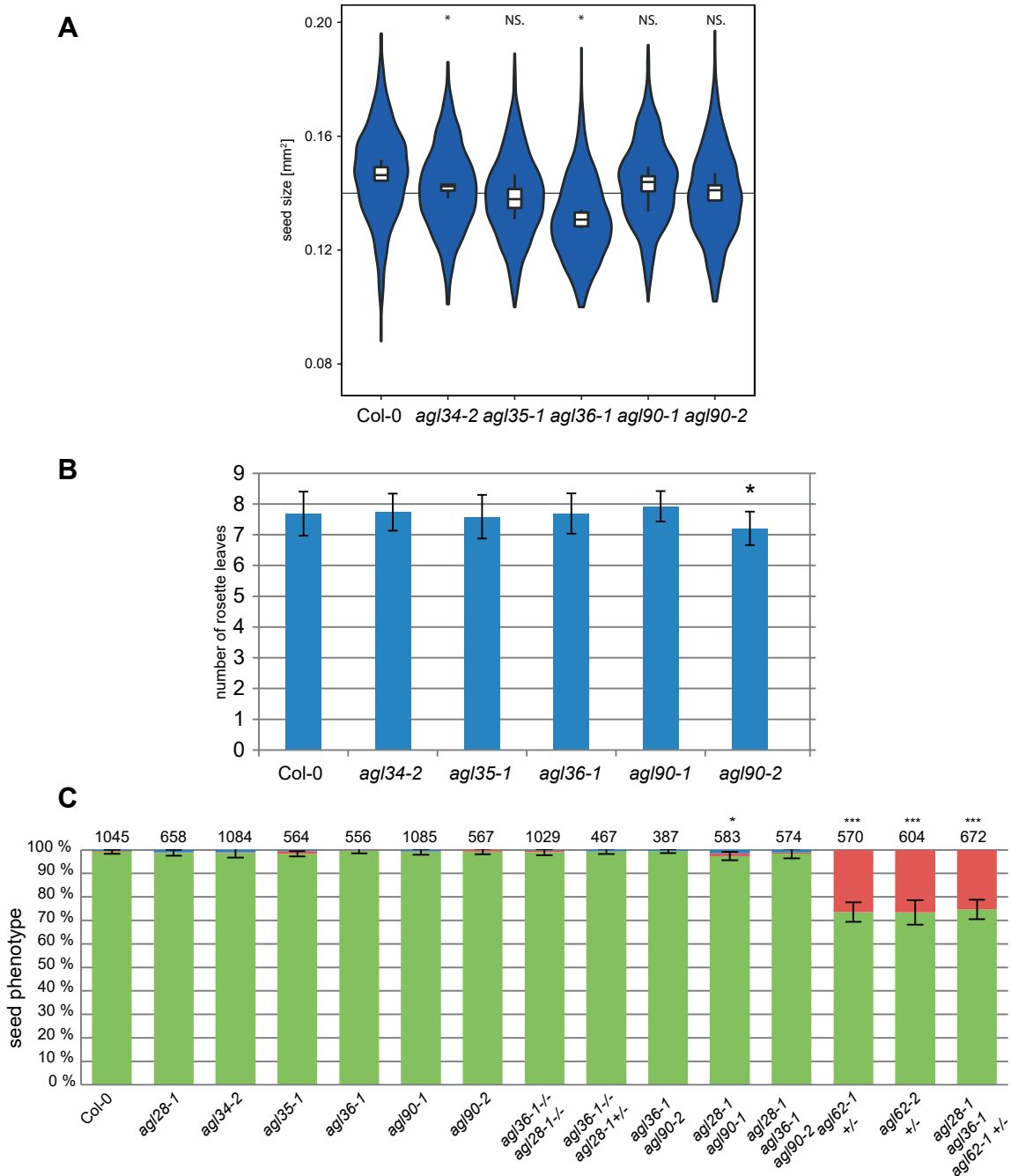


Figure S9: Phenotypic characterization of MADS-box transcription factor mutants. A) Violin plot showing seed size (area) of mutant lines *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1*, *agl90-2* and *agl28-1 agl36-1* with Col-0 as control. The boxplot represents the distribution of the mean seed size per plant (Col-0 N= 4 plants (989 seeds), *agl34-2* N=4 plants (965 seeds), *agl35-1* N= 4 plants (1545 seeds), *agl36-1* N= 4 plants (1661 seeds), *agl90-1* N= 4 plants (833 seeds), and *agl90-2* N= 4 plants (1398 seeds)(Kruskal-Wallis test: $p=0,0047$; Wilcoxon rank-sum test: * $p<0,05$ significant different from wildtype). **B)** Flowering time analysis for *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1*, and *agl90-2* with Col-0 as control. The number of rosette leaves at flowering was analyzed from 48 plants for each line. Significant differences from wild type were identified by t-test and are indicated by asterisks: * $p<0,05$; *** $p<0,001$. **C)** Seed phenotype analysis for the mutant lines *agl28-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1*, *agl90-2*, *agl28-1 agl36-1*, *agl28-1^{+/-} agl36-1* *agl36-1 agl90-2*, *agl28-1 agl90-1*, *agl28-1 agl36-1 agl90-2*, *agl62-1^{+/-}*, *agl62-2^{+/-}*, and *agl28-1 agl36-1 agl62-1^{+/-}*. All lines are homozygous unless indicated with ^{+/-}. Green color indicates percent live seeds, red indicates percent aborted seeds, and blue is percent unfertilized seeds. Number of seeds observed indicated on top of columns. Significant differences from wild type were identified by t-test and are indicated by asterisks: * $p<0,05$; *** $p<0,001$. Note that the triple mutant *agl28-1 agl36-1 agl62-1^{+/-}* display the same fraction of aborted seeds as previously reported for *agl62-1^{+/-}* (Kang et al. 2008).

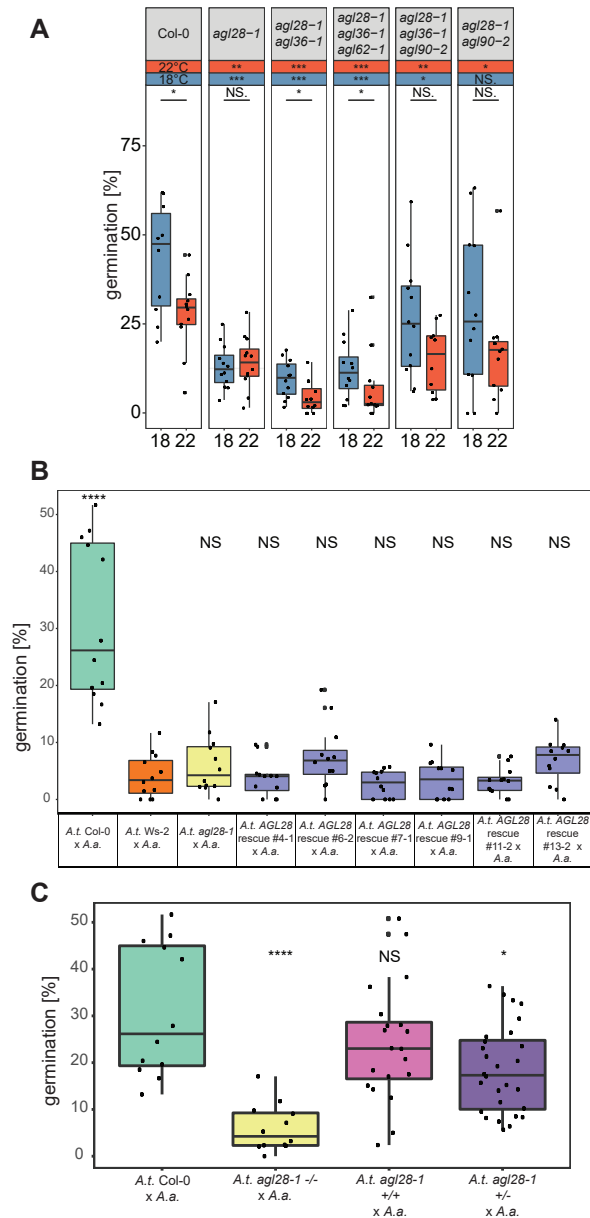


Figure S10: Characterization of a genetic background effect in mixed *A. thaliana* accessions crossed to *A. arenosa*. **A.** Single mutant *agl28-1*, double mutants *agl28-1 agl36-1* and *agl28-1 agl90-2*, and triplemutants *agl28-1 agl36-1 agl62-1* and *agl28-1 agl36-1 agl90-2* crossed to *A. arenosa* at 18°C and 22°C. The Col-0 crossed to *A. arenosa* control is identical to Figure 5. 18°C: Col-0 N= 10 BR (536 seeds), *agl28-1* N= 12 BR (664 seeds), *agl28-1 agl36-1* N= 12 BR (632 seeds), *agl28-1 agl36-1 agl62-1* N= 12 BR (594 seeds), *agl28-1 agl36-1 agl90-2* N= 12 BR (516 seeds), *agl28-1 agl90-2* N= 12 BR (579 seeds). 22°C: Col-0 N= 12 BR (578 seeds), *agl28-1* N= 12 BR (753 seeds), *agl28-1 agl36-1* N= 12 BR (572 seeds), *agl28-1 agl36-1 agl62-1* N= 12 BR (503 seeds), *agl28-1 agl36-1 agl90-2* N= 10 BR (359 seeds), *agl28-1 agl90-2* N= 12 BR (613 seeds). (Wilcoxon rank-sum test: NS.: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). **B.** Genomic complementation of the *agl28-1* mutant in Col-0/Ws-2 background do not revert to Col-0 germination levels. All lines tested resemble the Ws-2 background. All crosses were performed at 22°C. All rescue lines are in the *agl28-1* *-/-* mutant background and are from unique transformants. Col-0 N= 12 BR (637 seeds), Ws-2 N= 12 BR (727 seeds), *agl28-1* N= 12 BR (545 seeds), AGL28 rescue #4-1 N= 12 BR (600 seeds), AGL28 rescue #6-2 N= 12 BR (597 seeds), AGL28 rescue #7-1 N= 12 BR (537 seeds), AGL28 rescue #9-1 N= 12 BR (645 seeds), AGL28 rescue #11-2 N= 12 BR (715 seeds), AGL28 rescue #13-2-1 N= 12 BR (631 seeds). (Kruskal-Wallis test: $p < 9.5e-07$; Wilcoxon rank-sum test: NS.: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$). **C.** The observed reduced germination frequency in *A. thaliana* x *A. arenosa* hybrid seeds using the Ws-2 background is linked to the *agl28-1* region. Descendants of heterozygous *agl28-1* plants introgressed with Col-0 were genotyped for the presence or absence of the T-DNA causing the *agl28-1* mutation. The two classes of plants, *agl28-1* hemizygous (*agl28-1* *+/-*) or not carrying the *agl28-1* T-DNA insert (*agl28-1* *+/+*) were crossed with *A. arenosa* pollen. The plants lacking the T-DNA mutation (*agl28-1* *+/+*) are expected to have a Col-0 genetic background surrounding the AGL28 locus. Note that hybrid seeds generated from *A. arenosa* crosses with plants lacking the T-DNA mutation (*agl28-1* *+/+*) are not significantly different from *A. arenosa* crosses with the Col-0 control. Crosses of *A. arenosa* with hemizygous (*agl28-1* *+/-*) germinate at a significantly lower frequency than crosses with Col-0, however at a higher frequency than the homozygous *agl28-1* control (*agl28-1* *-/-*). All crosses performed at 22°C. Col-0 N= 12 BR (637 seeds), *agl28-1* N= 12 BR (545 seeds), *agl28-1* *+/+* N= 20 BR (1029 seeds), *agl28-1* *+/-* N= 28 BR (1406 seeds). BR, biological replicas. (Kruskal-Wallis test: $p < 8.2e-06$; Wilcoxon rank-sum test: NS.: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$).

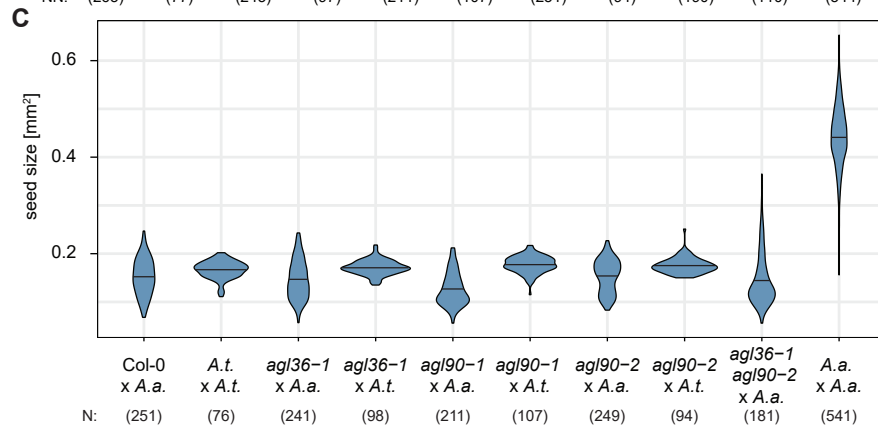
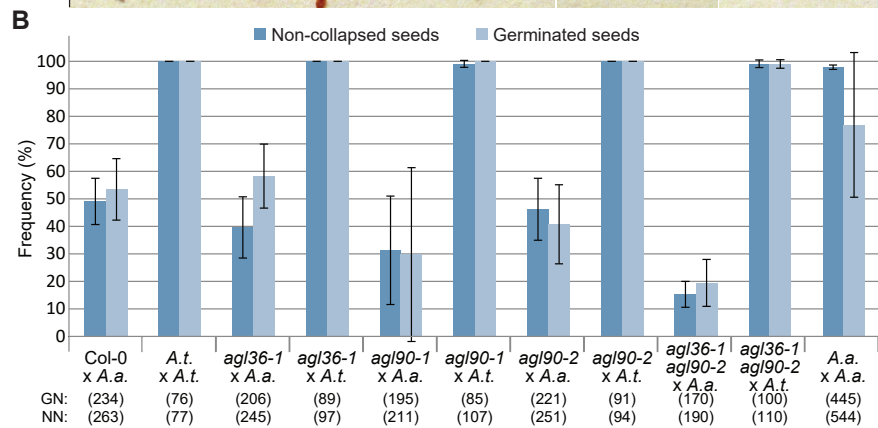
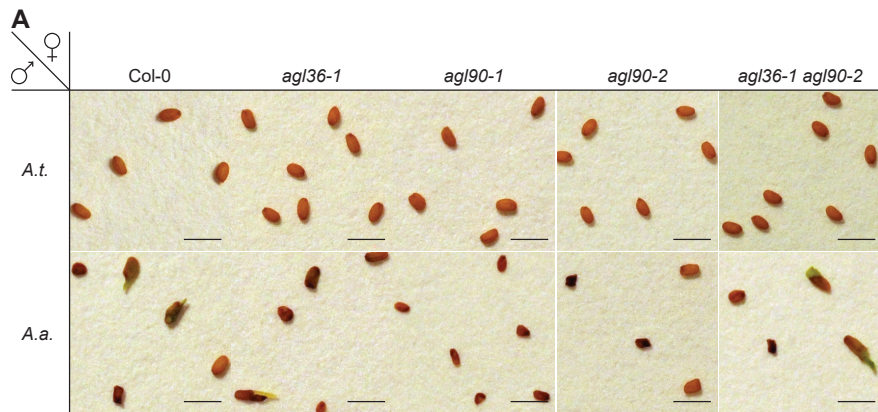


Figure S11: F1 hybrid seed phenotypes from *A. arenosa* crosses to *A. thaliana* Col-0, *agl36-1*, *agl90-1*, *agl90-2* and *agl36-1 agl90-2* at 18°C. **A) Representative selection of seeds from mutants *agl36-1*, *agl90-1*, *agl90-2* and *agl36-1 agl90-2* crossed to the *A. arenosa* MJ09-4 line at 18°C. Scale bar= 1 mm. **B)** Percent non-collapsed seeds and seed germination rate in F1 hybrid seeds of mutant lines crossed to *A. thaliana* and *A. arenosa*. GN=number of seeds tested for germination. NN=number of seeds classified as non-collapsed or collapsed seeds. **C)** Violin plot showing the distribution and mean area (mm²) of F1 hybrid seeds from mutant lines crossed to *A. thaliana* and *A. arenosa* at 18°C. Area was determined using ImageJ. N=number of seeds.**

Table S1: Schematic overview of restriction digest set-up.

A)

Gene	Ecotype cross	Restriction enzyme	Undigested	Digested cDNA	Primers	T _M °C
AGL28	Col-0 x Tsu-1	HinfI cuts in Col0, not in Tsu	143 bp	45 bp + 98 bp	AGL28_SNP306_Tsu_LP	58,6
					AGL28_SNP306_Tsu_RP	58,2
AGL35	Col-0 x Tsu-1	TauI cuts in Tsu-1 and not Col-0.	451 bp	83 bp + 368 bp.	AGL35-ASP-SNP	56,7
					AGL35-SP-SNP	56,3
AGL36	Col-0 x Tsu-1	MseI cuts in Col0, not in Tsu-1	191 bp	49 bp + 142 bp	AGL36_SNP590_Tsu_LP	57,1
					AGL36_SNP590_Tsu_RP	57,1
AGL90	Col-0 x Tsu-1	TauI cuts in Col-0, not in Tsu-1	208 bp	166 bp + 42 bp	AGL90_SALK_RP	51,2
					AGL90-SP3 subcloning	62,1

B)

Gene	Cross	Restriction enzyme	Undigested	Digested cDNA	Primers	T _M °C
AGL36	<i>nprp1</i> x Col-0	AlwNI (Cail) cuts	399 bp	165 bp + 234 bp	AGL36_SP7_SNP	53,6
					AGL36 ASP6 SNP	53,1

C)

Gene	Cross	Restriction enzyme	Undigested	Digested cDNA	Primers	T _M °C
AaAGL36	<i>A. arenosa</i> x <i>A. arenosa</i>	BceAI cuts AaAGL36 without SNP	487 bp	152 bp + 336 bp	AaAGL36 SP SNP	57,7
					AaAGL36 ASP SNP	56,1
AtAGL36	<i>A. thaliana</i> x <i>A. arenosa</i>	AlwNI (Cail) cuts in AtAGL36, not AaAGL36	AaAGL36: 768 bp, AtAGL36: 819 bp	611 bp + 208 bp	AGL36-146-LP	54,2
					AGL36-AS1 cloning primer	57,0
AIAGL36	<i>A. lyrata</i> x <i>A. arenosa</i>	TauI cuts AIAGL36, not AaAGL36	768 bp for both AIAGL36 and AaAGL36	495 bp + 276 bp	AGL36-146-LP	54,2
					AGL36-AS1 cloning primer	57,0
AaAGL36	<i>A. arenosa</i> x <i>A. lyrata</i>	EcoRI cuts AaAGL36, not AIAGL36	768 bp for both AIAGL36 and AaAGL36	560 bp + 208 bp	AGL36-146-LP	54,2
					AGL36-AS1 cloning primer	57,0

A-C) Overview of restriction digest set-up used in verification of imprinting of *AGL28*, *AGL35*, *AGL36*, and *AGL90* in seeds from reciprocal crosses of Col-0 and Tsu-1 (**A**), *AGL36* imprinting in seeds from *nprp1* crossed to Col-0 (**B**), *AGL36-related* genes in F1 seeds from *A. thaliana* x *A. arenosa* (**C**), *AGL36-related* genes in F1 seeds from *A. arenosa* and *A. arenosa* reciprocally crossed to *A. lyrata* (**C**). Note that amplification of *AGL36-related* genes from F1 seeds from *A. arenosa* x *A. lyrata* results in a fragment of same length and the restriction digest distinguishes the origin of the fragment. T_M=Melting Temperature using Nearest Neighbor.

Table S2 Oligonucleotide name, sequence and description.

Name/ experiment	Oligo Sequence	Comments
Real-time primers and Probes:		
AGL28-22-LP	TGCTCAATGAGTCTTTAACTGAGG	AGL28
AGL28-22-RP	AGCTAAGTTGAGTTCTGTTGGAGAG	AGL28
AGL34-Forward-Bemer2010	AGGAAGATGATGAATCAAGAAACGT	AGL34
AGL34-Reverse-Bemer2010	AGTCACCGTGTTCTTTAAGGAT	AGL34
AGL35 qPCR ASP2	GAAGAACCCATCTCAACCACAGGC	AGL35
AGL35 qPCR SP2	ATGAACCATCTTCCGCCCTCG	AGL35
AGL36-160-LP	AGGTGGCTTCAAGGTTTCTG	AGL36
AGL36-160-RP	GATCCATCATCTTCTTGGTTCTG	AGL36
AGL90-qpcr-left	TGGTGATGAGTCGTTTTCCGA	AGL90
AGL90-qpcr-right	ACCTGATTCATATTCGCATTTCG	AGL90
ACT11 77 RP	TGTCTTCACCATCTGCCATT	ACTIN11
ACT11 77 LP	CAAAAACACACACCCGTACCA	ACTIN11
Probes from Universal Probe Library (Roche):		
#160		AGL36 real-time PCR
#22		AGL28 real-time PCR
#77		ACT11 real-time PCR
Genotyping primers:		
Hook1	CTACACTGAATTGGTAGCTCAAACCTGTC	Left border T-DNA primer for KONCZ lines
AGL36 AS2 KONCZ	GGATGGTAGTTGATGGTAGTTGTAGCAGTT	Antisense primer for agl36-1 genotyping
AGL34 AS3 KONCZ	GGCACTGATTAATCGATATAGGCCAAAG	agl34-1 T-DNA with Hook1
AGL34 S1 KONCZ	ATGAAATTACTAAGAGTAATTTCTAGCTTACCACAA	Sense primer for genomic sequence of AGL34
AGL34-SP2-SNP	TATATGAATCAACAACAACC	T-DNA for agl34-3
Barb_193_FBL17_T-DNA	ATATTGACCATCATACTCATTGC	T-DNA for agl34-3
AGL34-ASP1-SNP	GTTGAAACGGTTCGATGCAGA	T-DNA agl34-2 leftborder with LB3
pAGL36-Forw1	AGAACAAAGTAAGCTAGAAATT	T-DNA agl34-2 rightborder with LB3
AGL34 AS4 KONCZ	GAAAATGTTATTAATAAAAAATCGACCTAGGCACTGATTAATCGAT	Antisense primer for genomic sequence of AGL34
agl35_Salk_033801_RP	AAACCAAAGTTTTGCCACTAAGAC	Antisense primer for agl35-1 genotyping with J504
agl35_Salk_033801_LP	ATTTTTCAGTCAAGATTACCCACC	Sense primer for agl35-1 genotyping
J504	GCGTGGACCGCTTGCTGCAACTCTCTCAGG	Left border T-DNA primer for Salk lines
LBa1	GTTCACGTAGTGGGCCATC	Left border T-DNA primer for Salk lines
LB3	TAGCATCTGAATTTTATAACCAATCTCGATACAC	T-DNA agl34-2 right- and leftborder primer.
PC23_SAIL_1242_G12_RP	TTTCTTTAATGGACTCGATCCTTC	agl90-1 and agl90-2 genotypings with LBa1.
PC21_SALK_008897_RP	AATCGTCTAACCTGTAATTCTCGG	agl90-1 genotyping
PC21_SALK_008897_LP	TAAGAAACCTTTTTGGCAGAAAAG	agl90-1 genotyping
PC23_SAIL_1242_G12_LP	TTTGCAATTTTCAAATTGTTTGC	agl90-2 genotyping with LBa1
AGL28-SP1	TGGCGAGAAAGAATCTTGGTCGTAGAA	T-DNA left border in agl28-1
pDSLOX_1F	TGGAAGGCACGCAACGCCTACGACTGGACG	T-DNA primer for agl28-1 leftborder
AGL28ASP1	TTGAGCATTTGAATACGGAGCTTTGTGC	T-DNA alg28-1 rightborder
RB FST T-DNA	GGGTTGGGGTTTCTACAGGACG	T-DNA primer for agl28-1 rightborder
AGL28 rc/1-482 SP	ATGGCGAGAAAGAATCTTGG	Genomic sequence of AGL28
AGL28 rc/1-482 asp	CCCACCAAATTTTCATTCT	Genomic sequence of AGL28
A.thaliana SNP primers:		
AGL36-SP7-SNP	GAACCGTTTCAACACCTTGTTCC	AGL36 leftborder <i>nrdp1x</i> Col-0 SNP PCR primer
AGL36-ASP6-SNP	CTAGTTCTTCTCCCTCTTTTGTG	AGL36 rightborder <i>nrdp1x</i> Col-0 SNP PCR primer

AGL28_SNP306_Tsu_LP	ACTTTGCTGAAAGCCGCACAAA	AGL28 leftborder SNP PCR primer
AGL28_SNP306_Tsu_RP	CCACTTCTCAGCGTCCTTGTTC	AGL28 rightborder SNP PCR primer
AGL36_SNP590_Tsu_LP	CATTTTCCGACTCTCCTATTCATGC	AGL36 leftborder Tsu-1 xCol-0 SNP PCR primer
AGL36_SNP590_Tsu_RP	CGGTTTCATGCAGATTTTGACTCATA	AGL36 rightborder Tsu-1 xCol-0 SNP PCR primer
AGL34-SP-SNP8	GAAACTGGTGATGAGTCGTCTTCC	AGL34 Tsu-1 xCol-0 SNP PCR primer
AGL34-ASP-SNP8	CATGAAAGGAATGCTCTCACGC	AGL34 Tsu-1 xCol-0 SNP PCR primer
AGL35-ASP-SNP	GAATTCTTTGTGGCGTCCCG	AGL35 Tsu-1 xCol-0 SNP PCR primer
AGL35-SP-SNP	GAAGAACCCATCTCAACCACAGG	AGL35 Tsu-1 xCol-0 SNP PCR primer
AGL90_SALK_RP	TGGGTTTCTTGATCCATCATC	AGL90 rightborderTsu-1 xCol-0 SNP PCR primer
AGL90-SP3 subcloning	CCTTCATGAAGAGGAAAAACGGGATATTC	AGL90 leftborderTsu-1 xCol-0 SNP PCR primer
A.arenosa and A.lyrata amplification:		
AGL36-SP	ATGAAGAAGGTGAAGCTATCTTT	Amplifying AaAGL36 and AIAGL36
AGL36-146-LP	AAGCAAAAGAGCAACTAAAGAACC	Amplifying AaAGL36 and AIAGL36
AGL36-AS1-Cloning primer	AGAGATTATTGTTGATGTAAGGATCATAGACA	Amplifying AaAGL36 and AIAGL36
AaAGL36 SP SNP	CTCATTCATGAAGAGGAAGAACGGG	Aa AGL36 SNP analysis
AaAGL36 ASP SNP	TGCAACATCAGGAATATCTACAACCC	AaAGL36 SNP analysis
GateWay cloning primers:		
AGL28 rescue SP2	TATGGAAAGTGGAAGATAGT	Amplifying genomic AGL28
AGL28 rescue ASP2	AACTAAACGATAGAGAGAAGTC	Amplifying genomic AGL28
attB1 AGL28 rescue SP	GGGGACAAGTTTGTACAAAAAAGCAGGCTTATATGGAAAGTG GAAGATA	Amplifying genomic AGL28 with att-site
attB2 AGL28 rescue ASP	GGGGACCACTTTGTACAAGAAAGCTGGGTAAACTAAACGATA GAGAGA	Amplifying genomic AGL28 with att-site

Table S3: Characterization of mutants used in this study.

Gene name	AGI no.	Mutant line	T-DNA localization	Primers
AGL23	AT1G65360	<i>agl23-1</i> (SALK_147048) ¹	The right border of the T-DNA is 84 bp downstream of the start codon (ATG). 565 bp has been deleted at the left border of the T-DNA.	T-DNA: -Left border: RB FST T-DNA and AGL28-ASP1. -Right border: pDSLOX_1F and AGL28-SP1. Genomic: - AGL28 rc/1-482 SP and AGL28 rc/1-482 ASP T-DNA: -Left border: AGL34_ASP1_SNP and LB3 -Right border: pAGL36_forw1 and LB3. Genomic: -AGL34_S1_KONCZ and AGL34 AS3 KONCZ. T-DNA: -AGL34-SP2-SNP and Barb_193-FBL17_tDNA. Genomic: -AGL34 SP SNP and AGL34 AS3 KONCZ. T-DNA: -J504 and <i>agl35_Salk_033801_RP</i> . Genomic: - <i>agl35_Salk_033801_LP</i> and <i>agl35_Salk_033801_RP</i> .
AGL28	AT1G01530	<i>agl28-1</i> (FLAG_386G1) ²		
AGL34	AT5G26580	<i>agl34-2</i> (SAIL_1242_G12)	AGL34 right border: 71 bp upstream of the start codon (ATG). AGL34 left border: 85 bp upstream of start codon (ATG). Two T-DNA inserts and therefore two T-DNA left borders.	T-DNA: -PC23_SAIL_1242_G12_LP with the T-DNA primer LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP. T-DNA: -Left border: PC23_SAIL_1242_G12_RP and LBA1. -Right border: PC21_SALK_008897_RP and LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP.
AGL35	AT5G26630	<i>agl34-3</i> (WiscDsLoxHs012_11C)	T-DNA is downstream of the AGL34 stop codon.	
AGL36	AT5G26650	<i>agl36-1</i> (Koncz line) ³	18 bp downstream of the start codon (ATG).	T-DNA: -PC23_SAIL_1242_G12_LP with the T-DNA primer LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP. T-DNA: -Left border: PC23_SAIL_1242_G12_RP and LBA1. -Right border: PC21_SALK_008897_RP and LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP.
AGL90	AT5G27960	<i>agl90-1</i> (SALK_008897)		
AGL62	AT5G60440	<i>agl62-1</i> (SALK_137707) ⁴ <i>agl62-2</i> (SALK_022148) ⁴	T-DNA left border 421 bp upstream of the start codon (ATG). Two T-DNA inserts and therefore two T-DNA left borders.	T-DNA: -PC23_SAIL_1242_G12_LP with the T-DNA primer LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP. T-DNA: -Left border: PC23_SAIL_1242_G12_RP and LBA1. -Right border: PC21_SALK_008897_RP and LBA1. Genomic: -PC21_SALK_008897 RP and PC21_SALK_008897 LP.

All T-DNA insertional mutants were thoroughly verified and characterized using PCR and sequencing. Primer sequences used to amplify T-DNA borders are listed in Table S2.¹ characterized by Colombo et al. (2008),² characterized by Yoo et al. (2006),³ characterized by Shirzadi et al. (2011),⁴ characterized by Kang et al. (2008). All mutant lines are in the Col-0 accession except *agl28-1* which was back crossed 6 times to Col-0 from the original Wassilewskija (Ws-2) background.

Table S4: Segregation of the mutant alleles *agl28-1*, *agl34-2*, *agl35-1*, *agl36-1*, *agl90-1* and *agl90-2* in self crosses and in reciprocal crosses to wild type (Col-0).

	Mutant line	SM	N	% segregation	SD	H₀	χ²
<i>agl28-1</i> × Self	FLAG_386G1	BASTA	543	76,80	4,55	3:1	0,93
<i>agl28-1</i> ^{+/-} × Wild type			1041	54,66	10,05	1:1	9,04**
Wild type × <i>agl28-1</i> ^{+/-}			876	49,77	11,64	1:1	0,02
<i>agl34-2</i> × Self	SAIL_1242_G12	Km	1069	77,36	4,48	3:1	3,18
<i>agl34-2</i> ^{+/-} × Wild type			2029	52,49	1,88	1:1	5,03*
Wild type × <i>agl34-2</i> ^{+/-}			385	54,55	5,55	1:1	3,18
<i>agl35-1</i> × Self	SALK_033801	Genotyping	50	78,00	NA	3:1	0,24
<i>agl35-1</i> ^{+/-} × Wild type			100	52,00	NA	1:1	0,16
Wild type × <i>agl35-1</i> ^{+/-}			99	53,54	NA	1:1	0,49
<i>agl36-1</i> × Self	Koncz line	Hyg	1025	77,46	2,08	3:1	3,32
<i>agl36-1</i> ^{+/-} × Wild type			661	50,98	3,88	1:1	0,26
Wild type × <i>agl36-1</i> ^{+/-}			1015	50,05	4,02	1:1	0,00
<i>agl90-1</i> × Self	SALK_008897	Genotyping	51	76,47	NA	3:1	0,06
<i>agl90-1</i> ^{+/-} × Wild type			100	58,00	NA	1:1	2,56
Wild type × <i>agl90-1</i> ^{+/-}			118	64,41	NA	1:1	9,80**
<i>agl90-2</i> × Self	SALK_092748	Km	2355	77,15	4,59	3:1	5,83*
<i>agl90-2</i> ^{+/-} × Wild type			213	46,95	NA	1:1	0,79
Wild type × <i>agl90-2</i> ^{+/-}			210	48,10	NA	1:1	0,30

Selfed mutant crosses are expected to give a segregation frequency of 75 % (3:1) of the mutant allele. Segregation of 50 % (1:1) is expected of seeds from heterozygous mutant plants crossed to wild type (WT). Using a chi square test (df=1), the observed values were tested against the expected 3:1 and 1:1 segregation ratios (H₀). The null hypothesis is rejected at p<0,05 (X²>3,84). SM: selection medium, N: number of seeds in segregation test.

Table S5: Crossing scheme for various *A. thaliana* to *A. arenosa* experiments.

A) Seed development comparison of *A. thaliana* crossed to *A. arenosa* at 18°C and 22°C.

<i>A.thaliana</i> line (N of plants)	<i>A.arenosa</i> line (N plants)	Siliques per experiment:	Figure
Col-0 (6)	MJ09-4 (3)	-1 silique per plant for imaging and phenotyping 20 days after pollination. -Four siliques per plant for germination.	5 A, 5 B
Col-0 (6)	SN1 (3)	-1 silique per plant for imaging and phenotyping 20 days after pollination. -4 siliques per plant for germination.	5 A, 5 B and 5 C

B) Effect of natural variation on hybridization using multiple populations of *A. arenosa* and accessions of *A. thaliana* at 18°C and 22°C.

<i>A.thaliana</i> line (N of plants)	<i>A.arenosa</i> line and N plants	Sub-lines	Temperature	Siliques per experiment:	Figure
Col-0 (3)	MJ09-4 (3)	2	18	3	5 C
Col-0 (3)	MJ09-4 (3)	1	22	4	5 C
Col-0 (3)	MJ09-1 (3)	2	18	3	5 C
Col-0 (3)	MJ09-1 (3)	1	22	4	5 C
Col-0 (3)	SN1 (3)	1	18	3	5 C
Col-0 (3)	SN1 (3)	1	22*	4	5 C
Col-0 (3)	SN2 (3)	1	18	3	5 C
Col-0 (3)	SN2 (3)	1	22	4	5 C
Col-0 (3)	MJ09-4 (3)	1	18	4	5 D
Col-0 (3)	MJ09-4 (3)	1	22	4	5 D
C24 (3)	MJ09-4 (3)	1	18	4	5 D
C24 (3)	MJ09-4 (3)	1	22	4	5 D
Ler-1 (3)	MJ09-4 (3)	1	18	4	5 D
Ler-1 (3)	MJ09-4 (3)	1	22	4	5 D
Ws-2 (3)	MJ09-4 (3)	1	18	4	5 D
Ws-2 (3)	MJ09-4 (3)	1	22	4	5 D, S10 B

* repeated twice

C) Effects of *A. thaliana* mutants of the MADS-box type I transcription factors crossed to *A. arenosa* at 18°C and 22°C on seed development by germination test.

<i>A.thaliana</i> line (N of plants) (mother)	<i>A.arenosa</i> line (N of plants) (father)	Experimental temperature (N of plants):	Siliques pr plant:	Figure
Col-0 (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7, S10 A, B, C
<i>agl23-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl28-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	S10 A, B, C
<i>agl34-2</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl35-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl36-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl62-1</i> (6)	MJ09-4 (3)	18°C (2) 22°C (2)	4	7
<i>agl90-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl90-2</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl28-1 agl36-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	S10 A
<i>agl28-1 agl90-2</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	S10 A
<i>agl36-1 agl62-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	7
<i>agl28-1 agl36-1 agl90-2</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	S10 A

<i>agl28-1 agl36-1 agl62-1</i> (6)	MJ09-4 (3)	18°C (3) 22°C (3)	4	S10 A
AGL28 rescue in <i>agl28-1</i> #4-1 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
AGL28 rescue in <i>agl28-1</i> #6-2 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
AGL28 rescue in <i>agl28-1</i> #7-1 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
AGL28 rescue in <i>agl28-1</i> #9-1 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
AGL28 rescue in <i>agl28-1</i> #11-2 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
AGL28 rescue in <i>agl28-1</i> #13-2 (3)	MJ09-4 (3)	22°C (3)	4	S10 B
<i>agl28-1 +/-</i> x Col-0 (12)	MJ09-4 (1)	22°C (12)	4	S10 C

D) Seed imaging, seed size measurements and phenotype/ germination correlation of *A.thaliana* wildtype and mutants crossed to *A.arenosa*.

<i>A.thaliana</i> line and N plants (mother)	Father and N plants	Experiment:	Siliques per plant:	Figure
<i>agl90-1</i> (3)	<i>A. arenosa</i> MJ09-4 (3)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl90-1</i> (1)	<i>A. thaliana</i> Col-0 (1)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl90-2</i> (3)	<i>A. arenosa</i> MJ09-4 (3)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl90-2</i> (1)	<i>A. thaliana</i> Col-0 (1)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl36-1</i> (3)	<i>A. arenosa</i> MJ09-4 (3)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl36-1</i> (1)	<i>A. thaliana</i> Col-0 (1)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl36-1 agl90-2</i> (3)	<i>A. arenosa</i> MJ09-4 (3)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C
<i>agl36-1 agl90-2</i> (1)	<i>A. thaliana</i> Col-0 (1)	-Seed imaging and seed size measurement with ImageJ. -Germination assay	2	S11 A, B, C

A) Plant lines used for seed development comparison of *A. thaliana* crossed to *A. arenosa* at 18°C and 22°C. **B)** Plant lines used to examine the effect of natural variation on hybridization at 18°C and 22°C. **C)** Plant lines used to examine the effect of *A. thaliana* MADS-box type I mutants crossed to *A. arenosa* at 18°C and 22°C on seed germination. 84 MADS box mutant plants were crossed to *A.arenosa* (MJ09-4) and half was grown at 18°C, while the other half was grown at 22°C. 18 plants from 6 unique AGL28 complementation lines in the *agl28-1* background were crossed to *A. arenosa* (MJ09-4) at 22°C. Twelve plants from a cross between *agl28-1 +/-* and Col-0 were crossed to one *A. arenosa* (MJ09-4) plant. Four siliques were harvested and sown out individually on MS-2 plates before being scored for germination. **D)** Plant lines used for seed imaging, seed size measurements and correlation germination of *A.thaliana* wildtype and mutants crossed to *A.arenosa*. Two siliques from each cross were imaged and analyzed using ImageJ to calculate average seed size. The seeds were sterilized and transferred to MS plates and stratified for four days before being placed under long day conditions. Germination was scored for each silique and hybrid seedlings transferred to soil for further investigation.

