Signalling mechanisms and epigenetic regulation of seed development in *Arabidopsis thaliana*

Rita Adriano Batista
Faculty of Natural Resources and Agricultural Sciences
Department of Plant Biology
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2019

Acta Universitatis agriculturae Sueciae
2019:39

Cover: *Arabidopsis thaliana* ovules expressing the PcG protein MSI1 tagged with GFP (green). Cell walls and gametophyte cells are stained with propidium iodide (magenta).
(photo: Rita Adriano Batista)

ISSN 1652-6880
ISBN (print version) 978-91-7760-396-2
ISBN (electronic version) 978-91-7760-397-9

Print: SLU Service/Repro, Uppsala 2019

Signalling mechanisms and epigenetic regulation of seed development in *Arabidopsis thaliana*

Abstract

In flowering plants, seed development starts with a double fertilisation event, leading to the formation of the embryo and the endosperm. Though only the embryo contributes to the next generation, control of endosperm proliferation, as well as the timing of its cellularisation are crucial for embryo viability. Both the embryo and the endosperm are enclosed by the seed coat, a maternal tissue that serves as a protective layer and is important for various aspects of seed development. Despite not arising directly from a fertilization event, the seed coat responds to fertilization by expanding and differentiating, accompanying the growth of the embryo and the endosperm. Therefore, successful seed development requires a set of signalling networks that coordinate growth between the embryo, the endosperm, and the seed coat. In this thesis, I show that auxin is a key player in these signalling networks. Its biosynthesis occurs in the endosperm, depends on the activity of imprinted paternally expressed genes, and serves two purposes: i) it triggers and sustains endosperm proliferation, and ii) it is exported to the integuments, which kick-starts seed coat development. In later stages of seed development, a reduction of auxin activity in the endosperm is required to initiate endosperm cellularisation, and the seed abortion phenotypes observed in interploidy crosses are connected with atypical levels of this hormone. The present work also identified two endosperm-expressed MADS-box transcription factors (TFs) as the upstream transcriptional regulators of auxin biosynthesis and transport genes. Interestingly, one of these TFs – the imprinted paternally expressed *PHERES1* – controls the expression of several other imprinted genes, and accesses its targets through specific DNA-binding sites, which are significantly associated with transposable elements (TEs). TEs were previously implicated in providing epigenetic landscapes conducive to imprinted gene expression. I further propose that these elements have been co-opted as *cis*-regulatory sequences that facilitate *PHERES1* binding to promoters of imprinted genes. Overall, this thesis makes important contributions to further increase our understanding of seed development: not only it uncovers the hormone auxin as a key factor that coordinates the development of different seed components, but also clarifies the previously elusive role of MADS-box genes in the transcriptional networks controlling endosperm development.

Keywords: seed, endosperm, seed coat, auxin, imprinting, MADS-box, *PHERES1*

Author's address: Rita Adriano Batista, SLU, Department of Plant Biology, Linnean Centre for Plant Biology, P.O. Box 7800, SE-75007, Uppsala, Sweden

Signalmekanismer och epigenetisk reglering av fröutveckling i *Arabidopsis thaliana*

Vetenskaplig sammanfattning

I blommande växter börjar fröutveckling med en dubbel fertilisering som leder till att embryo och endosperm bildas. Även om endast embryot bidrar till nästa generation, är delning av cellkärnor i blivande endosperm, liksom tidpunkt för dess cellbildning, avgörande för embryots livskraft. Både embryo och endosperm omsluts av ett fröskal, en vävnad från moderplantan som fungerar som skydd och även är viktig för fröutveckling. Trots att fröskalet inte uppstår direkt vid befruktning, reagerar det genom att expandera och differentiera tillsammans med embryo och endosperm. Därför kräver en lyckad fröutveckling en uppsättning signalkedjor som samordnar tillväxt av embryo, endosperm och fröskal. Detta arbete visar att auxin har en nyckelroll i dessa signalkedjor. Auxin som bildas i endospermet beror av aktiviteten av genetiskt präglade gener från faderplantan och har två funktioner: i) det initierar och upprätthåller endospermutveckling och ii) det transporteras till fröhöljet vilket kick-startar utveckling av fröskalet. Senare i fröutvecklingen krävs en lägre auxinaktivitet i endospermet för att cellbildning ska initieras, och fröabortering som observerats i korsningar mellan individer med olika ploiditet kan kopplas till onormala auxinnivåer. I detta arbete identifierades också två endosperm-uttryckta MADS-box-transkriptionsfaktorer (TFs), som uppströms regulatorer av auxin-biosyntes och transportgener. Intressant är att en av dessa TFs - den epigenetiskt reglerade paternellt uttryckta *PHERES1* - styr uttrycket av flera andra präglade gener och binder sina mål-gener genom specifika bindningsställen i DNAt, vilka är signifikant associerade med transposoner (TEs). Vi vet sedan tidigare att TEs använts till att skapa epigenetiska landskap för präglat genuttryck. Vidare föreslår jag att dessa TEs har samordnats som cis-regulatoriska sekvenser som underlättar *PHERES1*-bindning till promotorer av präglade gener. Sammanfattningsvis bidrar denna avhandling till att ytterligare öka vår kunskap om fröutveckling: det avslöjar inte bara hormonet auxin som en nyckelfaktor som samordnar utvecklingen av olika frökomponenter, utan klargör även MADS-boxgenernas gäckande roll i de transkriptionella nätverk som kontrollerar endospermutveckling.

Nyckelord: frö, endosperm, fröskal, auxin, genetisk prägling, MADS-box, *PHERES1*

Författarens adress: Rita Adriano Batista, SLU, Department of Plant Biology, Linnean Centre for Plant Biology, P.O. Box 7800, SE-75007, Uppsala, Sweden

Mecanismos de sinalização e regulação epigenética do desenvolvimento da semente em *Arabidopsis thaliana*

Sumário

Em angiospermas, a formação de uma semente começa com o evento da dupla fertilização, que gera o embrião e o endosperma. Apesar de só o embrião contribuir para a formação da nova geração, o controlo da proliferação do endosperma, bem como da sua celularização, são cruciais para a viabilidade do embrião. O embrião e o endosperma estão rodeados pelo tegumento, um tecido materno importante para diversos aspectos do desenvolvimento da semente. O tegumento não é fertilizado, no entanto, através da sua expansão e diferenciação, é capaz de acompanhar o crescimento do embrião e do endosperma. Assim, torna-se evidente que o desenvolvimento da semente requer mecanismos de sinalização precisos que coordenem o crescimento do embrião, do endosperma, e do tegumento. Neste estudo demostrámos que a hormona auxina é um factor central nestes mecanismos de sinalização. A produção de auxina ocorre no endosperma, depende da actividade de genes *imprinted* e expressos paternalmente, e serve dois propósitos: i) inicia e mantém a proliferação do endosperma, e ii) fornece auxina que será exportada para o tegumento, iniciando o seu desenvolvimento. Este estudo mostra também que em fases mais tardias do desenvolvimento da semente, um decréscimo da actividade da auxina é essencial para iniciar a celularização do endosperma. Este estudo levou também à identificação de dois factores de transcrição do tipo MADS-box, que controlam a expressão de genes de biossíntese e transporte de auxina no endosperma. Um destes factores de transcrição é *PHERES1*, um gene *imprinted* e expresso paternalmente, que controla a expressão de inúmeros outros genes *imprinted*. *PHERES1* acede aos seus genes-alvo através de sequências de DNA que estão contidas em transposões. A presença de transposões foi previamente associada à geração de configurações epigenéticas passíveis de criar genes *imprinted*. Aqui proponho que estes transposões foram domesticados pela planta hospedeira, providenciando sequências regulatórias que facilitam o acesso de *PHERES1* aos seus genes-alvo. No geral, esta tese gerou uma contribuição significativa para melhorar a compreensão do desenvolvimento de sementes: não só revelou que a auxina é um factor central que coordena o crescimento dos vários componentes da semente, como também clarificou que os genes MADS-box são importantes mediadores das redes transcricionais que controlam o endosperma.

Palavras-chave: semente, endosperma, tegumento, auxina, *imprinting*, MADS-box, *PHERES1*

Morada da autora: Rita Adriano Batista, SLU, Department of Plant Biology, Linnean Centre for Plant Biology, P.O. Box 7800, SE-75007, Uppsala, Sweden

Dedication

Aos meus avós Emília e Elvino, que tantas saudades deixaram.

Aos meus avós Senhorinha e Feliciano.

À minha irmã.

Aos meus pais.

Contents

List of publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Figueiredo, D. D., **Batista, R. A.**, Roszak, P. J. & Köhler*, C. (2015). Auxin production couples endosperm development to fertilization. *Nature Plants*, vol 1 (12), pp. 15184.

II Figueiredo, D. D., **Batista, R. A.**, Roszak, P. J., Hennig, L. & Köhler*, C. (2016). Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *eLife*, vol 5, pp. 1-23

III **Batista**§**, R. A.**, Figueiredo§, D. D., Santos-González, J. & Köhler*, C. (2019). Auxin regulates endosperm cellularisation in *Arabidopsis*. *Genes & Development*, vol 33, nos. 7-8, pp. 466-476.

IV **Batista, R. A.**, Moreno-Romero, J., van Boven, J., Qiu, Y., Santos-González, J., Figueiredo, D. D., & Köhler*, C. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposable elements.

Paper I is reproduced here as a peer-reviewed author's version, following Springer Nature's copyright guidelines. Paper II is reproduced as the published version, under the Creative Commons Attribution 4.0 International Public License. Paper III is reproduced as a peer-reviewed author's version, following Cold Spring Harbour Laboratory Press' copyright guidelines. Paper IV is reproduced as an author's version.

* Corresponding author.
§ These authors contributed equally to this work

The contribution of Rita Adriano Batista to the papers included in this thesis was as follows:

I Performed experimental design; Carried out experimental procedures; Analysed data and discussed results; Commented on the manuscript.

II Performed experimental design; Carried out experimental procedures; Analysed data and discussed results; Commented on the manuscript.

III Performed experimental design; Carried out experimental procedures; Analysed data and discussed results; Wrote the manuscript.

IV Performed experimental design; Carried out experimental procedures; Analysed data and discussed results; Wrote the manuscript.

1 Introduction

Seeds are ubiquitous in nature, and their evolutionary origin correlates with a dramatic increase in reproductive success and the colonisation of new habitats. In fact, most land plants existing today are seed-bearing, among which flowering plants – or angiosperms – constitute around 90% of all known plant species (Corlett, 2016). The appearance of seeds in flowering plants is associated with a double fertilisation event, generating not one, but two distinct fertilisation products – the embryo and the endosperm. Thus, the endosperm is a biparental sexual structure, which nevertheless is ephemeral, and does not contribute to the establishment of the new generation. For this reason, the nature of the endosperm has long fascinated researchers – What is its evolutionary origin? What is its function during seed development? And how does its growth synchronise with other seed components, such as the embryo and seed coat? Furthermore, researchers often wondered about the biparental composition of the endosperm: several works have suggested this tissue is a battlefield, where maternal and paternal interests collide. One example of this is the presence of genomic imprinting in the endosperm, a mechanism by which certain genes are expressed in a parental-specific manner. But what is the significance of this epigenetic phenomenon?

These questions remain relevant to this day, and this thesis aims to contribute to finding the answers. The main goals of this work were to uncover the factors regulating endosperm development, to identify the signalling mechanisms acting between different seed components, as well as to understand the regulation of genomic imprinting. In this introduction, I will describe the developmental processes leading to the formation of a seed **(section 1.1.1 and 1.1.2)**, discuss the current knowledge on endosperm development **(section 1.1.4)** and remaining seed components **(sections 1.1.3 and 1.1.5)**, as well as the known communication pathways between them **(section 1.1.6)**. Additionally, I will provide some context on two key players in endosperm development – MADS-box transcription factors **(section 1.2)**, and the phytohormone auxin **(section**

1.3). I will also focus on the epigenetic regulation of seed development, paying special attention to genomic imprinting **(section 1.4)**. Finally, I will elucidate how deregulation of endosperm developmental pathways contributes to the establishment of reproductive barriers that can be involved in driving plant speciation **(section 1.5)**. Exploring these introductory concepts is essential for the understanding of the results and conclusions of this thesis. These are overviewed in **sections 3 through 5**, and explored in more detail in the manuscripts compiled in the last section.

1.1 Seed development

The life cycle of all multicellular plants can be described as an alternation of two generations – the sporophyte and the gametophyte. The sporophyte is a diploid (2n) structure capable of producing haploid (n) spores through meiosis. Each spore develops into a gametophyte, capable of producing haploid gametes through mitosis. During fertilisation, the fusion of two haploid gametes gives rise to the diploid sporophyte, thus perpetuating the cycle.

In early plants, such as some multicellular algae, liverworts, and mosses, the haploid gametophytic phase is dominant. The shift to a predominant sporophytic life cycle occurred later, with the evolution of land plants and the acquisition of traits such as vasculature and roots. This was followed by another significant breakthrough in plant evolution - the advent of seeds. The fact that most land plants existing today have a sporophytic-dominant life cycle and are seed-bearing, attests to the efficiency of both these strategies (Corlett, 2016; Baroux & Grossniklaus, 2018).

Several different characteristics of seeds have contributed to their evolutionary success: i) they are small dispersal units where the new generation is conveniently packed and can be spread by wind, animals, water, etc., thus allowing the colonisation of new environments; ii) seeds are able to remain dormant until proper conditions are met for their germination; iii) they are able to interpret environmental cues and evaluate if the ideal conditions for germination are met; iv) and finally, once germination starts, seeds are equipped with a sufficient nutrient storage to kick-start the growth of the new plant (Baroux & Grossniklaus, 2018).

Seed-bearing plants, called spermatophytes, can be divided into two main clades: the gymnosperms and the angiosperms. The formation of seeds in gymnosperms relies on nutrient allocation that is independent of fertilisation (Leslie & Kevin Boyce, 2012). Consequently, prior to fertilisation, the maternal plants have to accumulate resources in preparation for the growth of the future embryo. This can represent a large waste of resources, in case fertilisation is

unsuccessful. This was progressively circumvented in flowering plants (*i.e.* angiosperms), where multiple innovations have further improved their reproductive success (Leslie & Kevin Boyce, 2012).

Among the most striking traits developed in angiosperms is the double fertilisation event, where two fertilisation products are formed – the embryo and the endosperm. In order to achieve this, angiosperms also developed female gametophytes containing two gametes passible of fertilisation (Drews & Koltunow, 2011). Importantly, coupling of endosperm development to fertilisation ensures that allocation of resources only takes place in case an embryo is present. This energy-management strategy is likely one of the factors contributing for the ecological dominance of angiosperms (Baroux *et al.*, 2002).

In this section, I describe the development and morphology of female and male gametophytes in angiosperms. This overview is required for the understanding of both the double fertilisation event, and the final composition of an angiosperm seed, which I will also discuss. Finally, I will explore the development and function of each seed component, as well as the signalling mechanisms between them.

1.1.1 Gametogenesis

In angiosperms, the production of gametes occurs within the gametophyte, which exists in specialized sporophytic tissues – the flowers. Within the flower, the sporophytic anthers give rise to the male gametophyte (pollen grain), while the sporophytic pistil, containing the ovary, gives rise to the female gametophyte (embryo sac).

Female and male gamete production involves two sequential phases: i) sporogenesis, where meiosis occurs to form haploid undifferentiated spores, and ii) gametogenesis, where several rounds of mitosis, accompanied by cell differentiation, generate the mature gametophyte. These processes are described in detail below.

Female gametophyte development

Different angiosperm genera follow slightly distinct sporogenesis and gametogenesis strategies for female gamete formation. These differences reside mostly in the number of haploid spores that give rise to the gametophyte, as well as in the number of mitotic divisions required to achieve maturity. This leads to discrepancies in the final number of cells in the embryo sac; nevertheless, the different types of cells that are represented are usually the same (Maheshwari, 1950; Schmid *et al.*, 2015). For the sake of simplicity, the process described here is the one observed in *Arabidopsis thaliana* and other Brassicaceae, which is

characterized by a monosporic gametophyte of the *Polygonum* type – the most common within angiosperms (≥ 70% of known species) (Maheshwari, 1950; Yadegari & Drews, 2004).

Sporogenesis takes place inside the flower pistil, where finger-like protrusions of sporophytic tissue surround a single diploid cell – the megaspore mother cell (MMC). The MMC undergoes meiosis to produce four haploid spores, of which three degenerate, and only one survives – the functional megaspore (Christensen *et al.*, 1997; Bajon *et al.*, 1999; Drews & Koltunow, 2011). The functional megaspore then enters gametogenesis, undergoing three consecutive rounds of mitosis, to produce a total of eight haploid nuclei. This is followed by a cellularisation step, whereby these eight nuclei are organized into seven cells, forming the embryo sac (Christensen *et al.*, 1997; Drews & Koltunow, 2011). These cells are: one egg cell, which will be fertilised to give rise to the embryo **(section 1.1.3)**; two synergid cells, involved in pollen tube attraction and reception **(section 1.1.2)**; three antipodal cells that degenerate shortly after fertilisation, but whose function is still unclear (Song *et al.*, 2014); and two polar nuclei that fuse to generate one homodiploid central cell, which will be fertilised to give rise to the endosperm **(section 1.1.4)**. Consequently, the female gametophyte is composed of four different cell types, of which only two will contribute to the fertilisation products **(Figure 1)**.

Ovule sporophytic tissues

Throughout sporogenesis and gametogenesis, the female gametophyte is contained within a maternally-derived sporophytic tissue. This tissue initially develops as an epidermal layer that protrudes from the pistil, and encloses the MMC (Robinson-Beers *et al.*, 1992; Bajon *et al.*, 1999). At this stage a proximal-distal axis is established in this protrusion: the proximal region consists of the funiculus, responsible for attachment and communication with the mother plant, and the distal region contains the MMC (Christensen *et al.*, 1997; Drews & Koltunow, 2011). Meiosis of the MMC is accompanied by a periclinal division of the epidermal layer surrounding it. This division originates two layers: the inner and the outer integument (Schneitz *et al.*, 1995). During gametogenesis each integument layer creates a ring-like structure that progressively grows to surround the gametophyte (Schneitz *et al.*, 1995; Bajon *et al.*, 1999). Later, both integuments undergo additional periclinal divisions. In *Arabidopsis,* this leads to the formation of five cell layers that surround the gametophyte: two layers derived from the outer integument and three layers derived from the inner integument (Robinson-Beers *et al.*, 1992; Schneitz *et al.*, 1995). Throughout this process, the ovule – consisting of the embryo sac and its surrounding integuments – remains permanently connected to the maternal tissues via the

funiculus. Upon fertilization, the ovule integuments will differentiate into the seed coat, whose growth accompanies seed development and is essential for successful seed formation, dormancy, and germination **(section 1.1.5)**.

Male gametophyte development

The male gametophyte is produced in the flower anthers. The anthers are formed from sporophytic tissue, and are characterized by a bilobed structure with two distinct chambers, a common feature among angiosperms (Maheshwari, 1950). In each chamber of the anther resides a population of diploid pollen mother cells. Sporogenesis starts with the meiotic division of these cells, culminating in the production of a tetrad of haploid spores (Maheshwari, 1950; Owen & Makaroff, 1995). This tetrad is initially a syncytium enclosed in a particularly thick callose wall, derived from the pollen mother cell. Following meiosis, this wall is degraded, and the four individual haploid microspores are released (Mascarenhas, 1989; Owen & Makaroff, 1995).

This microspore release marks the beginning of gametogenesis, whereby the microspores undergo an asymmetric mitotic division. A first round of mitosis generates two cells: the larger vegetative cell, and the smaller generative cell. The latter is engulfed by the vegetative cell, thus effectively residing in its cytoplasm (Twell, 2011; Hafidh *et al.*, 2016). After this, an additional round of mitosis of the generative cell occurs to form two sperm cells. This last division can happen while the pollen is still in the anthers - as it is the case in *Arabidopsis* - or after pollen release (Maheshwari, 1950; Hafidh *et al.*, 2016). Thus, the mature pollen grain contains two identical haploid sperm cells, which will fertilize the egg and the central cell of the female gametophyte **(Figure 1)**. The companion vegetative cell generates the pollen tube, which transports the sperm cells towards the female gametophyte to achieve fertilization **(section 1.1.2)**.

1.1.2 Double fertilisation

As described in the previous section, both male and female gametophytes contain two gametes each. Upon arrival of the pollen tube to the ovule, a double fertilisation event takes place, whereby the two identical sperm cells will fertilise the egg cell and the central cell. Double fertilisation was first described about 120 years ago in several lily species (Nawaschin, 1898; Guignard, 1899), and is regarded as a distinguishing feature of angiosperms (Raghavan, 2006). However, this process is not unique to flowering plants, since double fertilisation is also observed in some gymnosperms, as is the case of several species of the Gnetales family (Carmichael & Friedman, 1996). The key differences between double fertilisation in the Gnetales and in angiosperms are the products of

fertilisation: in the Gnetales, two identical embryos are produced, of which only one will originate a new plant; while in angiosperms, only one embryo is produced, and the second fertilisation product consists of the endosperm. Several authors have proposed that the double fertilisation observed in angiosperms evolved from Gnetales, nevertheless there is still some controversy on whether this is the case (Raghavan, 2006).

While the evolutionary origin of double fertilisation in angiosperms still remains unresolved, the mechanistic details that govern this process have been greatly clarified thanks to the development of intricate live-imaging techniques and genetic approaches. After the pollen lands on the stigma, germination of the pollen grain takes place, and the vegetative cell will form a tube that grows inside the pistil and towards the ovules, transporting the sperm cells inside. This pollen tube has remarkable mechanisms that enable it to interpret chemical signals derived from the maternal pistil tissues, as well as from the ovule integuments and synergids (Mizuta & Higashiyama, 2018; Zhou & Dresselhaus, 2018). Chemical signalling is essential for the guidance of the pollen tube, its navigation within the pistil, and its arrival at the ovule. Once there, the pollen tube navigates through an opening at the integuments, and enters the female gametophyte through one of the two synergid cells. Upon entering, the pollen tube bursts, triggering the death of the synergid cell and the release of the two sperm cells: one of which will fuse with the egg, forming the embryo; while the other will fuse with the central cell to form the endosperm (Hamamura *et al.*, 2011; Zhou & Dresselhaus, 2018).

The products of the double fertilisation event – embryo and endosperm – will then develop synchronously while encased in the maternal seed coat. A detailed description of the development of these three structures is given in **sections 1.1.3 to 1.1.6**, and is represented in **Figure 1**. For the sake of simplicity, this description will focus on the model species *Arabidopsis thaliana*, and knowledge from other species will occasionally be referred to for comparative purposes.

1.1.3 Embryo

Karyogamy of the egg and sperm cell nuclei is followed by elongation of the zygote and a subsequent asymmetric division (Mansfield & Briarty, 1991; Palovaara *et al.*, 2016). This division leads to the establishment of a large basal cell and a small apical cell, which will have distinct fates during seed development: the apical cell gives rise to the embryo proper, while the basal cell originates the hypophysis and the suspensor **(Figure 1- stage1)**. The suspensor does not contribute directly to the embryo proper. Nevertheless, it provides

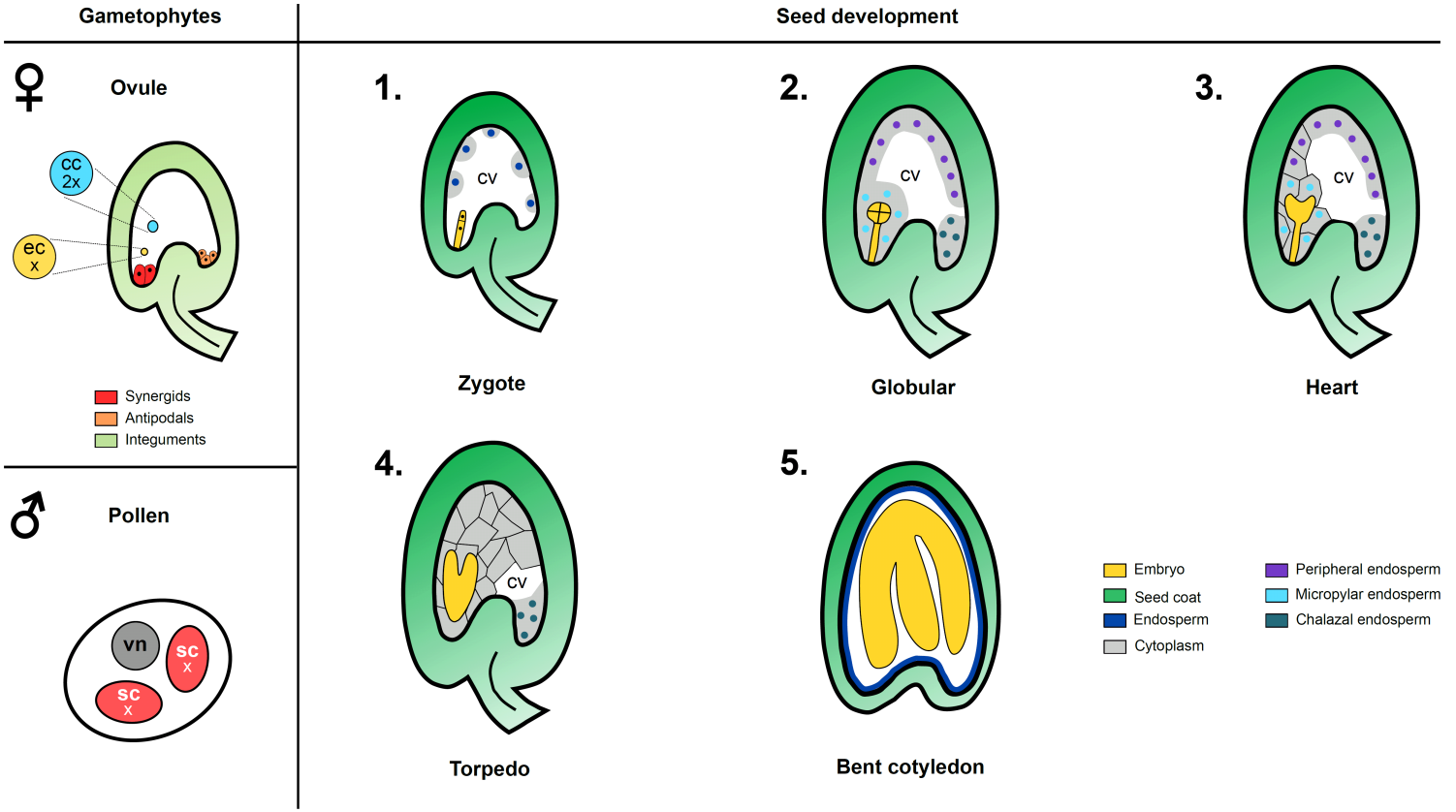

Figure 1. **Seed development in** ***Arabidopsis thaliana.*** Mature female and male gametophytes are shown in the left panel, while different stages of seed development are shown in the right panel. The ploidy of female and male gametes is indicated with x. Seed development stages are ordered from 1 (shortly after fertilisation), to 5 (maturity). The corresponding embryo stages are shown below each seed. Endosperm cellularisation is represented by the deposition of cell walls (stage 3, black lines). cc: central cell, ec: egg cell, vn: vegetative nucleus, sc: sperm cell, cv: central vacuole.

support to the growing embryo, and connects it to the maternal tissues and to the endosperm. This connection to the maternal tissues has been hypothesized to facilitate sucrose import during early stages of development (Kawashima & Goldberg, 2010). Furthermore, the suspensor has been shown to supply growth regulators that are essential for proper embryo identity (*e.g.* auxin) (Friml *et al.*, 2003; Palovaara *et al.*, 2016; Robert *et al.*, 2018).

The mature embryo derives from a series of cellular divisions of the embryo proper and of the hypophysis. These divisions are tightly coordinated in space and time, and rely on the integration of multiple developmental signals, which are extensively detailed in several reviews (Laux *et al.*, 2004; Capron *et al.*, 2009; Palovaara *et al.*, 2016; de Vries & Weijers, 2017). Initially, these divisions lead to a globular shaped embryo **(Figure 1- stage 2)**. Later, the development of the two cotyledon primordia, followed by their growth and elongation, gives rise to a heart shaped embryo (Mansfield & Briarty, 1991; Capron *et al.*, 2009). This coincides with the initiation of endosperm cellularisation, and programmed cell death of the suspensor **(Figure 1- stage 3 and 4)** (Mansfield & Briarty, 1990b; Bozhkov *et al.*, 2005). In the final stages of seed development, a pronounced growth of the cotyledons and of the hypocotyl causes the embryo to bend on itself, in order to fit in the seed cavity **(Figure 1- stage 5)**. This fast growth is accompanied by accumulation of storage compounds, such as lipids and proteins, and by consumption of the surrounding endosperm (Baud *et al.*, 2008). At maturity, the seed cavity will be mostly composed of the bent dicotyledonar embryo, which enters dormancy until the proper conditions for germination are met (Baud *et al.*, 2008).

1.1.4 Endosperm

The endosperm has long been hypothesized to be an accessory nourishing tissue that aids embryo growth and development. Numerous studies have corroborated this idea and have described the endosperm as a sink tissue within the seed, which accumulates sugars, lipids, and proteins (Maheshwari, 1950; Bhatnagar & Sawhney, 1981; Fincher, 1989). In Brassicaceae, a complex mechanism of sugar transport, from maternal tissues to the endosperm, has been described (Chen *et al.*, 2015). Once in the endosperm, sucrose is converted and stored in the form of hexose, and is subsequently transferred to the embryo as the need arises (Hill *et al.*, 2003; Morley-Smith *et al.*, 2008; Hehenberger *et al.*, 2012; Lafon-Placette & Köhler, 2014). In this family, the endosperm is almost fully consumed by the growing embryo, and therefore contributes very little to the final volume of the seed (Olsen, 2004; Yan *et al.*, 2014). However, in other species, an additional role of assisting germination and seedling growth requires

that the endosperm is preserved in the mature seed, making up most of its volume (Bhatnagar & Sawhney, 1981; Olsen, 2004; Yan *et al.*, 2014). For example in cereals, upon germination, the endosperm was shown to respond to signals derived from the embryo. These signals trigger breakdown of storage compounds in the endosperm, making them available for the emerging seedling (Fincher, 1989).

Besides its nourishing function, the endosperm also functions as a signalling mediator that coordinates seed development. This is exemplified by studies that used the *kokopelli* (*kpl*) mutant as a pollen donor: *kpl* pollen only contains one sperm cell and, as a consequence, only one fertilisation event takes place. Interestingly, the absence of an endosperm leads to embryo arrest (Ron *et al.*, 2010). Notably, if no endosperm is produced, the maternal integuments do not develop into a seed coat (Roszak & Köhler, 2011). These observations support the role of the endosperm as a nourishing entity for the embryo, but also indicate that upon fertilisation, an endosperm-derived signal is necessary for the differentiation of the integuments into a seed coat **(section 1.1.6)**.

Endosperms of flowering plants can be divided into three major groups, according to their morphology and mode of cell division: cellular, nuclear and helobial endosperm (Maheshwari, 1950; Bhatnagar & Sawhney, 1981). Cellular endosperm is characterised by rounds of mitosis which are always accompanied by cell wall deposition. Although this is not the most common type of endosperm, it occurs in several relevant crop species, such as those of the Solanaceae family (*e.g.* potato, tomato, tobacco, among others) (Lester & Kang, 1998). Helobial endosperm is the rarest type of endosperm and is characterised by an initial mitotic division accompanied by cell wall deposition. This division gives rise to two daughter cells, each of which will constitute a distinct endosperm domain – the micropylar domain (originated from the cell closest to the embryo), and the chalazal domain (derived from the cell furthest from the embryo). As a general rule, the micropylar domain will form a nuclear endosperm that undergoes cellularisation later in development, while the chalazal domain will develop as a purely nuclear endosperm (Maheshwari, 1950; Bhatnagar & Sawhney, 1981).

Nuclear endosperm is the most prevalent among angiosperms, most notably it is present in cereals, as well as in the Brassicaceae (Olsen, 2004), and is characterised by several rounds of mitosis without cell wall deposition, meaning the early endosperm exists as a syncytium. During this phase, the nuclear divisions are tightly regulated, both in time and space, and distinct endosperm domains are formed (Mansfield & Briarty, 1990a; Boisnard-Lorig *et al.*, 2001; Olsen, 2004). These domains show different morphological characteristics, as well as distinct transcriptional profiles (Belmonte *et al.*, 2013). They can be

divided into three types: micropylar, peripheral, and chalazal endosperm **(Figure 1- stage 2)**. The micropylar endosperm surrounds the developing embryo, and is hypothesized to have an important role in the communication between embryo and endosperm **(section 1.1.6)**; the peripheral endosperm consists of uniformly distributed nuclei in the periphery of the seed cavity; and the chalazal endosperm is characterized by a high level of endoreduplication, a very dense cytoplasm, and is hypothesised to have a role in unloading resources derived from the maternal tissues (Brown *et al.*, 1999; Olsen, 2004). During the syncytial endosperm stage, a large central vacuole is formed and functions as a sink for sucrose, which is imported from the maternal tissues and converted into hexoses by the action of vacuole-specific invertases (Hill *et al.*, 2003; Morley-Smith *et al.*, 2008; Hehenberger *et al.*, 2012; Lafon-Placette & Köhler, 2014). The large volume of the vacuole pushes the dividing endosperm nuclei to the periphery of the seed cavity. After this initial phase, the endosperm undergoes cellularisation, where cell walls are deposited to define uninucleate cells **(Figure 1- stage 3)**.

Sustainable division of the syncytial endosperm requires unknown paternal factors

After karyogamy of sperm and central cell nuclei, the nuclear endosperm starts rapid mitotic divisions, even before the first zygote division takes place (Faure *et al.*, 2002). These rapid divisions were noticed by several botanists investigating the double fertilisation event, leading to the hypothesis that fertilisation conferred a potent "growth-stimulus" (Sargant, 1900). Nevertheless, the nature of this "growth-stimulus" has remained elusive. Several studies have paved the way for a better understanding of this process: it was shown that sperm cell delivery to the central cell, without occurrence of karyogamy, is able to trigger spontaneous mitotic divisions – a processed termed autonomous endosperm formation (Aw *et al.*, 2010). Additionally, release of pollen tube cytoplasm into the female gametophyte also leads to spontaneous division of the unfertilised central cell (Susaki *et al.*, 2016). However, it is worth noting that the autonomous endosperm division reported in both these studies is limited to very few mitotic divisions. These cannot be sustained over time and, importantly, do not accompany embryo growth. In parallel, other studies have shown that in mutants where a single fertilisation event occurs, the absence of a zygote still allows the fertilised central cell to form a viable endosperm (Aw *et al.*, 2010; Roszak & Köhler, 2011).

Together these studies indicate that: i) endosperm division after fertilisation does not require a zygote-derived signal; ii) pollen tube burst, either by delivery of its cytoplasmic contents, delivery of sperm cells, or through another unknown mechanism, is able to stimulate mitotic divisions of the central cell; and iii)

formation of actively-dividing endosperm that accompanies embryo development is dependent on fertilisation of the central cell, and therefore requires the presence of the paternal genome.

Understanding the trigger of endosperm proliferation, as well as the mechanisms sustaining it over time, is not only important to gain a better understanding of endosperm development, but is also essential to comprehend the seed abortion events that underlie several endosperm-based reproductive barriers **(section 1.5, paper III, and paper IV)**. For these reasons, one of the aims of this thesis was to identify and characterise this mysterious "growth-stimulus", which is conferred by the paternal genome **(paper I)**.

The mechanisms regulating endosperm cellularisation are poorly understood

After the syncytial stage, the endosperm undergoes an important developmental transition – cellularisation (Mansfield & Briarty, 1990b). During this process cell walls are deposited gradually, starting from the micropylar endosperm surrounding the embryo, and then progressing to the peripheral endosperm **(Figure 1- stages 3 and 4)**. This is followed by a centripetal pattern of cell wall deposition (Brown *et al.*, 1999; Olsen, 2004). As the endosperm divides and cellularises, the size of the central vacuole decreases, giving space to the growing embryo and endosperm (Brown *et al.*, 1999; Olsen, 2004). The physiological implication of cellularisation is still somewhat unclear, with some authors suggesting it allows for the embryo to replace the vacuole as the main energy sink in the seed. This would lead to the endosperm transitioning from a sink tissue into a source tissue (Hehenberger *et al.*, 2012; Lafon-Placette & Köhler, 2014). Several studies have demonstrated that the occurrence of cellularisation, as well as its timing, are crucial for proper seed development. This is illustrated by mutants of the FERTILISATION INDEPENDENT SEED-POLYCOMB REPRESSIVE COMPLEX 2 (FIS-PRC2) **(section 1.4)**, by crossings of parents of different ploidies **(section 1.5)**, as well as in certain instances of interspecies crosses **(section 1.5)**. In all these cases, the timing of endosperm cellularisation is disrupted, leading to seed abortion. *In vitro* cultivation of embryos derived from these aborting seeds allows embryo survival, suggesting endosperm cellularisation directly impacts on embryo development (Brink & Cooper, 1947; Hehenberger *et al.*, 2012; Rebernig *et al.*, 2015; Lafon-Placette *et al.*, 2017).

The observations that endosperm cellularisation is preponed or postponed in seeds where endosperm proliferation is reduced or increased (Scott *et al.*, 1998), has led to the hypothesis that cellularisation and endosperm proliferation are intrinsically linked processes that share common regulatory mechanisms. However, several pieces of evidence challenge this idea. First, seeds with similar numbers of endosperm nuclei can initiate cellularisation at different time-points

(Ishikawa *et al.*, 2011). Conversely, cellularisation can occur at similar time-points in seeds that contain dramatically different amounts of endosperm nuclei (Scott *et al.*, 1998; Bushell *et al.*, 2003). Furthermore, in mutants where early endosperm cellularisation is concomitant with reduced nuclei numbers, preventing endosperm cellularisation by genetic means does not lead to changes in nuclear proliferation (Garcia *et al.*, 2003). All these pieces of evidence suggest that proliferation and cellularisation are processes that can be uncoupled, and which may be regulated independently. Several alternative hypotheses to explain the onset of endosperm cellularisation have been suggested: Garcia *et al.* (2003) proposed that cellularisation is triggered when a specific ratio between number of endosperm nuclei and cytoplasmic volume is reached. On the other hand, Beauzamy *et al.* (2016) show that turgor pressure exerted by the endosperm is high during early stages of endosperm development, and decreases upon cellularisation. Thus, the authors propose that this shift in mechanical properties of the endosperm could be causal to cellularisation. Notwithstanding, the mechanisms triggering endosperm cellularisation remain unknown. Uncovering the molecular pathways involved in this developmental transition was one of the aims of this thesis **(paper III)**.

1.1.5 Seed coat

The seed coat is a maternally-derived tissue that, despite not being fertilised, has the ability to respond to fertilisation by expanding and differentiating. It functions as a protective layer for the developing embryo and endosperm, and also has a crucial role in seed dormancy and germination (Baroux & Grossniklaus, 2018). Immediately after fertilisation, the seed coat of *Arabidopsis* consists of five distinct cell layers: three derived from the inner integument, and two derived from the outer integument. Additionally, a specialised seed coat region is present, which has a direct connection with the funiculus and the maternal plant – the chalazal seed coat (Beeckman *et al.*, 2000). All these structures are present in the mature ovule and develop during female gametogenesis (Schneitz *et al.*, 1995), meaning their identity is defined before fertilisation and is under exclusive maternal control.

The chalazal seed coat region is unique since it is the only region in the seed with direct communication with the maternal plant. It contains xylem and phloem vessels that connect to the funiculus. Thus, water and nutrients can be loaded into the seed via this region (Khan *et al.*, 2014; Millar *et al.*, 2015). This is supported by the unique transcriptional profile of this seed coat domain, where numerous genes coding for proteins involved in the transport of water, sugars, amino acids, and lipids are expressed (Khan *et al.*, 2014; Millar *et al.*, 2015).

This strongly suggests that the chalazal seed coat acts as an unloading region for maternally-derived nutrients, which can then be transferred to the remaining seed coat cell layers, as well as to the endosperm and embryo.

After fertilisation, the seed coat expands and creates a cavity to accommodate the growing embryo and endosperm. This expansion is mostly achieved through cell elongation, rather than cell division (Garcia *et al.*, 2005). In parallel to their growth, the different layers of the seed coat undergo specific differentiation events. In the innermost layer - the endothelium - production of proanthocyanidins (PAs) starts immediately after fertilisation (Debeaujon *et al.*, 2003). PA production is part of the complex flavonoid biosynthetic pathway, of which multiple players have been thoroughly described (Debeaujon *et al.*, 2003; Lepiniec *et al.*, 2006). PAs are important for the establishment of seed dormancy, to facilitate germination of seeds after long-term storage, as well as for defence against pathogens (Debeaujon *et al.*, 2002; Shirley, 2008). Moreover, PAs are responsible for the characteristic brown colour of *Arabidopsis* and other Brassicaceae seeds, which appears after the oxidation of these compounds, in later stages of seed development.

In the outermost layers of the seed coat, fertilisation triggers the accumulation of numerous starch granules. As seed development progresses, the outermost cell layer secretes mucilage into the space between the plasma membrane and the outer cell wall (Windsor *et al.*, 2000). This leads the cytoplasm to become restricted within the centre of the cell, forming a structure denominated columella. Progressively, the accumulated starch will be converted into secondary cell walls, reinforcing the columella structure (Windsor *et al.*, 2000).

During later stages of seed development, a gradual collapse of the different seed coat cell layers is observed: the two outermost layers of the inner integument are the first to collapse, followed by the innermost layer of the outer integument (Beeckman *et al.*, 2000). Thus, the mature seed coat is composed of two cell layers, the outermost layer of the outer integument, which is rich in mucilage, and the innermost layer of the inner integument. In between these, an amorphous layer composed of the collapsed cells accumulates the produced PAs (Beeckman *et al.*, 2000; Windsor *et al.*, 2000). Together, PAs and mucilage are thought to isolate the seed from the outside environment, so that germination is only triggered when the appropriate conditions are met (Haughn & Chaudhury, 2005). When the seed contacts with water, and in preparation for germination, the mucilage is rapidly released from the seed coat, providing a humid environment that aids germination (Windsor *et al.*, 2000; Haughn & Chaudhury, 2005).

1.1.6 Seed development requires coordinated communication between seed components

Embryo patterning, endosperm cellularisation, and seed coat differentiation are interconnected, and occur at specific time-points during seed development (Baroux & Grossniklaus, 2018). This strongly suggests the existence of a developmental coordination between these different seed components. Communication is important to ensure that the nutrients taken up by the seed are transported to the correct seed compartments. Additionally, signalling between the endosperm and the seed coat is required to trigger seed coat development. Therefore, communication between seed components must take place not only to transmit developmental cues, but also to allow for dynamic allocation of resources throughout seed development.

Nonetheless, this communication is restricted by several physical barriers: no symplastic connection is observed between endosperm and embryo, endosperm and seed coat, as well as between the inner and outer integuments of the seed coat (Stadler *et al.*, 2005; Ingram, 2010). Furthermore, apoplastic communication can potentially be constrained by the presence of cuticles in the embryo proper and in the endothelium (Ingram, 2010). All these observations suggest that signalling is most likely selective for certain molecules, and that it relies on active transport of substances across cells.

A substantial number of transporters have been described to play a role in sucrose transport within and across different seed compartments. By analysing the expression of several of these transporters, Chen *et al.* (2015) proposed that sucrose is sequentially unloaded form the phloem into the outer integuments, and from there to the inner integuments, endosperm, and embryo. Within the endosperm, the embryo surrounding region (ESR), seems to play a central role in communication with the embryo. This is illustrated by the fact that many transporters are active in this region (Baud *et al.*, 2005; Chen *et al.*, 2015). Furthermore, mutants lacking these transporters, although viable, produce seeds with small and malnourished embryos (Baud *et al.*, 2005; Chen *et al.*, 2015). This reinforces the idea that transport of resources across different compartments is important for correct seed development.

Seed size regulation is dependent on communication between seed components, and both the endosperm and the seed coat are able to control the growth of one another in a non-cell-autonomous manner. This can be illustrated by the analysis of several mutants where seed size is affected. These can be divided into two categories, according to their effect: i) mutants affecting endosperm growth in a zygotic manner, such as *haiku1* and *2* (*iku1*, *iku2*) (Garcia *et al.*, 2003, 2005), and the related mutants *miniseed3* (*mini3*) (Luo *et al.*, 2005), and *short hypocotyl under blue1* (*shb1*) (Zhou *et al.*, 2009); ii) mutants affecting

seed coat division and/or expansion in a maternal sporophytic manner, such as *apetala2* (*ap2*) (Ohto *et al.*, 2005), and *transparent testa glabra2* (*ttg2*) (Debeaujon *et al.*, 2003; Garcia *et al.*, 2005). Overall, these studies show that reduced growth of the endosperm leads to a reduction in seed coat expansion, and vice-versa; while increased seed coat expansion promotes endosperm growth. Thus, and even though the specific signalling mechanisms between endosperm and seed coat are not fully elucidated, it is clear that the development of these two compartments is tightly intertwined.

Several communication mechanisms, such as small peptide signalling and mechanical sensing, have been recently explored. A staggering amount of genes coding for small peptides are expressed during seed development (Ingram & Gutierrez-Marcos, 2015). However, so far, only a few have been implicated in intercellular communication. Most notable is the cysteine-rich peptide ESF1, which accumulates in the ESR region of the endosperm and is important for correct embryo patterning (Costa *et al.*, 2014). This observation suggests that additional small peptides might have an important role in non-cell-autonomous signalling between seed components, but future work is necessary to unravel their function.

Mechanical signalling, on the other hand has been more thoroughly explored. It was suggested that the high turgor pressure generated by the early endosperm could potentially act as a mechanical stimulus for seed coat cell elongation (Beauzamy *et al.*, 2016). Furthermore, the same team found that the adaxial epidermis of the outer integument is able to sense pressure exerted by the endosperm, triggering cell wall thickening, and dampening gibberellic acid response (GA), which is involved in cell elongation (Creff *et al.*, 2015). Thus, the authors suggest that this is a maternal mechanism of sensing endosperm growth and restricting seed coat expansion.

Collectively, these data constitute a strong body of evidence for the existence of communication between different seed components, and indicate that several signalling mechanisms might take place simultaneously. However, one central question that remains is how seed coat development is initiated. In single fertilisation events leading to seeds that only contain an embryo, the seed coat does not develop (Roszak & Köhler, 2011). The same is observed in seeds where the endosperm was ablated (Weijers, 2003). Additionally, in mutants where fertilisation is not accompanied by karyogamy of the sperm cell and central cell, transient autonomous endosperm of maternal origin is formed, but this is not sufficient to trigger seed coat formation (Aw *et al.*, 2010). Consequently, these data suggest that seed coat development requires the presence of endosperm and of the paternal genome (Figueiredo & Köhler, 2016). Moreover, initiation of seed coat development is likely under the control of the endosperm-expressed

MADS-box transcription factor AGL62, as *agl62* mutant seeds fail to initiate seed coat development (Roszak & Köhler, 2011). Understanding the source and nature of this endosperm-generated signal that triggers seed coat development was one of the goals of this thesis **(paper II)**.

1.2 MADS-box transcription factors

MADS-box transcription factors (TFs) are ubiquitous among eukaryotes and are defined by the presence of a MADS protein domain. The acronym MADS derives from the original MADS-box TFs identified in different organisms: *MINICHROMOSOME MAINTENANCE1* (*MCM1*, yeast), *AGAMOUS* (*AG*, *Arabidopsis*), *DEFICIENS (DEF A*, snapdragon), and *SERUM RESPONSE FACTOR (SRF,* human). MADS-box TFs are thought to have originated in a common ancestor of extant eukaryotes (Gramzow & Theissen, 2010), and can be found in vertebrates, plants, insects, nematodes, and fungi (Messenguy & Dubois, 2003). Nevertheless, the number of these TFs varies dramatically across species: while most eukaryotes show a low number, plant genomes can encode several hundred members (Gramzow & Theissen, 2010; Theissen & Gramzow, 2016). The reason for this discrepancy is still unknown. However, within the plant kingdom, increased MADS-box TF numbers correlate with developmental innovations, such as the appearance of seeds and flowers (Gramzow & Theissen, 2010; Theissen & Gramzow, 2016). The importance of MADS-box TFs for eukaryotes is illustrated by the wide range of developmental processes they regulate (*e.g*. cell cycle transition in yeast; cell proliferation, differentiation, and response to growth factors in mammals; and establishment of floral organ identity in plants) (Shore & Sharrocks, 1995; Messenguy & Dubois, 2003).

The distinguishing feature of this diverse group of TFs is the presence of the MADS protein domain, responsible for binding to specific DNA sequences – CArG boxes (C-A-rich-G-box), with variations on the following sequence [$CC(A/T)_6GG$] (Shore & Sharrocks, 1995). The MADS domain is usually 55-60 aa in size and is hypothesized to originate from a duplication of a prokaryotic topoisomerase IIA subunit A gene, which later gave rise to a MADS-coding gene in eukaryotes (Gramzow *et al.*, 2010). Additional protein domains can be identified in plant MADS-box TFs, such as the intervening domain (I, involved in specifying protein dimerization properties), the keratin-like domain (K, involved in the formation of higher order complexes between MADS-box TF dimers), and the C-terminal domain (C, which can have different functions, such as stabilisation of dimers, and transactivation) **(Figure 2)** (Kaufmann *et al.*, 2005).

MADS-box TFs can be subdivided into Type I and Type II, according to the protein motifs they contain **(Figure 2)**. In plants, type I TFs contain a MADS- and a C- domain, as well as several unknown domains; while type II TFs contain MADS-, I-, K-, and C- domains (Alvarez-Buylla *et al.*, 2000; Nam *et al.*, 2004;

Kaufmann *et al.*, 2005). Thus, type II TFs can also be described in the literature as MIKC-type. These structural variations are translated in distinct sets of biological functions for plant type I and type II TFs: while the former are involved in gamete and seed development, the latter are implicated in many developmental processes, most prominently the control of flowering time morphology (Theissen & Gramzow, 2016).

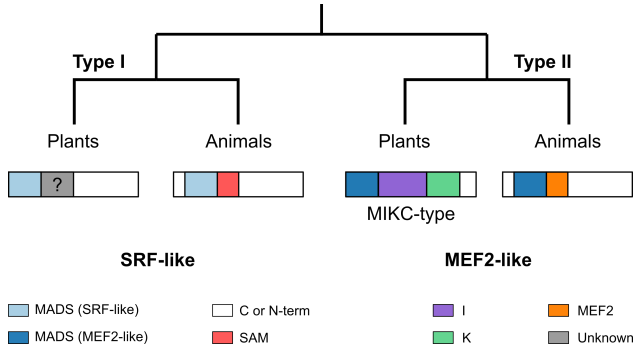

Figure 2. **Structure of animal and plant MADS-box transcription factors.** Each domain is indicated by a distinct colour. Unknown domains (?) are present in the C-terminal region of type I plant MADS-box TF. Adapted from Nam *et al.*, (2004), and Alvarez-Buylla *et al.*, (2000).

As mentioned above, MADS-box proteins contain several domains responsible for protein dimerization. Therefore, it comes as no surprise that these proteins are able to form both homo- and heterodimers with other MADS-box proteins, binding as dimers to CArG boxes (Riechmann *et al.*, 1996). This results in the presence of complex protein interaction and gene regulatory networks (Immink & Angenent, 2002; de Folter *et al.*, 2005). Interestingly, in *Arabidopsis* these proteins most often dimerize with MADS-box proteins of the same type, even though interaction between proteins of different types is also possible (de Folter *et al.*, 2005). Additionally, formation of higher-order structures - such as interactions between two MADS-box dimers - is also possible (Kaufmann *et al.*, 2005). So far, multimerization has only been observed for type II MADS-box TFs and it is thought to allow simultaneous accessibility to two different CArG boxes. Additionally, it is hypothesized that multimerization stabilises the interaction between DNA and the TF multimer, bringing distant CArG-boxes together, and thus promoting DNA conformation changes that allow access to distal promoter regions (Kaufmann *et al.*, 2005).

MADS-box TFs have also been shown to interact with other proteins, such as non-MADS TFs, as well as chromatin remodellers. This is widely observed in animals and yeast (Messenguy & Dubois, 2003), and a few similar cases have

been described in plants (Sridhar *et al.*, 2006; Smaczniak *et al.*, 2017). Yet, in plants, this has only been observed for type II MADS-box TFs and thus it is currently unknown if type I proteins are able to establish similar interactions. Nonetheless, the fact that MADS-box TFs are able to interact with chromatin remodellers motivated their classification as 'pioneer factors', which can trigger chromatin accessibility changes through their binding to DNA, contributing for the transcriptional regulation of the target loci (Pajoro *et al.*, 2014; Yan *et al.*, 2016).

As it is apparent from the last paragraphs, the majority of knowledge on plant MADS-box TFs is derived from type II TFs. Despite this lack of information on type I TFs, it has become recently evident that they play important roles in specific plant developmental stages, such as gametogenesis and seed development. Nevertheless, their exact functions, target genes, and developmental impact are mostly unknown (Theissen & Gramzow, 2016). One of the aims of this thesis was to uncover the biological role of some of these type I TFs, with particular emphasis on their involvement in endosperm and seed development. In the next section I will further characterise the current knowledge on these TFs, illustrate how they relate to seed development, as well as formulate some of the questions this thesis work aimed to address.

1.2.1 Type I MADS-box transcription factors have an important but elusive function in endosperm development

On average, there are about one hundred genes coding for MADS-box proteins in genomes of different angiosperms (Gramzow & Theissen, 2010) – 107 in *Arabidopsis thaliana* (Pařenicová *et al.*, 2003). Of those, about 50% code for type I TFs (Gramzow & Theissen, 2010). This abundance of type I genes seems to coincide with the appearance of flowering plants, as gymnosperms and other basal plant species show a strong underrepresentation of these genes (Gramzow & Theissen, 2010; Nystedt *et al.*, 2013; Gramzow *et al.*, 2014). Interestingly, it has been shown that type I MADS-box genes arise more frequently than type II genes. Generally, this is achieved through segmental duplications that cannot be explained by whole genome duplications (Nam *et al.*, 2004). Furthermore, type I genes show signs of higher purifying selection when compared to type II genes, which can explain their higher rate of "birth and death", as well as their high variability within and across species (Nam *et al.*, 2004). This reveals a strong evolutionary dynamics of type I MADS-box TFs.

Due to the intrinsic properties of these TFs, their identification and functional characterization in plant genomes has been delayed. Firstly, gene duplications

are common, and lead to high functional redundancy. Thus, single gene mutations lead to no observable phenotypes, making it challenging to identify these genes through forward genetic screens (Gramzow & Theissen, 2010). Secondly, the expression of many of these genes often falls below the detection thresholds, and for those genes where expression is detected, its levels are quite low (Leseberg *et al.*, 2006; Arora *et al.*, 2007; Bemer *et al.*, 2010). And finally, type I MADS-box genes are expressed during very short temporal windows within plant development, such as gametogenesis and seed development (Leseberg *et al.*, 2006; Arora *et al.*, 2007; Bemer *et al.*, 2010). Nevertheless, whole genome sequencing of *Arabidopsis thaliana* and of other plant species was the turning point for type I MADS-box research, as several genes were identified through homology searches, opening the doors for their functional characterisation.

Presently, several type I MADS-box TFs have been characterised in *Arabidopsis: AGAMOUS-LIKE 80* (*AGL80*) and *DIANA* (*AGAMOUS-LIKE 61, AGL61*) were shown to be involved in specifying central cell identity (Portereiko *et al.*, 2006; Bemer *et al.*, 2008), while *AGAMOUS-LIKE 23* (*AGL23*) is involved in female MMC development, and in chloroplast biogenesis in the embryo (Colombo *et al.*, 2008). Additionally, several type I MADS-box genes have been implicated in endosperm development. *AGAMOUS-LIKE 62* (*AGL62*) is an important example, since mutants for this endosperm-specific gene show a striking phenotype, where endosperm proliferation is reduced, cellularisation happens prematurely, and seed coat development is not initiated (Kang *et al.*, 2008; Roszak & Köhler, 2011). For these reasons, AGL62 was proposed to control the timing of endosperm cellularisation (Kang *et al.*, 2008), and to be responsible for generating a fertilisation-dependent signal that drives seed coat development **(section 1.1.6)** (Roszak & Köhler, 2011). Nevertheless, the specific mechanisms through which AGL62 controls these processes remain unknown.

Besides *AGL62*, additional type I MADS-box *AGL* genes have been associated with endosperm development, in particular with the timing of cellularisation. This association stems from multiple observations where endosperm-based seed abortion events are linked with the deregulation of numerous endosperm-expressed *AGL* genes. These observations took place in several plant species where endosperm cellularisation is disrupted in response to interploidy crosses (Erilova *et al.*, 2009; Tiwari *et al.*, 2010; Lu *et al.*, 2012), interspecies crosses (Walia *et al.*, 2009; Ishikawa *et al.*, 2011; Rebernig *et al.*, 2015; Roth *et al.*, 2018; Tonosaki *et al.*, 2018), or in mutants where genomic imprinting is lost, such as mutants for components of FIS-PRC2 **(section 1.4)** (Köhler *et al.*, 2003, 2005; Kang *et al.*, 2008; Erilova *et al.*, 2009; Tiwari *et al.*,

2010). The common feature in these observations is that the expression level and dynamics of type I MADS-box genes is somewhat predictive of the outcome of endosperm cellularisation, and consequently of seed viability: in seeds where cellularisation is delayed, *AGL* expression is increased and prolonged in time; the opposite happens in seeds where endosperm cellularisation occurs prematurely **(section 1.5)**. Among the type I genes showing this expression pattern, is *PHERES1* (*PHE1*, *AGL37*), which was initially identified as upregulated in FIS-PRC2 mutants (Köhler *et al.*, 2003). Decreasing *PHE1* expression partially alleviates the seed abortion phenotype present in these mutants (Köhler *et al.*, 2003), suggesting that this TF might have a role in controlling endosperm development.

Despite the body of evidence implicating type I MADS-box TFs in female gametogenesis and seed development, these data are mostly derived from mutant and gene expression analysis. While important to pave the way for further research, these studies do not address the mechanistic details through which these TFs control these developmental processes. One of the main aims of this thesis was to uncover these mechanisms. Special emphasis was given to uncover the roles of *AGL62* and *PHE1*, both in the context of endosperm and seed coat development **(paper I, II)**, but also in the context of interploidy crosses, where these genes seem to play an important role **(section 1.5; papers III and IV)**.

1.3 Auxin

In 1880 Charles Darwin, working with his son Francis, observed that seedling coleoptiles responded to light and gravity by changing their growth pattern and bending towards the source of the stimulus. Furthermore, the capacity to respond to these stimuli depended solely on the apex of the coleoptile, even though the changes in growth were observed throughout the whole organ. From this, the Darwins concluded that a signal of unknown nature travels form the apical to the basal part of the coleoptile, in order to generate the growth response (Darwin, 1880). Later on, the work of several biologists led to a progressive understanding of Darwin's observations, culminating with the identification of the hormone auxin in plants, in 1946 (reviewed in Abel & Theologis, 2010).

In fact, besides controlling tropic responses, as the Darwins and others observed, auxin influences a multitude of physiological processes. Among those are the control of cell expansion, division and differentiation, thus impacting key developmental processes such as shoot and root development, apical dominance, leaf abscission, among many others (Davies, 2010).

In the following sections, I will briefly introduce important aspects of auxin biosynthesis, transport, and signalling. Additionally, I will explore the potential role of auxin in controlling seed development, as well as formulate some of the hypotheses that this thesis aimed to address.

1.3.1 Mechanisms of biosynthesis, transport, and signalling

Even though indole-acetic acid (IAA) is the most common form of auxin in plants, other analogous molecules exist, which can trigger auxin-like responses (Woodward & Bartel, 2005). In line with this, there are distinct auxin biosynthesis pathways, which differ in their substrates, as well as in their final products. Two main pathways can be defined, according to their dependence on the amino acid tryptophan (Trp) as the initial precursor of auxin: Trp-dependent and Trp-independent (Kasahara, 2016; Zhao, 2018). Trp-dependent pathways are by far better characterised; nevertheless, Trp auxotrophs in *Arabidopsis* and maize accumulate higher levels of IAA when compared to wild-type plants, leading to the proposal of the existence of a Trp-independent auxin biosynthesis pathway (Wright *et al.*, 1991; Normanly *et al.*, 1993). Notwithstanding, there is still a debate on whether this putative pathway is of biological significance (Nonhebel, 2015).

Several routes to IAA production exist within Trp-dependent pathways, with the main differences being the enzymes mediating the reactions, as well as the intermediate products (Kasahara, 2016). For the purpose of this thesis, I will focus on what is considered the main Trp-dependent pathway: the one catalysed

by tryptophan aminotransferases (TAAs) and flavin monooxygenases (YUCCAs). In this pathway, Trp is first converted to indole-3-pyruvate by the action of TAAs (Mashiguchi *et al.*, 2011; Won *et al.*, 2011). This compound is then converted into IAA through the activity of YUCCAs (Mashiguchi *et al.*, 2011; Won *et al.*, 2011). In *Arabidopsis*, there are several genes coding for TAA and YUCCA proteins (*TAA1*, *TAR1*, *TAR2*; *YUC1-11*). However, these genes are not ubiquitously expressed. Instead, each has a specific expression pattern across different plant tissues and developmental stages, attesting to the importance of tightly regulating auxin biosynthesis during plant development (Kasahara, 2016; Zhao, 2018).

Fine-tuning of the spatial distribution of auxin, for example through the establishment of auxin gradients, is often required for a wide range of developmental processes. This can be achieved through control of auxin biosynthesis, but also of its intercellular transport. Transport of auxin can be described through the chemiosmotic model initially formulated by Rubery, Sheldrake, and Raven, in the 1970s (Rubery & Sheldrake, 1974; Raven, 1975), and further elaborated by Goldsmith (1977). This model can be explained by two factors: i) passive diffusion of protonated IAA (IAAH) across the cell membrane, due to pH differences between the apoplast and cytoplasm; and ii) influx and efflux of anionic IAA (IAA^-) via active transport (Goldsmith, 1977; Petrasek & Friml, 2009). Recent studies have elucidated the details of this model, with particular emphasis on uncovering the role of auxin transporter proteins. Presently, there are three types of protein families known to transport auxin: the AUXIN1/LIKE-AUXIN1 (AUX/LAX) importers (Bennett *et al.*, 1996; Yang *et al.*, 2006; Swarup *et al.*, 2008), the ATP-BINDING CASSETTE/P-GLYCOPROTEIN (ABCB/PGP) exporters (Noh *et al.*, 2001), and the PIN-FORMED/PIN-LIKES (PIN/PILS) transporters (Petrásek *et al.*, 2006; Barbez *et al.*, 2012). The latter are involved in export from the cell (PINs), but also in intracellular compartmentalisation of auxin (PILS), which is hypothesized to allow for a finer control of the intracellular levels of this hormone (Barbez *et al.*, 2012). PIN exporters are typically polarly distributed in the cell, showing an apical or basal preference depending on the cell type (Wisniewska *et al.*, 2006). Consequently, the polar positioning of PINs leads to auxin flow in a coordinated direction within a set of cells (Wisniewska *et al.*, 2006). This directionality is important to establish auxin maxima and minima within tissues and organs, which in turn leads to the modulation of several developmental processes, such as the establishment of morphogenetic fields (*e.g.* in root, embryo, or shoot apical meristem patterning) (Petrasek & Friml, 2009).

The developmental significance of establishing auxin gradients can be explained by the fact that the auxin response machinery is sensitive to changes in the cellular concentration of auxin (Weijers & Wagner, 2016). AUXIN RESPONSE FACTORS (ARF) are TFs capable of recognising specific DNA sequences – auxin response elements (AuxRE) (Weijers & Wagner, 2016). In the presence of low concentrations of auxin, ARFs form complexes with the transcriptional coregulators Aux/IAAs (AUXIN/INDOLE-3-ACETIC-ACID). This impairs the transcriptional activity of ARFs, and consequently inhibits auxin responses (Enders & Strader, 2015; Weijers & Wagner, 2016). In contrast, when auxin concentration increases, Aux/IAA proteins interact preferentially with the $SCF^{TIR1/AFB}$ complex. The SCF complex (Skp, Cullin, F-box), in conjunction with the F-box proteins TRANSPORT INHIBITOR RESISTANT1 (TIR1) or AUXIN SIGNALLING F-BOX (AFB), polyubiquitinates Aux/IAA proteins, leading to their degradation by the 26S proteasome (Enders & Strader, 2015; Weijers & Wagner, 2016). The degradation of Aux/IAAs renders ARF proteins free to transcriptionally control target genes, thus initiating the auxin response (Enders & Strader, 2015; Weijers & Wagner, 2016).

Although the mechanism of auxin signalling seems rather straightforward, a considerable amount of studies have shown it to be more nuanced than initially thought (Abel & Theologis, 2010). For example, the affinity of Aux/IAA proteins to the $SCF^{TIR1/AFB}$ complex varies among Aux/IAA proteins, and also depends on the intracellular concentration and on the type of auxin present in the cell (Enders & Strader, 2015; Weijers & Wagner, 2016). Furthermore, several dozen ARFs are encoded in the *Arabidopsis* genome, and while the majority is reported to act as transcriptional activators, some can act as transcriptional repressors (Guilfoyle & Hagen, 2007; Weijers & Wagner, 2016). Besides this, ARFs, alone or complexed with Aux/IAAs, are reported to interact with additional proteins, such as chromatin remodellers (Yamamuro *et al.*, 2016), and other TFs. Among these, is the type II MADS-box TF SEPALLATA3 (SEP3) (Smaczniak *et al.*, 2012; Weijers & Wagner, 2016). This suggests that the auxin signalling system is able to recruit other players in order to increase the plasticity and range of responses the system is able to trigger.

1.3.2 Fertilisation leads to increased auxin levels in seeds

Among all the developmental processes controlled by auxin, are those related to reproductive development, of which embryogenesis is an important example: changes in auxin biosynthesis, transport, or signalling lead to dramatic developmental phenotypes and ultimately, to unviable embryos (Möller & Weijers, 2009). Nevertheless, the role of auxin in the development of the

remaining seed components – endosperm and seed coat – remains unexplored in *Arabidopsis*.

It is interesting to note that auxin-related genes, namely the biosynthesis genes *TAA1, TAR1,* and *YUC10,* are expressed in the fertilised endosperm, and subject to genomic imprinting in *Arabidopsis*, exhibiting paternally-derived expression **(section 1.4.3)** (Hsieh *et al.*, 2011; Wolff *et al.*, 2011; Pignatta *et al.*, 2014). Even though imprinted genes tend to show low conservation among species (Waters *et al.*, 2013; Hatorangan *et al.*, 2016), *TAA1* is imprinted in the *Arabidopsis* genus, in *Capsella rubella,* and in maize (Waters *et al.*, 2013; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016), while *YUC10* orthologues are imprinted in rice, and maize (Luo *et al.*, 2011; Waters *et al.*, 2013). This suggests that auxin production upon fertilisation might be a conserved feature among flowering plants. Notwithstanding, the functional relevance of coupling auxin biosynthesis to fertilisation via paternal expression of auxin biosynthesis genes, has not yet been explored. Clarifying this was one of the goals of this thesis **(paper I, II and III)**.

In line with the aforementioned pattern of expression of auxin biosynthesis genes, reporters for auxin signalling show that auxin activity increases in seeds following the fertilisation event (Dorcey *et al.*, 2009). One of the hypotheses that could explain this, is that this hormone is required as a fertilisation-derived signal that promotes seed development. The fact that there are no symplastic connections between endosperm and seed coat **(section 1.1.6)**, coupled with the ability of auxin to be actively transported across cells, make this hormone a promising candidate for the endosperm-derived signal that kick-starts seed coat growth. Testing this hypothesis was a central part of this thesis **(see paper II)**.

1.4 Epigenetics of reproductive development

The DNA molecule carries a vast amount of information in its sequence, but an additional a layer of information is added by the conformation of the chromatin, a complex composed of DNA, histones, and additional scaffold proteins. The basic unit of the chromatin is the nucleosome, which consists of 146 bp of DNA wrapped around a core of 8 histones (Zhou *et al.*, 2019). During the past decades, several chemical modifications, affecting either the DNA or histones, have been implicated in modifying the structure of these units. These modifications affect the way the information contained in the DNA sequence is used, mainly by influencing gene expression and genome stability, and constitute the basis of epigenetic information (Feng & Jacobsen, 2011; Zhang *et al.*, 2018). In this chapter I will give a brief overview of epigenetic modifications that are relevant in the context of plant reproduction, more specifically those important for the understanding of genomic imprinting.

DNA methylation

In eukaryotes, DNA can be methylated through the addition of a methyl group to cytosine bases, and in plants DNA methylation can occur in three different sequence contexts – symmetric CG and CHG, and asymmetric CHH (where H corresponds to A, T, or G) (Law & Jacobsen, 2010). DNA methylation is mainly directed at transposable elements (TEs) (Zhang *et al.*, 2006), but it can also occur in promoter regions, where it is often linked to tissue-specific repression of transcription, as well as in gene bodies, whose impact remains somewhat elusive (Bewick & Schmitz, 2017; Zilberman, 2017). *De novo* deposition of DNA methylation largely relies on the RNA-directed DNA methylation pathway (RdDM), whose homology-based strategy allows to take advantage of the repetitive nature of its target sequences (*i.e.* TEs) to efficiently silence them across the genome. A brief description of the RdDM pathway is presented below and is based on the comprehensive reviews published by Law & Jacobsen, (2010), Matzke & Mosher, (2014), Zhang *et al.*, 2018, and references therein. In this pathway, a specific plant RNA polymerase - RNA POLYMERASE IV (Pol IV) - mediates the transcription of single stranded RNA from TEs and related repeat sequences. These transcripts are converted into double stranded RNA by the RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), and then diced into small 24-nt interfering RNAs (siRNAs) by DICER-LIKE (DCL) proteins. These siRNAs are subsequently loaded into ARGONAUTE (AGO) proteins, which are

guided to target loci. In these loci, transcription by RNA POLYMERASE V (Pol V) produces nascent transcripts which share complementarity with the siRNAs loaded in AGO proteins. The AGO proteins then interact with the DNA methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), which catalyses *de novo* methylation of target loci in all three cytosine contexts.

An additional non-canonical RdDM pathway exists, and is instead dependent on transcription by RNA POLYMERASE II (Pol II), RDR6-mediated double stranded RNA synthesis, and a different set of AGO proteins (Panda & Slotkin, 2013). In this pathway, a distinct class of siRNAs is produced (*i.e.* 21-nt), which are subsequently integrated in the canonical RdDM pathway. This non-canonical RdDM pathway is hypothesized to take advantage of TE transcription to kick-start the silencing of new TE insertions (Panda & Slotkin, 2013).

Maintenance of DNA methylation is required during DNA replication and cell division, and several methyltransferases are responsible for this task. In the CG context, hemi-methylated cytosine is recognised by VARIANT IN METHYLATION (VIM) proteins, which are hypothesized to recruit METHYLTRANSFERASE 1 (MET1) to catalyse cytosine methylation in the newly-formed DNA strand (Woo *et al.*, 2008). CHG methylation often co-occurs with di-methylation of lysine 9 on histone H3 (H3K9me2). This histone mark is recognized and bound by the DNA methyltransferase CHROMOMETHYLASE3 (CMT3), facilitating maintenance of this DNA methylation context (Du *et al.*, 2012). On the other hand, some H3K9 methyltransferase proteins, such as SUPRESSOR OF VARIEGATION 3-9 HOMOLOGS (SUVHs), recognize CHG methylation. Lack of one of these marks leads to absence of the other, thus pointing to CHG being maintained by a self-reinforced loop, which is H3K9me2-dependent (Johnson *et al.*, 2007). While symmetric methylation contexts facilitate the recognition of hemi-methylated sites, the maintenance of CHH methylation is carried out by distinct pathways, depending on the genomic region. In RdDM target sites, CHH is maintained by the action of DRM2, as described above. While in pericentromeric and centromeric heterochromatic regions, from where RdDM is excluded, the chromatin remodeller DECREASED DNA METHYLATION 1 (DDM1) enforces silencing through the action of CHROMOMETHYLASE 2 (CMT2), which deposits CHH independently of the RdDM pathway (Zemach *et al.*, 2013).

Developmental responses require modulation of DNA methylation levels. Thus, coordination between DNA methylation deposition, maintenance, and demethylation is required to achieve this modulation. DNA demethylation is carried out by a set of DNA glycosylases, which are able to excise methylated

cytosine. These proteins are DEMETER (DME), REACTIVATION OF SILENCING 1 (ROS1), and DEMETER-LIKE 2-3 (DML2-3) (Law & Jacobsen, 2010). These four enzymes demethylate cytosine in any sequence context, and while ROS1 and DML2-3 are active throughout most stages of plant development (Gong *et al.*, 2002; Ortega-Galisteo *et al.*, 2008), DME activity is restricted to gametes and proliferating cells (Choi *et al.*, 2002; Kim *et al.*, 2008; Schoft *et al.*, 2011), and is required for genomic imprinting **(section 1.4.1)**.

Polycomb Group-mediated histone modifications

Contrary to the histone modification H3K9me2, which is mainly associated with TEs and CHG methylation, tri-methylation of lysine 27 on histone H3 (H3K27me3) mostly targets genes, and is negatively correlated with the presence of DNA methylation (Weinhofer *et al.*, 2010; Roudier *et al.*, 2011; Deleris *et al.*, 2012). This histone modification is established by Polycomb Group proteins (PcG), it leads to inhibition of transcription and compaction of the chromatin, and has a critical role in controlling developmental transitions in plants and animals (Mozgova & Hennig, 2015). The importance of this histone modification was first revealed in *Drosophila*, where the activity of several key homeotic genes is under the control of PcG proteins (Jürgens, 1985). These proteins are organized in multimeric complexes, designated POLYCOMB REPRESSIVE COMPLEX (PRC). These complexes can generally be divided in PRC1- and PRC2-type, with PRC2 being responsible for the establishment of H3K27me3. In *Arabidopsis* there are three PRC2 complexes, with distinct components, as well as distinct functions within plant development (Mozgova *et al.*, 2015). The catalytic subunits of PRC2 are the histone lysine methyltransferases MEDEA (MEA), SWINGER (SWN) and CURLY LEAF (CLF), and the remaining PRC2 members are: FERTILISATION INDEPENDENT SEED 2 (FIS2), FERTILISATION INDEPENDENT ENDOSPERM (FIE), MULTICOPY SUPRESSOR OF IRA (MSI1), EMBRYONIC FLOWER 2 (EMF2), and VERNALIZATION 2 (VRN2) (Mozgova & Hennig, 2015). Together, these components form three different complexes – FIS-PRC2 (composed by FIS2, MEA/SWN, FIE, and MSI1); EMF-PRC2 (EMF2, CLF/SWN, FIE, and MSI1); and VRN-PRC2 (VRN2, CLF/SWN, FIE, and MSI1) (Mozgova & Hennig, 2015).

PRC2 complexes act in a wide range of processes, namely vegetative development, vernalisation, gametogenesis, and seed development (Mozgova *et al.*, 2015). During gametogenesis and seed development, FIS-PRC2 is required to establish genomic imprinting **(section 1.4.1)**, as well as to supress the development of seeds without fertilisation, along with EMF-PRC2 and VRN-PRC2. This is illustrated by the phenotypes of mutants for gametophytic FIS-

PRC2 subunits, such as *mea*, *fis2* and *fie*, where autonomous division of the central cell generates a purely maternal endosperm that can proliferate to varying degrees (Chaudhury *et al.*, 1997; Kiyosue *et al.*, 1999; Ohad *et al.*, 1999). In *msi1* mutants, parthenogenic embryos are also formed occasionally (Guitton & Berger, 2005). Similarly, in sporophytic VRN-PRC2 and EMF-PRC2 mutants (*fie*, *vrn2* and *emf2*), the maternal integuments spontaneously differentiate into seed coat tissue (Roszak & Köhler, 2011). These observations indicate that PcG proteins exert an effective repression on seed development, and are thus required to couple the initiation of seed growth with the fertilisation event. The phenotypes of these mutants mimic the apomictic seeds produced in several plant species (Bicknell & Koltunow, 2004), and although the autonomous seed phenotypes of PRC2 mutants have been known for many years, the molecular mechanisms that lead to the development of these fertilization-independent seeds remain undiscovered. Elucidating some of these pathways was one of the aims of **papers I and II**.

1.4.1 Genomic imprinting is established via the asymmetric deposition of epigenetic modifications on parental genomes

Genomic imprinting can be described as an epigenetic phenomenon leading to the differential expression of a given allele, depending on its parental origin. Thus, a gene can be preferentially expressed from the maternally-, or from the paternally-derived allele. These genes are termed maternally-expressed genes (MEGs), or paternally-expressed genes (PEGs), respectively. Imprinting evolved independently in mammals and flowering plants, and was first identified in maize, where maternal inheritance of the *R* allele leads to a darkly-pigmented kernel, but paternal inheritance of this allele leads to a mottled kernel (Kermicle, 1970). In plants, imprinting takes place in the endosperm, and evidence for its occurrence in the embryo are scarce (Köhler *et al.*, 2012; Gehring, 2013).

Genomic imprinting can be largely explained through the action of DNA methylation and histone modifications. These marks are differentially deposited in female and male genomes during gametogenesis, and inherited in the endosperm, leading to parent-of-origin-specific gene expression. Imprinting often relies on the action of DME: in the central cell of the female gametophyte, DME promotes an extensive demethylation of the maternal genome (Hsieh *et al.*, 2009), which preferentially targets TEs and short repetitive sequences (Gehring *et al.*, 2009; Ibarra *et al.*, 2012). Similarly, in the vegetative nucleus of the male gametophyte, loss of DDM1 coupled with the presence of DME, ROS1, and DML2-3 (Slotkin *et al.*, 2009; Schoft *et al.*, 2011; Calarco *et al.*, 2012), lead to a global loss of DNA methylation (Ibarra *et al.*, 2012). DNA demethylation

in the vegetative nucleus coincides with reactivation of TEs and production of siRNAs (Slotkin *et al.*, 2009), which are potentially transported to the adjacent sperm cells to reinforce TE silencing (Martínez *et al.*, 2016). In agreement with this, these loci are hypermethylated in the sperm cells when compared to the vegetative nucleus (Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Imprinted genes are often flanked by TEs (Wolff *et al.*, 2011; Pignatta *et al.*, 2014; Hatorangan *et al.*, 2016). Consequently, DME-dependent demethylation of TEs and repetitive sequences can have an indirect effect in the expression of these genes (Gehring *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Together, these data point to a scenario where in the endosperm, a given locus is maternally inherited as hypomethylated and paternally inherited as hypermethylated **(Figure 3)**.

This pattern of DNA methylation is important for the expression of imprinted genes, either directly or indirectly. Cases of direct influence can be exemplified by a subset of MEGs, where maternal hypomethylation leads to expression of this allele, and paternal hypermethylation leads to repression (Kinoshita, 2004; Gehring *et al.*, 2006; Jullien *et al.*, 2006; Hsieh *et al.*, 2011). In many instances however, DNA methylation has an indirect role on imprinting patterns that is dependent on the interaction with H3K27me3. FIS-PRC2 is active in the central cell and H3K27me3 deposition in the maternal genome occurs at hypomethylated DME loci (Moreno-Romero *et al.*, 2016). In addition, FIS-PRC2 is also present in the developing endosperm, leading to the possibility of *de novo* H3K27me3 establishment in the paternal genome. Nevertheless, the exclusion of PRC2 from DNA methylated regions (Weinhofer *et al.*, 2010; Deleris *et al.*, 2012), likely prevents the methylated paternal alleles to be targeted. Paternal-specific gene expression can largely be explained by this pattern: while the maternal allele is silenced by H3K27me3, the paternal allele is protected from such silencing by DNA methylation **(Figure 3)** (Gehring & Satyaki, 2017). This is supported by the observation that paternal inheritance of hypomethylated alleles, as a consequence of a *met1* mutation, causes repression of paternal alleles of PEGs (Hsieh *et al.*, 2011). Furthermore, maternal mutations in components of FIS-PRC2 lead to loss of imprinting at PEG loci, while also affecting imprinting and expression level of some MEGs (Hsieh *et al.*, 2011), thus suggesting that H3K27me3 is required to establish and regulate both paternal- and maternal-specific gene expression **(Figure 3)**.

Interestingly, a recent study found that not only the presence of H3K27me3 in maternal alleles, but also of H3K9me2 and CHG methylation, can be used as predictive factors for the identification of PEGs (Moreno-Romero *et al.*, 2019), **(Figure 3)**. Nevertheless, the mechanistic relevance for the unexpected coexistence of these marks remains to be explored. Despite the fact that similarities exist among the imprinting mechanisms leading to MEGs, these

seem to be more variable than those leading to PEGs. Consequently, to this point, no predictive model based on epigenetic marks could be ascertained for the majority of MEGs.

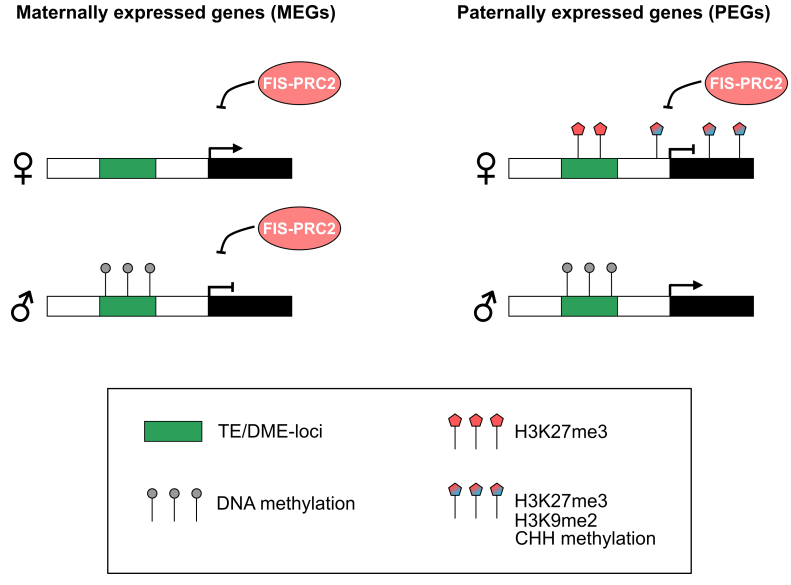

Figure 3. **Epigenetic mechanisms regulating imprinted gene expression.** Epigenetic modifications associated with maternally and paternally expressed genes are shown. For simplicity purposes, only one maternal allele is represented.

The transcriptional output of imprinted genes reflects the asymmetric epigenetic landscapes at both parental genomes, showing that a cross-talk between epigenetics and transcription must occur at these loci. Presently, there is no knowledge on which TFs control the expression of imprinted genes, neither on which are the promoter binding sites used by these TFs. Consequently, it is not possible to directly infer how the epigenetic landscape at imprinted genes impacts the accessibility of TFs to their binding sites. One of the main goals of this thesis was to explore and better understand the transcriptional regulation of imprinted genes, as well as the interplay between transcriptional and epigenetic regulation at these genes **(paper IV)**.

1.4.2 Transposable elements are involved in the establishment of imprinting

Most efforts in understanding genomic imprinting have focused on understanding the epigenetic landscapes that lead to parental-specific gene

expression. Strikingly, the presence of epigenetic marks in the vicinity of imprinted genes often correlates with the presence of TEs: these elements, rather than the imprinted genes themselves, are associated with DME-derived differentially methylated regions (DMRs), both in female and male gametes (Gehring *et al.*, 2009; Ibarra *et al.*, 2012). In pollen, siRNAs derived from TEs are correlated with the presence of DMRs near imprinted genes, especially MEGs (Calarco *et al.*, 2012); and in the endosperm, siRNAs are associated with TEs in the vicinity of imprinted genes (Pignatta *et al.*, 2014). Furthermore, in some Brassicaceae, imprinted genes show an enrichment of TEs in their neighbouring regions (Wolff *et al.*, 2011; Pignatta *et al.*, 2014; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016), with specific TE superfamilies, such as RC/Helitron and DNA/MuDR, being overrepresented (Wolff *et al.*, 2011; Hatorangan *et al.*, 2016). Additionally, in some species of the *Capsella* genus, increased genomic TE content correlates with an increase in the number of imprinted genes, more specifically of PEGs (Lafon-Placette *et al.*, 2018).

In line with this, specific epigenetic signatures associated with the presence of TEs have been directly implicated in parentally biased expression of several imprinted genes (Pignatta *et al.*, 2014). One of those genes codes for the type I MADS-box TF PHE1. *PHE1* is a PEG, and its imprinting depends on the presence of repeat sequences at the 3' flanking region. In the paternal allele of *PHE1* these repeats are DNA methylated, likely preventing deposition of H3K27me3 by FIS-PRC2. On the other hand, these regions are demethylated in the maternal allele, leading to H3K27me3 deposition and silencing of this allele (Makarevich *et al.*, 2008; Villar *et al.*, 2009).

TE insertions are known to mobilize *cis*-regulatory sequences that have the potential to generate novel promoters, and affect neighbouring gene expression (Lisch, 2012). This could be pertinent when considering the enrichment of RC/Helitrons in the vicinity of imprinted genes, since these TEs exhibit a mode of transposition that is associated with capture of host-derived sequences (Kapitonov & Jurka, 2007). Notwithstanding, the hypothesis that TEs could provide *cis*-regulatory sequences that facilitate transcription of imprinted genes remains untested – a knowledge gap that **paper IV** aimed to address.

The data presented here constitute a strong body of evidence that implicates TEs in generating specific epigenetic landscapes, which are conducive for the establishment of imprinting. But why are TEs targeted by specific epigenetic modifications? And why are these elements found abundantly in the vicinity of imprinted genes? To answer these questions, one must reflect on the intrinsic properties of TEs and on the relationship with their host. TEs are often seen as selfish elements that can propagate in the host genome, with potentially critical consequences. The ability of these elements to cut or copy themselves and insert

into new locations is an obvious threat to genome stability (Austin & Trivers, 2009). New insertions can be a source of mutagenesis by disrupting genes or their regulatory sequences, and the repetitive nature of TEs can often lead to undesired chromosomal rearrangements (Austin & Trivers, 2009). Nevertheless, virtually all eukaryotic species contain TEs, and in some cases these elements constitute significant fractions of the genome (Biémont & Vieira, 2006; Chénais *et al.*, 2012). This is coupled with the existence of intricate defence strategies set up by the host, in order to epigenetically silence those TEs, with the goal of preventing their transposition (Fultz *et al.*, 2015). Such strategies largely depend on the combined action of histone modifications and DNA methylation, with the latter being primarily enforced by the RdDM pathway through its ability to recognize TEs and derived repeat sequences (Panda & Slotkin, 2013).

Given the wide range of potential deleterious effects of TE insertions, their frequent presence near imprinted genes seems paradoxical, and their maintenance likely depends on several factors: i) the insertion and its effects on neighbouring gene activity contribute positively to the fitness of the host and are therefore under positive selection, or ii) the affected gene *per se*, or in combination with any effects caused by the TE insertion, has no strong impact for host fitness and is therefore under relaxed selection. Hence, it can be envisioned that imprinting itself is not relevant for plant fitness, and that it simply occurs as a consequence of TE-derived epigenetic modifications, enforced as a host defence against TEs (Gehring, 2013). In the next section I will reflect on this and other hypotheses that attempt to explain the evolution of genomic imprinting, while further exploring the role of imprinting in the context of plant reproduction.

1.4.3 Imprinting is relevant for endosperm development, especially in the context of reproductive barriers

There are several theories that aim to explain the evolutionary origin and maintenance of genomic imprinting (Patten *et al.*, 2014; Rodrigues & Zilberman, 2015). When considering these theories, it is important to bear in mind the nature of the endosperm as a terminal tissue, responsible for the allocation of resources to the progeny – the embryo. Additionally, the biparental composition of the endosperm should also be taken into account: even though this is not the case in *Arabidopsis thaliana*, many angiosperms are outcrossers, meaning the parents of a given seed are two distinct individuals. In the previous section, I discussed how TE insertions can be regarded as an important mechanism to generate imprinted genes. Here, I will discuss the kinship theory (Haig & Westoby, 1989),

which provides further theoretical framework for the understanding of the evolution of imprinting.

Given that a female can simultaneously bear progeny derived from several different males, and that this progeny consists of half-siblings, the kinship theory proposes that maternal and paternal interests are distinct: while the mother promotes equal distribution of resources to all progeny, the father diverts as much resources as possible to its own progeny, in order to confer advantage over its half-siblings. These distinct interests are then reflected in the development of parentally-biased gene expression. Some experimental evidence has provided support to the kinship theory: for instance, the maternal and paternal genomes exert different influences in seed development, with the paternal genome promoting growth of the endosperm, and the maternal genome showing the opposite effect (Scott *et al.*, 1998). Furthermore, analysis of imprinted genes revealed that MEGs and PEGs have distinct functions – while MEGs are enriched for genes involved in primary metabolism and for transcription factors, PEGs are enriched for chromatin regulators (Wolff *et al.*, 2011; Hatorangan *et al.*, 2016). Similarly, in mammals, MEGs are associated with growth inhibition, while PEGs are associated with growth stimulation (Haig, 2004). This might suggest that, as predicted by the kinship theory, antagonistic selective pressures act on MEGs and PEGs.

This theory is based on the assumption that genomic imprinting has an impact on the fitness of the produced offspring – the seed, in case of flowering plants. If this is the case, it is expected that loss of imprinting has deleterious effects in seed development. Interestingly, mutations of single imprinted genes usually lead to no aberrant phenotypes (Bai & Settles, 2014; Wolff *et al.*, 2015), with the exception of epigenetic mutants where imprinting is lost in a more systematic manner. Mutating the maternally-expressed FIS-PRC2 components MEA and FIS2 leads to over-proliferation of the endosperm, lack of cellularisation and, consequently, to seed abortion (Luo *et al.*, 2000). This is accompanied by the transcriptional activation of the maternal alleles of many PEGs (Hsieh *et al.*, 2011), likely due to the absence of maternal repression by FIS-PRC2. These phenotypes show that global overexpression of PEGs leads to increased growth, pointing to a paternal interest in proliferation, as predicted by the kinship theory. In parallel, they also demonstrate that the presence of MEGs, such as FIS-PRC2 components, is required to counteract the paternal effect. Additional studies focusing on the selective pressures acting on imprinted genes suggest that a subset of these genes are under positive selection, especially those that are conserved across several species (Spillane *et al.*, 2007; Wolff *et al.*, 2011; Waters *et al.*, 2013; Hatorangan *et al.*, 2016; Tuteja *et al.*, 2019). Together, these data show that imprinting is required for correct endosperm development, and

repeal the idea that this phenomenon is a mere by-product of TE-derived epigenetic landscapes that would have no impact for plant fitness. Still, further work characterising the functions of imprinted genes is required, in order to gain a better understanding of the effects of imprinting in seed development.

It is also important to note that many studies focusing on identifying imprinted genes, or exploring their function in seed development through mutagenesis, have been performed in diploid selfing species, such as *Arabidopsis thaliana* (Wolff *et al.*, 2011; Bai & Settles, 2014; Pignatta *et al.*, 2014). However, when considering the kinship theory, it is expected that the functional relevance of imprinting is more substantial in situations where parental conflict is more pronounced. This could be the case in species where genetic diversity is high, such as in outcrossing plants (Brandvain & Haig, 2005). In these species, genetic differences between parents are expected to be higher than in selfing species and could arise, among others, from differences in genomic TE composition (Boutin *et al.*, 2012). Another scenario where parental conflict can be manifested is in interploidy crosses: even in a selfing species, increasing genomic dosage of one of the parents leads to unbalanced parental composition in the endosperm (Haig & Westoby, 1991).

In fact, seed abortion derived from interploidy or interspecies crosses has often been associated with deregulation of imprinted gene function (Josefsson *et al.*, 2006; Walia *et al.*, 2009; Kirkbride *et al.*, 2015; Wolff *et al.*, 2015; Florez-Rueda *et al.*, 2016; Lafon-Placette *et al.*, 2018). This shows that, besides having relevance for endosperm development in the context of low parental conflict (*i.e.* selfing diploid species), imprinting can take part in establishing endosperm-based reproductive barriers. In the next section I will introduce some of these reproductive barriers, giving special attention to interploidy crosses, and explaining the molecular mechanisms driving them.

1.5 Endosperm-based reproductive barriers

The biological concept of species, as proposed by Ernst Mayr, relies on the presence of reproductive isolation: individuals of different species cannot interbreed (Mayr, 1996). Reproductive barriers can be manifested at different levels, and at different stages of an organism's development. In general, they can be defined as pre-zygotic, or post-zygotic, depending on whether they occur before or after fertilisation (Seehausen *et al.*, 2014). In plants, different seasonal flowering times, or pollen-pistil incompatibility are examples of pre-zygotic barriers; while seed inviability, or F1 hybrid sterility are examples of post-zygotic barriers (Baack *et al.*, 2015). In this section, I will focus on post-zygotic reproductive barriers taking place in hybrid seeds.

These hybridization barriers have been studied by breeders for decades, who observed that they can occur in crosses between different species (interspecies crosses), or in crosses between plants of different ploidies (interploidy crosses) (Brink & Cooper, 1947; Håkansson, 1953; Johnston & Hanneman, 1982; Lin, 1984; Scott *et al.*, 1998; Sekine *et al.*, 2013; Rebernig *et al.*, 2015; Lafon-Placette *et al.*, 2017; Roth *et al.*, 2018). Both types of abortive crosses are present in a wide range of angiosperms (*e.g.* cereals, Solanaceae, Brassicaceae, among others), and exhibit a similar phenotype: seed abortion is usually accompanied by alterations in embryo development, as well as aberrant endosperm proliferation, and impaired endosperm differentiation. Interestingly, the embryo can be rescued when removed from the seed and grown in nourishing media, pointing to the endosperm as the cause for seed abortion (Brink & Cooper, 1947; Hehenberger *et al.*, 2012; Rebernig *et al.*, 2015).

Another observation derived from these studies is that seed abortion can be explained by unbalanced maternal and paternal contributions to the endosperm (Brink & Cooper, 1947). These authors argued that there is a quantitative balance between the two parental genomes, and when this is not met, the endosperm aborts. For example, in the case of *Arabidopsis thaliana*, the endosperm is composed of 2 maternal and 1 paternal genome copies **(section 1.1)**. Increasing the ploidy of one of the parents leads to unbalanced endosperm composition, and results in seed abortion (Scott *et al.*, 1998). This phenomenon is called the triploid block, alluding to the triploid nature of the embryo **(Table 1)** (Marks, 1996). Remarkably, the seed phenotypes vary depending on which parent contributes with the highest genome dosage: with increased maternal dosage there is less endosperm proliferation and early cellularisation; while increased paternal dosage leads to increased proliferation and delayed cellularisation (Lafon-Placette & Köhler, 2016). The same phenomenon can be observed in crosses between different species: even when both parents have the same ploidy,

Table 1. **Ploidies of different seed components derived from interploidy crosses.** The ploidy of the maternal and paternal plants are indicated for each cross (eg. 2x ♀ × 2x ♂ in the balanced cross, maternal plant always indicated first). ec: egg cell, cc: central cell, sc: sperm cells.

	Type of cross	Female gametes	Male gametes	Embryo	Endosperm
	Balanced cross 2x ♀ × 2x ♂	ec: x cc: 2x	sc: x	2x	3x
Triploid block	*Maternal excess* 4x ♀ × 2x ♂	ec: 2x cc: 4x	sc: x	3x	5x
Triploid block	*Paternal excess* 2x ♀ × 4x ♂	ec: x cc: 2x	sc: 2x	3x	4x

maternal excess-like and paternal excess-like reciprocal defects can be observed, which are similar to the ones seen in interploidy hybrid seeds (Brink & Cooper, 1947; Sekine *et al.*, 2013; Rebernig *et al.*, 2015; Lafon-Placette *et al.*, 2017; Roth *et al.*, 2018). Interestingly, such defects can be alleviated by modulating the ploidy of one of the parents, leading to the production of viable hybrid seeds (Johnston & Hanneman, 1982; Lafon-Placette *et al.*, 2017; Tonosaki *et al.*, 2018).

This led to the hypothesis that species have different intrinsic genome dosages even if having the same number of chromosome sets, theoretically framed as the concept of Endosperm Balance Number (EBN) (Johnston *et al.*, 1980). The initial goal of calculating a theoretical EBN was to predict the outcome of hybridizations between different *Solanum* species. For each analysed species, an EBN value was calculated based on the outcome of different hybrid crosses, as well as how this outcome would change in response to changes in the parent's ploidy (Johnston *et al.*, 1980). Despite the fact that EBN can successfully predict the result of these hybridizations, the exact factors determining the EBN of a given species still remain elusive. Several authors have proposed that genomic imprinting could be one of those factors. Considering the predictions of the kinship theory **(section 1.4.3)**, imprinting would be dependent on the level of parental conflict present in a given species, and this, in turn, would be dependent on the reproductive strategy of that species (*i.e.* selfer or outcrosser) (Brandvain & Haig, 2005). Therefore, the EBN could be derived from a direct measure of three factors: reproductive strategy, parental conflict level, and imprinting, which in turn are all interconnected (Lafon-Placette & Köhler, 2016). In parallel, Dilkes and Comai (2004), propose a more general hypothesis in which any differences in maternal and paternal gene dosage would drive differences in EBN. Assuming endosperm developmental

pathways require a certain stoichiometry of protein complexes, differential parental dosage of the genes encoding these complexes – either achieved through imprinting, differences in gene copy number, or differences in gene expression levels – could contribute to unbalanced stoichiometry and consequently, failure of endosperm development.

In plants, polyploidization is a common phenomenon, which is thought to be an important driver of speciation, namely of sympatric speciation (Otto & Whitton, 2002). Thus, postzygotic reproductive barriers, especially those that are endosperm-based could have a significant impact in facilitating novel speciation events. Moreover, reproductive isolation is often observed in breeding programs aimed at producing hybrids, leaving embryo rescue as the only available tool to produce viable seeds (Tonosaki *et al.*, 2016). Nevertheless, embryo rescue is a complex process that often requires extensive optimisation times (Haslam & Yeung, 2011). For all these reasons, it is imperative to better understand the molecular mechanisms underlying endosperm-based hybridization barriers. Part of this thesis focused on further characterising these mechanisms, specifically those involved in establishing the triploid block **(paper III, and IV)**.

1.5.1 The triploid block is associated with deregulation of imprinted genes and MADS-box transcription factors

To better understand the molecular mechanisms driving the seed abortion phenotype associated with interploidy crosses, several studies have made significant efforts in characterizing gene expression profiles of whole seeds and of endosperm derived from these crosses. In these studies, a vast array of genes is found to be deregulated, such as genes related to chromatin modifications, cell proliferation, hormonal pathways, and transcriptional regulation (Tiwari *et al.*, 2010; Stoute *et al.*, 2012). Among these, several type I MADS-box TFs, including *PHE1* and *AGL62,* are consistently identified in different studies as highly deregulated (Erilova *et al.*, 2009; Tiwari *et al.*, 2010; Lu *et al.*, 2012; Kradolfer *et al.*, 2013). The common feature in these observations is that the level and dynamics of expression of type I MADS-box genes relates to the outcome of endosperm cellularisation: in paternal excess 3x seeds, where endosperm cellularisation is delayed, the expression of these TFs is increased and prolonged in time, when compared to a 2x seed; the reverse happens in seeds where endosperm cellularisation occurs prematurely, (*i.e.* in maternal excess crosses) **(Figure 4)** (Erilova *et al.*, 2009; Tiwari *et al.*, 2010; Lu *et al.*, 2012; Schatlowski *et al.*, 2012; Stoute *et al.*, 2012). Together, these observations have

led to the proposal that type I MADS-box TFs are likely to be negative regulators of endosperm cellularisation, and that their deregulation in interploidy crosses is causal to the observed developmental defects. Nevertheless, a detailed mechanistic explanation on how these TFs could control endosperm cellularisation has not yet been provided. One of the goals of this thesis was to uncover the biological role of these TFs, namely in the context of interploidy crosses **(paper IV)**.

Similarly to type I MADS-box genes, imprinted genes, especially PEGs, are upregulated in paternal excess 3x seeds, as predicted by proponents of the kinship theory (Haig & Westoby, 1991; Gutierrez-Marcos *et al.*, 2003) **(Figure 4)**. Furthermore, mutating single PEGs can prevent the triploid block phenotype to varying degrees, which implicates them in establishing endosperm-based hybridization barriers (Kradolfer *et al.*, 2013; Wolff *et al.*, 2015; Jiang *et al.*, 2017; Martinez *et al.*, 2018; Wang *et al.*, 2018). Remarkably, some MADS-box TFs that are deregulated in 3x seeds are also classified as PEGs. Nevertheless, their specific impact on the 3x seed phenotype has not yet been assessed.

Ultimately, by providing a better understanding of the biological role of MADS-box TFs in the endosperm, as well as providing insights into how imprinted genes are regulated, this thesis aims at understanding if the deregulation of all these genes can explain the seed abortion phenotypes observed in 3x seeds. This information will also prove useful in understanding other postzygotic hybridization barriers (*i.e.* interspecies), as these may share the

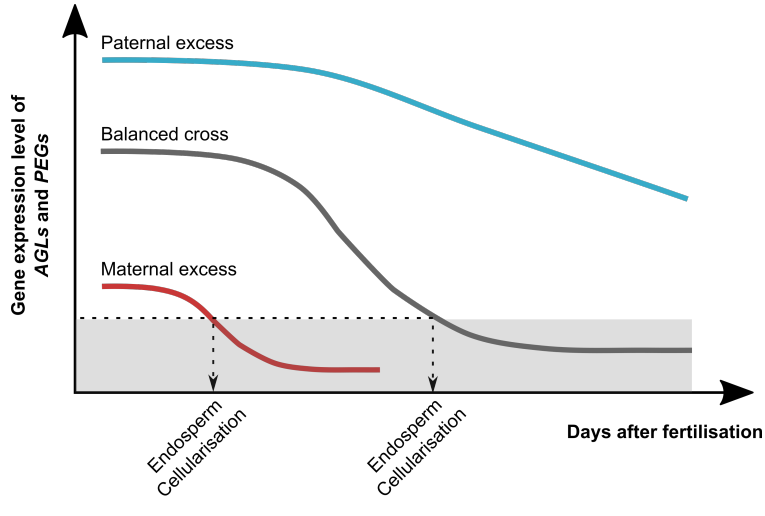

Figure 4. **Expression of *AGLs* and *PEGs* correlates with the timing of endosperm cellularisation.** Expression of *AGLs* and *PEGs* is shown for balanced, paternal excess, and maternal excess crosses. Initiation of cellularisation is associated with a decrease in expression of these genes below a certain level (grey box). Adapted from Schatlowski *et al.*, (2012).

same molecular phenotype as described for interploidy crosses (Walia *et al.*, 2009; Ishikawa *et al.*, 2011; Rebernig *et al.*, 2015; Lafon-Placette *et al.*, 2018; Roth *et al.*, 2018).

2 Aims of the study

The aims of this study are detailed in the introductory section and can be summarised as follows:

- To uncover paternally-derived factors driving endosperm proliferation
- To determine the nature of the endosperm-derived signal that triggers seed coat development
- To identify and characterise the factors controlling endosperm cellularisation
- To characterise the roles of the MADS-box TFs AGL62 and PHE1 in endosperm development
- To deepen the knowledge of the molecular mechanisms driving endosperm-based reproductive barriers

3 Results and Discussion

In this section, I will briefly describe and discuss the main findings obtained from this thesis work. Throughout the text, I refer to the original location of the results within the appropriate papers, which are compiled at the end of this thesis.

3.1 Paternal-derived auxin drives endosperm proliferation

In *Arabidopsis*, the auxin biosynthesis genes *TAR1*, *TAA1* and *YUC10* are expressed in the endosperm, and have been identified as PEGs (Hsieh *et al.*, 2011; Wolff *et al.*, 2011; Pignatta *et al.*, 2014). Through the analysis of fluorescent reporters, we confirmed the paternal-specific expression of *YUC10,* and showed that auxin activity is detected immediately after fertilisation, suggesting this hormone is accumulated in the early endosperm **(paper I - Figure 1)**.

Even though the mechanistic details still remain somewhat elusive, auxin is known to be a permissive signal for cell division (Perrot-Rechenmann, 2010). Because of this, we hypothesized that auxin production after fertilisation could be the trigger that leads to endosperm proliferation – the mysterious "growth-stimulus" postulated by Sargant over a century ago (Sargant, 1900). When analysing seeds from auxin biosynthesis or signalling mutants, we observed marked endosperm proliferation defects: some seeds arrested endosperm proliferation at early developmental stages; while others still showed endosperm proliferation, but at a lower rate when comparing with wt seeds **(paper I - Figure 2)**. This supports the idea that auxin is required for initiation, as well as for maintenance of endosperm proliferation throughout seed development.

Auxin biosynthesis genes are paternally expressed in a wide range of plants, such as in species of the *Arabidopsis* genus, as well as in *Capsella rubella*, rice, and maize (Luo *et al.*, 2011; Wolff *et al.*, 2011; Waters *et al.*, 2013; Pignatta *et*

al., 2014; Hatorangan *et al.*, 2016; Klosinska *et al.*, 2016), suggesting that this hormone could have a conserved role in endosperm development. Work performed in maize revealed that, unlike in *Arabidopsis,* auxin levels remain low during the proliferative phases of endosperm development. Later, the levels of this hormone are dramatically increased, correlating with the onset of endosperm endoreduplication and differentiation (Lur & Setter, 2016). On the other hand, the hormone cytokinin is highly abundant during early developmental stages, leading the authors to hypothesise that endosperm proliferation is mainly controlled by cytokinin. Notwithstanding, this does not rule out a role for auxin in early proliferation of the maize endosperm, nor a possible interaction with cytokinin in the control this process. Similarly to auxin, cytokinin is known to control cell proliferation and growth, often in conjunction with auxin (Schaller *et al.*, 2015). In *Arabidopsis*, cytokinin activity was detected in the endosperm, and the abnormal endosperm growth of *haiku* mutants is associated with abnormal cytokinin activity (Li *et al.*, 2013). Thus, it is possible that both these hormones interact to mediate the control of early endosperm proliferation; nevertheless, further work is required to substantiate this hypothesis.

Interestingly, the fact that all enzymes required for the Trp-dependent auxin biosynthesis pathway are coded by PEGs, reveals a straightforward developmental strategy that links endosperm proliferation with the fertilisation event. In fact, we observed that autonomous endosperm formation in the PRC2 mutants *fis2* and *fie,* correlates with the ectopic activation of maternal auxin biosynthesis genes in the central cell **(paper I - Figure 4, Supplementary Figure 3)**. Additionally, exogenous application of auxin to unfertilised ovules, as well as ectopic production of this hormone in the central cell, led to autonomous endosperm formation, reinforcing the idea that auxin stimulates endosperm proliferation **(paper I - Figure 3)**. More importantly, these results provide a mechanistic explanation for the imprinted expression of auxin biosynthesis genes: repression of maternal alleles via the action of FIS-PRC2 ensures that endosperm proliferation only occurs in the presence of the paternal genome.

Furthermore, these results uncover the molecular mechanisms behind autonomous endosperm formation in FIS-PRC2 mutants, which phenotypically resemble the fertilization-independent endosperm produced by some apomictic species. In apomicts, fertilization-independent seed formation generates new plants which are clonally derived from their mothers (Conner & Ozias-Akins, 2017). Understanding this process is of substantial applied interest, since it allows the fixation of desirable traits in hybrid plants (Hand & Koltunow, 2014). In order to apply apomixis to relevant species, researchers focus on identifying the pathways leading to this phenomenon, both in natural apomicts, as well as in

mutants that exhibit apomictic traits, such as FIS-PRC2 mutants. Revealing that auxin can drive autonomous endosperm proliferation opens the door for further exploration of this phytohormone as a component of apomictic pathways, which may advance clonal seed formation in relevant crop species.

3.2 Seed coat development is triggered by endosperm-derived auxin

In order to address the question of what are the signals triggering seed coat development, we compared the transcriptional profiles of wt seeds and *vrn2/-emf2*/+ ovules, which autonomously develop seed coat. We observed that auxin-related genes were significantly upregulated in both datasets, when compared to unfertilized wt ovules **(paper II - Table 1)**. This led us to hypothesise that the presence of auxin in the integuments could initiate the development of the seed coat. This would be in line with our previous observations that auxin is produced post-fertilization in the endosperm **(paper I – Figure 1)**, and that auxin activity is detected in the seed coat following fertilisation **(paper I – Figure 1)**. Indeed, application of this hormone to unfertilised ovules triggered seed coat development without fertilization **(paper II – Figure 2)**. The same phenotype could be observed when auxin was ectopically produced in the central cell of unfertilised ovules **(paper II - Figure 4)**. This indicates that auxin production in the central cell is sufficient to drive seed coat formation in a non-cell-autonomous manner. Accordingly, impairing auxin biosynthesis in the endosperm of fertilized seeds, but not auxin signalling, led to the formation of smaller seeds, where seed coat expansion is reduced or not initiated **(paper II - Figure 3)**. Additionally, analysis of auxin distribution in *agl62* mutant seeds, which fail to develop a seed coat, revealed a peculiar pattern: while auxin activity was abnormally high in the endosperm, no such activity could be detected in the integuments, contrasting with what happens in a wt situation **(paper II - Figure 5)**. This indicates that in the absence of AGL62, auxin is not adequately transported to the integuments, correlating with this mutant's inability to form a seed coat. Together, these data suggest that upon fertilisation, endosperm-derived auxin is quickly exported to the integuments, in an AGL62-dependent manner, where it drives seed coat initiation **(paper II - Figure 8)**.

Furthermore, we demonstrated that auxin-mediated seed coat development is achieved through the transcriptional downregulation of PcG-coding genes and activation of gibberellin (GA) signalling **(paper II - Figures 1 and 7)**. Thus, directly or indirectly, auxin removes the block on seed coat development, which is exerted by PcG proteins, providing a unique example of hormone-mediated transcriptional regulation of PRC2 **(paper II - Figure 8)**.

GA signalling and biosynthesis were previously shown to be active after fertilisation, and to act downstream of auxin in promoting fruit growth (Dorcey *et al.*, 2009). In our work, we further showed that production of auxin in the central cell is sufficient to trigger parthenocarpic fruit growth **(paper II - Figure 4)**. These data strongly support the presence of a hierarchical signalling pathway that starts with biosynthesis of auxin in the endosperm, and leads to activation of GA signalling and biosynthesis in maternal tissues, thus promoting and coordinating seed and fruit growth.

3.3 Auxin homeostasis regulates endosperm cellularisation

In **paper I** we demonstrated that auxin is a major regulator of endosperm development. As such, we decided to investigate if this hormone is involved in the aberrant endosperm phenotypes observed in interploidy crosses. These phenotypes are characterised by mistimed onset of endosperm cellularisation, which culminates in seed abortion **(section 1.5)**. When comparing the transcriptomic profiles of 2x and 3x seeds, we observed a significant deregulation of auxin related genes **(paper III - Table 1)**. In fact, a wide range of auxin biosynthesis, signalling, and transport genes were upregulated in paternal excess 3x seeds, consistent with increased auxin activity in these seeds **(paper III - Figure 1)**.

Interestingly, abortion of 3x seeds could be rescued by reducing auxin biosynthesis or signalling, and this rescue was accompanied by restoration of endosperm cellularisation **(paper III - Figure 4)**. Consistent with this, overproduction of auxin in the endosperm of 2x seeds mimicked the delayed cellularisation phenotypes of aborting paternal excess 3x seeds **(paper III - Figure 2)**. Furthermore, the expression of auxin-related genes was reduced in 3x seeds mutant for the PEG *ADMETOS* (*ADM*), when compared to wt 3x seeds, which correlates with the restoration of endosperm cellularisation in this mutant **(paper III – Figure 5)** (Kradolfer *et al.*, 2013; Jiang *et al.*, 2017). A similar reduction of expression of auxin-related genes was observed during the developmental stages preceding endosperm cellularisation in 2x seeds **(paper III - Figure 5)**. Together, our results suggest that auxin levels correlate with the timing of endosperm cellularisation, and that in a 2x seed, the levels of this hormone must be reduced in order to allow endosperm cellularisation. Moreover, these data imply that the lack of endosperm cellularisation observed in paternal excess cross seeds is caused, at least in part, by abnormally high levels of auxin.

Endosperm cellularisation is hypothesised to be connected to the rate of nuclear proliferation; nevertheless, several experimental evidence challenge this idea **(section 1.1.4)**. In this study, we demonstrated that reducing endosperm proliferation by genetic means in paternal excess 3x seeds did not increase their viability **(paper III - Figure S8)**. The observed seed abortion in these mutants suggests that the development of 2x seeds remains abnormal, even when endosperm proliferation is decreased. This reinforces the idea that endosperm proliferation and cellularisation can be uncoupled, and it demonstrates that endosperm proliferation defects do not underlie the seed abortion phenotype observed in these seeds. We propose that auxin's control on endosperm cellularisation is independent of its function on endosperm proliferation. Auxin has been previously shown to induce cell wall loosening through demethylesterification of pectin, a process required for meristem growth (Braybrook & Peaucelle, 2013). Pectin demethylesterification is increased in paternal excess 3x seeds where endosperm cellularisation fails (Wolff *et al.*, 2015). Thus, it is plausible to hypothesise that auxin has a positive impact on cell wall loosening in the endosperm, and that onset of cellularisation requires reduction of auxin levels, in order for cell walls to adequately form.

We further observed that overproducing auxin in the sporophytic seed coat mimics the paternal excess 3x seed phenotype, similarly to what happened when overproducing auxin in the endosperm **(paper III - Figure S10)**. In contrast, reducing auxin signalling or cell expansion in the seed coat, does not lead to an extensive rescue of the 3x seed phenotype **(paper III - Figure S8)**. These results show that the sporophytic presence of auxin influences endosperm development in a non-cell-autonomous manner, but that the triploid block phenotype results from a combined deregulation of endosperm and seed coat developmental pathways.

The upregulation of PEGs and AGLs has been implicated in the establishment of the endosperm cellularisation phenotype of 3x seeds **(section 1.5.1)**. In order to assess how auxin influences the activity of these genes, we measured their expression in seeds showing deregulated endosperm cellularisation. We observed that PEGs and AGLs are not upregulated in auxin-induced 3x-like seeds **(paper III - Figure 3)**, and that rescue of 3x seed abortion in auxin signalling mutants did not correlate with a downregulation of PEGs and AGLs **(paper III - Figure 4)**. These results suggest that abnormal endosperm development in 3x seeds is dependent on auxin, but that the effect of this hormone is likely downstream of PEG/AGL activity. This points to auxin possibly being a downstream effector of the pathways controlling the timing of endosperm cellularisation.

3.4 PHE1 regulates key endosperm developmental genes

In the work described before, we have inferred that auxin activity can likely be placed downstream of PEG and AGL activity in the endosperm. We thus hypothesized that expression of auxin-related genes could be under the direct control of PEGs and AGLs. The first evidence supporting this was the observation that in the *agl62* mutant, where auxin is retained in the endosperm, there is a strong downregulation of the gene coding for the putative auxin transporter PGP10 **(paper II - Figure 5)**. PGP10 could potentially facilitate auxin export from the endosperm to the seed coat, and lack of its expression in the *agl62* background suggests that this TF can potentially mediate the transcriptional activation of *PGP10* **(paper II - Figure 8)**.

Remarkably, when performing ChIP-seq of PHE1, the auxin biosynthesis PEGs *TAR1* and *YUC10*, as well as *PGP10*, were identified as direct targets of this TF **(paper IV - Extended Data Table 1)**. Additionally, genes downregulated in *agl62* seeds showed a significant overlap with PHE1 target genes, suggesting these two TFs can potentially control similar targets **(Figure 5)**. This is further substantiated by the fact that these type I MADS-box TFs were shown to interact and form heterodimers *in vitro* (de Folter *et al.*, 2005). PHE1 accesses DNA through CArG-boxes, which are similar to those identified for type II MADS-box TFs **(paper IV - Figure 1)**. Previous studies have shown that a single CArG-box motif can be bound by several different MADS-box TFs (Aerts *et al.*, 2018). Thus, it is likely that AGL62, PHE1, and potentially other type I MADS-box TFs, can share the role of controlling key seed developmental pathways, such as auxin-related genes and others.

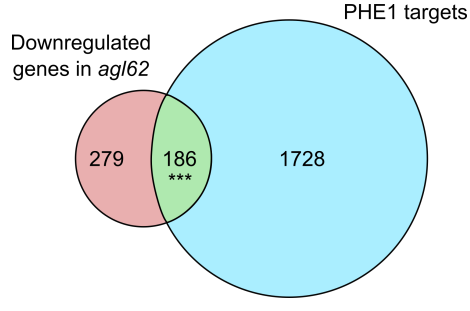

Figure 5. **PHE1 and AGL62 likely regulate similar genes.** Overlap between PHE1 target genes identified through ChIP-seq, and genes significantly downregulated in *agl62* seeds. Statistical significance of the overlap was assessed with a hypergeometric test (p = 3.99e-94).

Furthermore, PHE1 also targets genes previously implicated in endosperm development: such as genes involved in the HAIKU pathway, which regulate endosperm growth, as well as *ZHOUPI*, which codes for an endosperm-specific TF controlling embryo cuticle formation and endosperm breakdown after cellularisation **(paper IV - Extended Data Table 1)**. We also detected an enrichment of transcriptional regulators among PHE1 target genes, and within these, type I MADS-box TFs were overrepresented **(paper IV - Extended Data Figure 1)**. This points to the existence of a complex transcriptional network controlling endosperm development, which is likely mediated by a multitude of type I MADS-box TFs.

In addition to AGLs, we also observed that a significant number of imprinted genes are under the transcriptional control of PHE1, especially PEGs **(paper IV - Figure 2)**, many of which were previously implicated in establishing the triploid block phenotype **(paper IV - Extended Data Table 1)**. Given the significant overrepresentation of these genes among PHE1 targets, we assessed the impact of PHE1 in establishing the seed abortion phenotype observed in paternal excess 3x seeds. Consistent with PHE1 regulating many genes shown to be highly upregulated in 3x seeds **(paper IV - Figure 3)**, mutating PHE1 led to a rescue of the 3x seed abortion phenotype, which was accompanied by restoration of endosperm cellularisation, and reduction of expression of PHE1 targets **(paper IV - Figure 3, Extended Data Figure 8)**. Overall, these data show that PHE1 controls the expression of a significant proportion of AGLs and PEGs, among which are the auxin-related genes implicated in the control of endosperm cellularisation. Thus, the seed abortion phenotype observed in 3x seeds is likely enforced by deregulation of PHE1 and, consequently, its target genes, which ultimately impact the development of the endosperm.

Interestingly, paternal excess 3x seeds show a phenotype that resembles FIS-PRC2 mutant seeds **(section 1.4.3)**. Transcriptomic studies revealed that not only the seed phenotypes are similar, but also the molecular phenotypes, with largely overlapping sets of genes being deregulated in both cases (Erilova *et al.*, 2009; Tiwari *et al.*, 2010). From this, it can be hypothesised that the observed imprinted gene deregulation in 3x seeds could potentially be accompanied by a loss of imprinting, as is the case in mutants of FIS-PRC2. Nevertheless, when analysing parent-of-origin gene expression data we observed that imprinted genes largely retain their parental bias in 3x seeds **(paper IV - Figure 3)**. Concordantly, H3K27me3 deposition in maternal and paternal alleles is not changed in 3x seeds, pointing to a maintenance of imprinting **(paper IV - Figure 3)**.

3.5 PHE1 DNA-binding site accessibility in imprinted genes is conditioned by asymmetric epigenetic modifications

Imprinting is established by the asymmetric deposition of epigenetic modifications in maternal and paternal alleles **(section 1.4)**. The observation that PHE1 controls the expression of several imprinted genes **(paper IV – Figure 2)**, prompted us to evaluate how epigenetic marks are distributed in maternal and paternal alleles of PHE1 DNA-binding sites associated with imprinted genes. We observed that in PEGs and paternally biased genes, H3K27me3 shows a distinct distribution pattern along maternal and paternal PHE1 binding sites: in paternal alleles this histone mark is absent, in contrast with what happens in maternal alleles, where PHE1 binding sites are flanked by high levels of H3K27me3 **(paper IV - Figure 2)**. Curiously, we did not detect this mark at the centres of paternal binding sites **(paper IV - Figure 2)**. This correlates with biallelic binding of PHE1 at the tested PEGs, showing that absence of maternal H3K27me3 at the binding site is permissive for PHE1 binding **(paper IV - Figure 2, Extended Data Figure 5)**. We speculate that PHE1 binding to maternal alleles could serve one of two purposes: i) be required for the expression of maternal alleles of PEGs that do not show exclusive paternal expression; or ii) facilitate the recruitment of H3K27me3, in order to maintain the silencing of these alleles in the endosperm. Some MADS-box TFs have been proposed to act as "pioneer factors" able to recruit chromatin modifying enzymes to their binding sites **(section 1.2)**. Furthermore, the MADS-box SVP was shown to interact with the PRC1 component LHP1, recruiting it to target regions via CArG-boxes, which allows for modulation of H3K27me3 at these loci (Chen *et al.*, 2018). Thus, it can be hypothesized that PHE1 could have a dual role in the regulation of PEGs: besides contributing to the activation of paternal alleles (via dimerization with other AGLs), PHE1 could also enforce the repression of maternal alleles (via multimerization with PRC1/2 components). Nevertheless, the hypothesis that one TF could have a such a dual role within a specific tissue is unconventional, and remains to be tested.

Similarly to what happens in PEGs, distinct epigenetic landscapes were observed between the maternal and paternal alleles of PHE1 binding sites associated with MEGs: while paternal sites are enriched in CG methylation, maternal sites are devoid of it **(paper IV - Figure 2)**. Presence of CG methylation near the TSS has a repressive effect on gene expression (Niederhuth *et al.*, 2016). In agreement with this, we could not detect PHE1 binding in the paternal alleles of the tested MEGs **(paper IV - Figure 2)**, suggesting that the presence of CG methylation at these loci negatively impacts TF accessibility, and therefore transcription.

We propose a model where PHE1 accessibility to DNA-binding sites associated with imprinted genes is conditioned by the asymmetric presence of CG methylation and H3K27me3 in the maternal and paternal genomes **(paper IV - Extended Data Figure 6)**. Thus, these data provide a glimpse into how epigenetic and transcriptional regulation interact, in order to regulate imprinted gene expression.

3.6 Transposable elements facilitate gene targeting by PHE1

While analysing the genomic distribution of PHE1 binding sites we observed that these sites often show a spatial overlap with TEs, and that this overlap was predominantly occurring at RC/Helitrons **(paper IV - Figure 1)**. Interestingly, both the RC/Helitrons that co-localise with PHE1 binding sites, as well as those RC/Helitrons where this co-localisation is not detected, were enriched for PHE1 CArG-boxes **(paper IV - Figure 1)**. This suggests that RC/Helitrons can carry PHE1 DNA-binding motifs, and that these motifs are functionally relevant for PHE1 accessibility to its target genes.

TE insertions are known to affect neighbouring gene expression in different ways (Feschotte, 2008; Hirsch & Springer, 2017): i) they can mutate existent *cis*-regulatory regions, due to transposition; ii) neighbouring genes may be affected by internal TE promoters; iii) epigenetic modifications targeted at TEs can have indirect effects on gene expression, as is the case of imprinting **(section 1.4.2)**; iv) *cis*-regulatory sequences of the host may be mobilized to new genomic locations. Examples of the latter scenario can be abundantly found in animal genomes (Feschotte, 2008; Sundaram *et al.*, 2014). Nevertheless, examples of this situation in plants are uncommon. An exception to this is the case of stress-responsive TEs, such as *ONSEN*, which can recruit the host's heat-shock TFs, influencing nearby gene expression (Cavrak *et al.*, 2014; Hirsch & Springer, 2017). Interestingly, and similarly to what is described here for PHE1, Muiño *et al.*, (2016) found that binding sites for the type II MADS-box TF SEP3 are present within a specific class of LTR TEs of *Arabidopsis lyrata.* These LTR TEs are also present in *A. thaliana*, even though in less copy number and not containing SEP3 CArG-boxes; while in *Capsella* species they are completely absent (Muiño *et al.*, 2016). This suggests that transposition of these TEs in *A. lyrata* allowed to expand the repertoire of SEP3 target genes. The results presented in this thesis, together with the work of Muiño *et al.*, (2016), strongly suggest that TEs might significantly contribute to the generation of novel transcriptional networks in plants, through the spread of TF binding sites. The fact that SEP3 and PHE1 control key developmental genes involved in flowering

and seed development, shows that co-option of TEs can have a significant, yet unexplored impact in plant development.

RC/Helitrons have been previously shown to be enriched in the flanking regions of PEGs (Wolff *et al.*, 2011; Hatorangan *et al.*, 2016). In this work, we showed that PHE1 targets several PEGs **(section 3.4)**, and that RC/Helitrons are enriched for PHE1 DNA-binding motifs **(paper IV - Figure 1)**. In line with this, we detected an overrepresentation of PHE1 binding motifs in RC/Helitrons found in the vicinity of PHE1-targeted PEGs **(paper IV - Extended Data Figure 6)**. Additionally, an analysis of orthologous PHE1 PEG targets revealed that the presence of RC/Helitrons containing putative PHE1 binding motif correlates with the appearance of paternally biased expression in the Brassicaceae **(paper IV - Extended Data Figure 7)**. Together, these data show that the association of PEGs and RC/Helitrons likely facilitates their targeting by PHE1, conferring timely endosperm expression of PEGs. Furthermore, these TEs are known to facilitate the asymmetric deposition of epigenetic marks that lead to imprinting. This has been proposed to drive imprinting both in flowering plants **(section 1.4.2)**, and in therian mammals (Slotkin & Martienssen, 2007). In fact, a high TE content, as well as numerous CG islands are detected in imprinted regions of therian mammalian genomes (Pask *et al.*, 2009). This contrasts with the lack of these features in orthologous regions of the platypus genome, correlating with lack of imprinting in that species (Pask *et al.*, 2009). These data led to the proposition that novel TE insertions in therian mammals drove the appearance of imprinting, likely through the action of TE-mediated epigenetic modifications (Pask *et al.*, 2009). Nevertheless, the role of TEs in providing *cis*-regulatory sequences favourable for imprinted gene expression remains to be fully explored in therian mammals, and in flowering plants. Given the data presented here, we propose that in Brassicaceae RC/Helitrons have a dual role in establishing imprinted expression: they provide *cis*-regulatory sequences that facilitate endosperm transcription, and simultaneously promote the establishment of epigenetic landscapes conducive to imprinting.

4 Conclusions

Together, the findings of this thesis significantly contributed to a better understanding of seed development in *Arabidopsis thaliana*. A graphical summary of the conclusions derived from this thesis can be found in **Figure 6**.

We determined that auxin is produced in a paternal-dependent manner in the endosperm, driving its proliferation. In parallel, auxin is transported to the integuments, where it triggers removal of PRC2 and activates GA signalling, driving seed coat expansion and fruit growth. We further showed that the timing of endosperm cellularisation is determined through the modulation of auxin levels in the endosperm, and that auxin biosynthesis and transport genes are under the transcriptional control of the type I MADS-box TF PHE1, and likely AGL62. Besides controlling expression of auxin-related genes, PHE1 is also responsible for the expression of several other imprinted genes. The asymmetric epigenetic landscapes in maternal and paternal alleles of PHE1 binding sites influence the accessibility of this TF to DNA, leading to parent-of-origin-specific expression of imprinted targets. Interestingly, we found that through transposition, RC/Helitrons likely contributed to the spread of PHE1 binding sites, thus providing an example of TE domestication in a plant genome. This is especially relevant for PHE1 imprinted targets, where RC/Helitrons not only promote epigenetic modifications, but also allow transcriptional control by PHE1. Together, these data uncover the elusive role of type I MADS-box TFs, implicating them as key regulators of seed development. These results also explain the long-standing association between MADS-box gene deregulation and endosperm-based reproductive barriers. Deregulation of these genes leads to deregulation of their targets (*i.e.* PEGs and auxin-related genes), which in turn elicits endosperm cellularisation defects, explaining the seed abortion phenotypes observed in 3x seeds. Thus, the results obtained here not only contribute to a better understanding of endosperm and seed coat developmental pathways, but also show how the deregulation of these pathways enforces endosperm-based reproductive barriers.

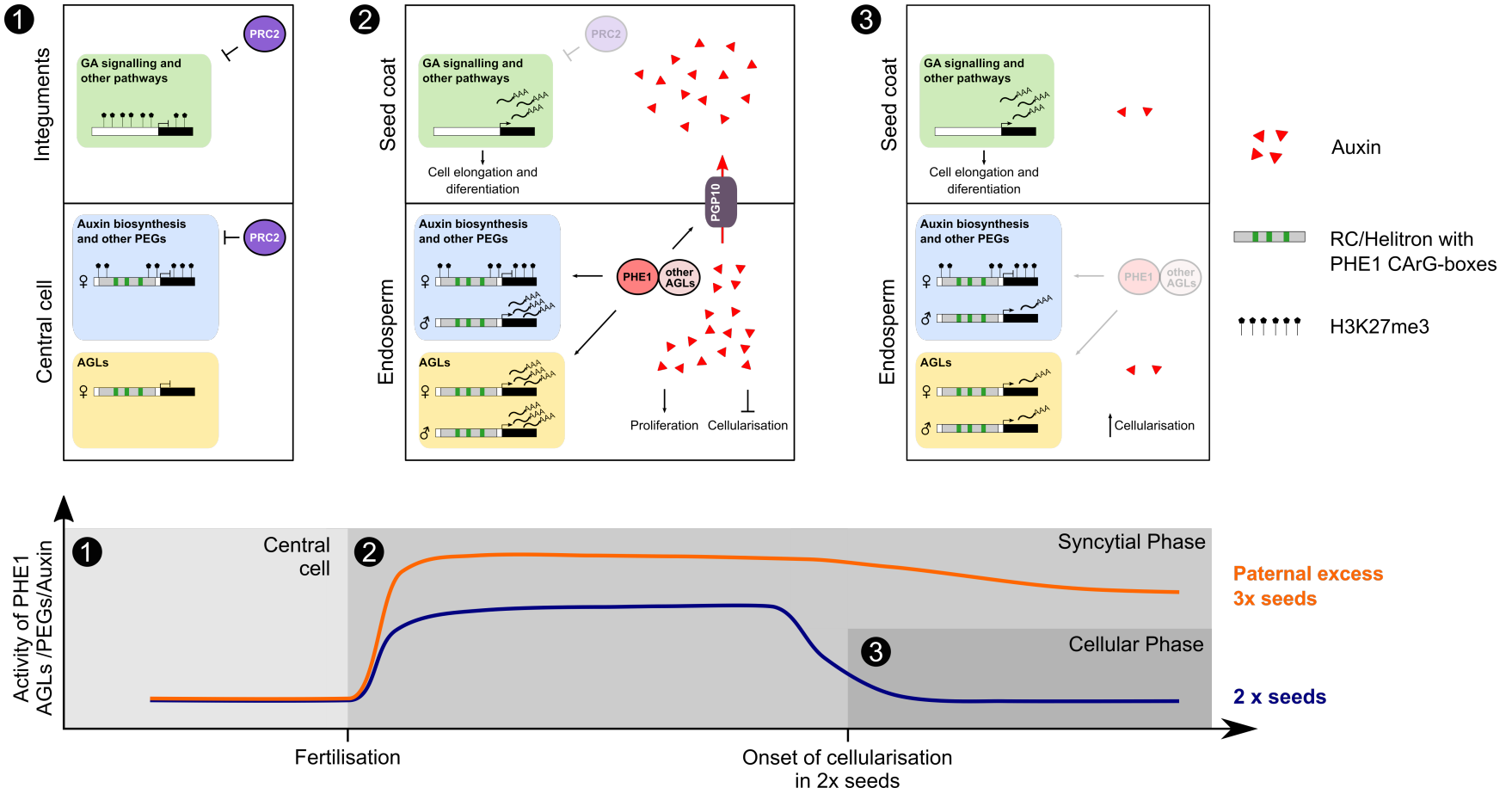
1
Integuments
GA signalling and other pathways
PRC2
Central cell
Auxin biosynthesis and other PEGs
PRC2
AGLs
2
Seed coat
GA signalling and other pathways
PRC2
Cell elongation and diferentiation
PGP10
Endosperm
Auxin biosynthesis and other PEGs
PHE1
other AGLs
AGLs
Proliferation
Cellularisation
3
Seed coat
GA signalling and other pathways
Cell elongation and diferentiation
Endosperm
Auxin biosynthesis and other PEGs
PHE1
other AGLs
AGLs
Cellularisation
Auxin
RC/Helitron with PHE1 CArG-boxes
H3K27me3
Activity of PHE1 AGLs /PEGs/Auxin
1
Central cell
2
Syncytial Phase
3
Cellular Phase
Paternal excess 3x seeds
2 x seeds
Fertilisation
Onset of cellularisation in 2x seeds

Figure 6. **PHE1 and auxin regulate endosperm and seed coat development.** The top panel represents the mature female gametophyte (1), and different components of a 2x seed throughout different stages of seed development (2-3). The bottom panel represents the dynamics of PHE1, AGLs, PEGs, and auxin activity in the central cell (1) and endosperm of 2x (blue) and 3x seeds (orange) (2-3). **(1)** Before fertilisation, FIS-PRC2 deposits repressive H3K27me3 in maternal alleles of PEGs. Endosperm-specific AGLs are not expressed. GA signalling and other unknown pathways are repressed in the integuments through the action of sporophytic PRC2. PHE1, endosperm-specific AGLs, and PEGs are not expressed in the central cell. Therefore, these proteins, as well as auxin, show no activity in this cell. **(2)** After fertilisation, PHE1, likely in conjunction with other AGLs, activates gene expression of *PEGs*, *AGLs* and *PGP10*. PHE1 accessibility to its targets is mediated by the presence of nearby RC/Helitrons containing PHE1 CArG-boxes. Activity of PHE1 target genes leads to production of auxin, which stimulates endosperm proliferation, and prevents cellularisation. Simultaneously, auxin is transported to the integuments, possibly via the action of PGP10. Presence of auxin in the integuments triggers removal of PRC2, and activation of GA signalling and other unknown pathways, which promote differentiation and growth of the seed coat. **(3)** In later stages of 2x seed development PHE1 expression is reduced, as is the expression of its target genes. Lack of auxin production causes a decrease in the levels of this hormone in the endosperm, allowing the onset of cellularisation. This does not occur in 3x seeds, where the activity of PHE1, AGLs, PEGs, and auxin is increased and extended in time. This prolongs the duration of the syncytial phase of endosperm development, and prevents the onset of cellularisation, which results in seed abortion.

5 Future perspectives

Even though this work contributed to the elucidation of several key aspects of seed development, it also opened the door for new questions and new hypotheses to be tested. First, it would be interesting to understand the specific mechanisms through which auxin drives endosperm proliferation, as well as to determine in more detail how this hormone can regulate cellularisation of the endosperm. It also remains to be tested if auxin has similar roles in other angiosperms that possess an endosperm of the nuclear type. If so, it is likely that endosperm-based reproductive barriers in a wide range of plants could be due to deregulation of auxin levels in the endosperm. The fact that auxin biosynthesis genes are imprinted in several species of angiosperms suggests a conserved function for this hormone; nevertheless, this should be adequately tested. It is tempting to envision that auxin could serve as a universal paternal stimulus to the proliferation of the endosperm. For this, it would be necessary to test imprinting of auxin-related genes, as well as to determine if their function is conserved in angiosperms with different endosperm types, as well as in basal angiosperms.

In this work, we showed that auxin induces the transcriptional downregulation of PcG-coding genes in the seed coat, a process required for the development of this structure. Nevertheless, the question of how this is achieved still remains. Since this is a unique example of transcriptional regulation of PcG gene expression, it would be interesting to uncover the mechanistic details behind it, and to determine if this mode of regulation is relevant for other PcG-mediated developmental transitions. Furthermore, ours and others' results clearly show that PcG activity in the integuments represses developmental pathways required for seed coat development, but the identity of these pathways remains unknown. Some of these are likely related to GA biosynthesis and/or signalling; nevertheless, other unknown pathways might be involved. Reduction of PcG activity in the seed coat is likely concomitant with epigenetic reprogramming of these tissues, namely a reduction of H3K27me3. Because seed coat growth and differentiation are achieved without cell division, reduction

of H3K27me3 in these cells could derive from a combination of reduced PcG activity, and active removal of this mark. In parallel, other epigenetic marks could be involved in modulating this rapid developmental transition. A thorough characterisation of the seed coat's epigenome would allow to test these hypotheses, and to identify the pathways responsible for seed coat development.

The role of type I MADS-box TFs in endosperm development has remained largely elusive. In this work, we identified PHE1 and AGL62 as central regulators of seed development, and our data suggests that these TFs cooperate in the control of at least some of their target genes. Given the heterodimerization properties of MADS-box TFs, this cooperation is likely extended to a wide range of other MADS-box TFs, and possibly to TFs of other families. Confirming this would require the identification of AGL62 target genes, as well as the identification of interaction partners of AGL62 and PHE1. In turn, exploring the interactome of these proteins would allow to answer the question of whether these TFs are able to recruit chromatin remodellers or PcG proteins to modulate the epigenetic status of their target genes.

Our results provide support to the idea that TE insertions, in particular of RC/Helitrons, trigger parentally biased expression of nearby genes. This likely stems from their ability to elicit a specific set of epigenetic modifications, coupled with the fact that they can trigger endosperm-specific expression of nearby genes, by providing *cis*-elements for PHE1 binding. Nevertheless, it remains to be experimentally demonstrated if, and how, new TE insertions can lead to imprinting.

Our results further suggest that TE domestication might have played a substantial role in generating transcriptional networks required for endosperm development. Nonetheless, and given the ability of MADS-box TFs to share binding sites, it is likely that molecular domestication of TEs containing CArG-boxes could have a prominent role during other developmental phases of the plant life cycle. Moreover, it could be envisioned that, as is the case in animals, co-option of TEs as TF binding sites could happen beyond the MADS-box TF family. However, this topic remains largely unexplored in plant biology. Further investigations into this hypothesis could potentially reveal that the contribution of TEs for plant development far surpasses what we currently imagine, challenging the paradigm that TE insertions are most often deleterious, and suggesting instead that TEs can coevolve with their hosts in a symbiotic fashion.

References

Abel, S. & Theologis, A. (2010). Odyssey of auxin. *Cold Spring Harbor perspectives in biology*.

Aerts, N., Bruijn, S. de, Mourik, H. van, Angenent, G. C. & Dijk, A. D. J. van (2018). Comparative analysis of binding patterns of MADS-domain proteins in Arabidopsis thaliana. *BMC Plant Biology*, 18(1), p 131 BioMed Central.

Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Pouplana, L. R. de, Martínez-Castilla, L. & Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences*, 97(10), pp 5328–5333 National Academy of Sciences.

Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K. & Kapoor, S. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC genomics*, 8, p 242 BioMed Central.

Austin, B. & Trivers, R. (2009). *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press.

Aw, S. J., Chen, Z., Hamamura, Y., Schnittger, A. & Berger, F. (2010). Sperm entry is sufficient to trigger division of the central cell but the paternal genome is required for endosperm development in Arabidopsis. *Development*, 137(16), pp 2683–2690 The Company of Biologists Ltd.

Baack, E., Melo, M. C., Rieseberg, L. H. & Ortiz-Barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, 207(4), pp 968–984.

Bai, F. & Settles, A. M. (2014). Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Frontiers in plant science*, 5, p 780 Frontiers Media SA.

Bajon, C., Horlow, C., Motamayor, J. C., Sauvanet, A. & Robert, D. (1999). Megasporogenesis in Arabidopsis thaliana L.: an ultrastructural study. *Sexual Plant Reproduction*, 12(2), pp 99–109.

Barbez, E., Kubeš, M., Rolčík, J., Béziat, C., Pěnčík, A., Wang, B., Rosquete, M. R., Zhu, J., Dobrev, P. I., Lee, Y., Zažímalovà, E., Petrášek, J., Geisler, M., Friml, J. & Kleine-Vehn, J. (2012). A novel putative auxin carrier family regulates intracellular auxin homeostasis in plants. *Nature*, 485(7396), pp 119–122 Nature Publishing Group.

Baroux, C. & Grossniklaus, U. (2018). Seeds—An evolutionary innovation underlying reproductive success in flowering plants. *Current Topics in Developmental Biology* pp 605–642.

Baroux, C., Spillane, C. & Grossniklaus, U. (2002). Evolutionary origins of the endosperm in flowering plants. *Genome Biology*, 3(9), p reviews1026 BioMed Central.

Baud, S., Dubreucq, B., Miquel, M., Rochat, C. & Lepiniec, L. (2008). Storage Reserve Accumulation

in Arabidopsis: Metabolic and Developmental Control of Seed Filling. *The Arabidopsis Book*, 6(6), p e0113.

Baud, S., Wuillème, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L. & Rochat, C. (2005). The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in Arabidopsis. *The Plant Journal*, 43(6), pp 824–836 John Wiley & Sons, Ltd (10.1111).

Beauzamy, L., Fourquin, C., Dubrulle, N., Boursiac, Y., Boudaoud, A. & Ingram, G. (2016). Endosperm turgor pressure decreases during early Arabidopsis seed development. *Development*, 143(18), pp 3295–9 Oxford University Press for The Company of Biologists Limited.

Beeckman, T., De Rycke, R., Mane, R. & Hzei, D. (2000). *Histological Study of Seed Coat Development in Arabidopsis thaliana*. *Journal of Plant Research*.

Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D., Le, B. H., Drews, G. N., Brady, S. M., Goldberg, R. B. & Harada, J. J. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proceedings of the National Academy of Sciences*, 110(5), pp E435–E444.

Bemer, M., Heijmans, K., Airoldi, C., Davies, B. & Angenent, G. C. (2010). An atlas of type I MADS box gene expression during female gametophyte and seed development in Arabidopsis. *Plant Physiology*, 154(1), pp 287–300.

Bemer, M., Wolters-Arts, M., Grossniklaus, U. & Angenent, G. C. (2008). The MADS Domain Protein DIANA Acts Together with AGAMOUS-LIKE80 to Specify the Central Cell in Arabidopsis Ovules. *THE PLANT CELL ONLINE*, 20(8), pp 2088–2101.

Bennett, M. J., Marchant, A., Green, H. G., May, S. T., Ward, S. P., Millner, P. A., Walker, A. R., Schulz, B. & Feldmann, K. A. (1996). Arabidopsis AUX1 Gene: A Permease-Like Regulator of Root Gravitropism. *Science*, 273(5277), pp 948–950.

Bewick, A. J. & Schmitz, R. J. (2017). Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, 36, pp 103–110.

Bhatnagar, S. P. & Sawhney, V. (1981). Endosperm—Its Morphology, infrastructure, and Histochemistry. *International Review of Cytology*, 73(C), pp 55–102.

Bicknell, R. A. & Koltunow, A. M. G. (2004). Understanding Apomixis: Recent Advances and Remaining Conundrums. *THE PLANT CELL ONLINE*, 16(suppl_1), pp S228–S245 American Society of Plant Biologists.

Biémont, C. & Vieira, C. (2006). Junk DNA as an evolutionary force. *Nature*, 443(7111), pp 521–524 Nature Publishing Group.

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(3), p 495.

Boutin, T. S., Le Rouzic, A. & Capy, P. (2012). How does selfing affect the dynamics of selfish transposable elements? *Mobile DNA*, 3(1).

Bozhkov, P. V., Filonova, L. H. & Suarez, M. F. (2005). Programmed Cell Death in Plant Embryogenesis. *Current Topics in Developmental Biology*, 67, pp 135–179 Academic Press.

Brandvain & Haig (2005). Divergent Mating Systems and Parental Conflict as a Barrier to Hybridization in Flowering Plants. *The American Naturalist*, 166(3), p 330.

Braybrook, S. A. & Peaucelle, A. (2013). Mechano-Chemical Aspects of Organ Formation in

Arabidopsis thaliana: The Relationship between Auxin and Pectin.(Grebe, M., Ed) *PLoS ONE*, 8(3), p e57813.

Brink, R. A. & Cooper, D. C. (1947). The endosperm in seed development. *The Botanical Review*, 13(9), pp 423–477.

Brown, R. C., Lemmon, B. E., Nguyen, H. & Olsen, O.-A. (1999). Development of endosperm in Arabidopsis thaliana. *Sexual Plant Reproduction*, 12(1), pp 32–42 Springer-Verlag.

Bushell, C., Spielman, M. & Scott, R. J. (2003). The basis of natural and artificial postzygotic hybridization barriers in Arabidopsis species. *The Plant cell*, 15(6), pp 1430–42.

Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijó, J. A., Becker, J. D. & Martienssen, R. A. (2012). Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell*, 151(1), pp 194–205 Cell Press.

Capron, A., Chatfield, S., Provart, N. & Berleth, T. (2009). Embryogenesis: pattern formation from a single cell. *The arabidopsis book*, 7, p e0126 American Society of Plant Biologists.

Carmichael, J. S. & Friedman, W. E. (1996). Double Fertilization in Gnetales : Implications for Understanding Reproductive Diversification among Seed Plants. *International Journal of Plant Sciences*, 157(6), pp 77–94.

Cavrak, V. V, Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L. M. & Mittelsten Scheid, O. (2014). How a Retrotransposon Exploits the Plant's Heat Stress Response for Its Activation. *PLoS Genetics*, 10(1), p e1004115 Public Library of Science.

Charles Darwin (1880). *The Power of Movement in Plants*. London: John Murray.

Chaudhury, A. M., Ming, L., Miller, C., Craig, S., Dennis, E. S. & Peacock, W. J. (1997). Fertilization-independent seed development in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 94(8), pp 4223–8.

Chen, C., Li, T., Zhu, S., Liu, Z., Shi, Z., Zheng, X., Chen, R., Huang, J., Shen, Y., Luo, S., Wang, L., Liu, Q.-Q. & E, Z. (2018). Characterization of Imprinted Genes in Rice Reveals Conservation of Regulation and Imprinting with Other Plant Species. *Plant physiology*, 177(4), pp 1754–1771 American Society of Plant Biologists.

Chen, L.-Q., Lin, I. W., Qu, X.-Q., Sosso, D., McFarlane, H. E., Londoño, A., Samuels, A. L. & Frommer, W. B. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. *The Plant cell*, 27(3), pp 607–19 American Society of Plant Biologists.

Chénais, B., Caruso, A., Hiard, S. & Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene*, 509(1), pp 7–15 Elsevier.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E. & Fischer, R. L. (2002). DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis. *Cell*, 110(1), pp 33–42 Cell Press.

Christensen, C. A., King, E. J., Jordan, J. R. & Drews, G. N. (1997). Megagametogenesis in Arabidopsis wild type and the Gf mutant. *Sexual Plant Reproduction*, 10(1), pp 49–64.

Colombo, M., Masiero, S., Vanzulli, S., Lardelli, P., Kater, M. M. & Colombo, L. (2008). AGL23, a type I MADS-box gene that controls female gametophyte and embryo development in Arabidopsis. *The Plant Journal*, 54(6), pp 1037–1048 John Wiley & Sons, Ltd (10.1111).

Conner, J. A. & Ozias-Akins, P. (2017). Apomixis: Engineering the ability to harness hybrid vigor in

crop plants. *Methods in Molecular Biology*. pp 17–34. Humana Press, New York, NY.

Corlett, R. T. (2016). Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Diversity*, 38(1), pp 10–16.

Costa, L. M., Marshall, E., Tesfaye, M., Silverstein, K. A. T., Mori, M., Umetsu, Y., Otterbach, S. L., Papareddy, R., Dickinson, H. G., Boutiller, K., VandenBosch, K. A., Ohki, S. & Gutierrez-Marcos, J. F. (2014). Central cell-derived peptides regulate early embryo patterning in flowering plants. *Science (New York, N.Y.)*, 344(6180), pp 168–72.

Creff, A., Brocard, L. & Ingram, G. (2015). A mechanically sensitive cell layer regulates the physical properties of the Arabidopsis seed coat. *Nature Communications*, 6(1), p 6382 Nature Publishing Group.

Davies, P. J. (2010). The Plant Hormones: Their Nature, Occurrence, and Functions. *Plant Hormones*. pp 1–15. Dordrecht: Springer Netherlands. ISBN 9780080961569.

Debeaujon, I., Léon-Kloosterziel, K. M. & Koornneef, M. (2002). Influence of the Testa on Seed Dormancy, Germination, and Longevity in Arabidopsis. *Plant Physiology*, 122(2), pp 403–414.

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

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D. & Jacobsen, S. E. (2012). Loss of the DNA Methyltransferase MET1 Induces H3K9 Hypermethylation at PcG Target Genes and Redistribution of H3K27 Trimethylation to Transposons in Arabidopsis thaliana.(Ferguson-Smith, A. C., Ed) *PLoS Genetics*, 8(11), p e1003062 Public Library of Science.

Dilkes, B. P. & Comai, L. (2004). A Differential Dosage Hypothesis for Parental Effects in Seed Development. *The Plant Cell*, 16(12), pp 3174–3180 American Society of Plant Biologists.

Dorcey, E., Urbez, C., Blázquez, M. A., Carbonell, J. & Perez-Amador, M. A. (2009). Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in Arabidopsis. *The Plant Journal*, 58(2), pp 318–32.

Drews, G. N. & Koltunow, A. M. G. (2011). The female gametophyte. *The Arabidopsis Book*, 9, p e0155.

Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, A., Hetzel, J., Wohlschlegel, J. A., Pradhan, S., Patel, D. J. & Jacobsen, S. E. (2012). Dual Binding of Chromomethylase Domains to H3K9me2-Containing Nucleosomes Directs DNA Methylation in Plants. *Cell*, 151(1), pp 167–180 Cell Press.

Enders, T. A. & Strader, L. C. (2015). Auxin activity: Past, present, and future. *American Journal of Botany*, 102(2), pp 180–196.

Erilova, A., Brownfield, L., Exner, V., Rosa, M., Twell, D., Scheid, O. M., Hennig, L. & Köhler, C. (2009). Imprinting of the Polycomb group gene MEDEA serves as a ploidy sensor in Arabidopsis.(Copenhaver, G. P., Ed) *PLoS Genetics*, 5(9), p e1000663 Public Library of Science.

Faure, J. E., Rotman, N., Fortuné, P. & Dumas, C. (2002). Fertilization in Arabidopsis thaliana wild type: Developmental stages and time course. *Plant Journal*, 30(4), pp 481–488 John Wiley & Sons, Ltd (10.1111).

Feng, S. & Jacobsen, S. E. (2011). Epigenetic modifications in plants: an evolutionary perspective. *Current Opinion in Plant Biology*, 14(2), pp 179–186 Elsevier Current Trends.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews*

Genetics, 9(5), pp 397–405 Nature Publishing Group.

Figueiredo, D. D. & Köhler, C. (2016). Bridging the generation gap: communication between maternal sporophyte, female gametophyte and fertilization products. *Current Opinion in Plant Biology*, 29, pp 16–20.

Fincher, G. B. (1989). Molecular and Cellular Biology Associated with Endosperm Mobilization in Germinating Cereal Grains. *Annual Review of Plant Physiology and Plant Molecular Biology*, 40(1), pp 305–346 Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA.

Florez-Rueda, A. M., Paris, M., Schmidt, A., Widmer, A., Grossniklaus, U. & Städler, T. (2016). Genomic Imprinting in the Endosperm Is Systematically Perturbed in Abortive Hybrid Tomato Seeds. *Molecular Biology and Evolution*, 33(11), pp 2935–2946 Narnia.

de Folter, S., Immink, R. G. H., Kieffer, M., Parenicová, L., Henz, S. R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M. M., Davies, B. & Angenent, G. C. (2005). Comprehensive interaction map of the Arabidopsis MADS Box transcription factors. *The Plant Cell*, 17(5), pp 1424–33.

Friml, J., Vieten, A., Sauer, M., Weijers, D., Schwarz, H., Hamann, T., Offringa, R. & Jürgens, G. (2003). Efflux-dependent auxin gradients establish the apical–basal axis of Arabidopsis. *Nature*, 426(6963), pp 147–153 Nature Publishing Group.

Fultz, D., Choudury, S. G. & Slotkin, R. K. (2015). Silencing of active transposable elements in plants. *Current Opinion in Plant Biology*, 27, pp 67–76 Elsevier Ltd.

Garcia, D., Fitz Gerald, J. N. & Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. *The Plant Cell*, 17(1), pp 52–60 American Society of Plant Biologists.

Garcia, D., Saingery, V., Chambrier, P., Mayer, U., Jürgens, G. & Berger, F. (2003). Arabidopsis haiku mutants reveal new controls of seed size by endosperm. *Plant Physiology*, 131(4), pp 1661–70 American Society of Plant Biologists.

Gehring, M. (2013). Genomic Imprinting: Insights From Plants. *Annual Review of Genetics*, 47(1), pp 187–208 Annual Reviews.

Gehring, M., Bubb, K. L. & Henikoff, S. (2009). Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science (New York, N.Y.)*, 324(5933), pp 1447–51 American Association for the Advancement of Science.

Gehring, M., Huh, J. H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (2006). DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. *Cell*, 124(3), pp 495–506 Cell Press.

Gehring, M. & Satyaki, P. R. (2017). Endosperm and Imprinting, Inextricably Linked. *Plant Physiology*, 173(1), pp 143–154.

Goldsmith, M. H. M. (1977). The Polar Transport of Auxin. *Annual Review of Plant Physiology*, 28(1), pp 439–478.

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldán-Arjona, T., David, L. & Zhu, J.-K. (2002). ROS1, a Repressor of Transcriptional Gene Silencing in Arabidopsis, Encodes a DNA Glycosylase/Lyase. *Cell*, 111(6), pp 803–814 Cell Press.

Gramzow, L., Ritz, M. S. & Theißen, G. (2010). On the origin of MADS-domain transcription factors. *Trends in Genetics*.

Gramzow, L. & Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biology* BioMed Central.

Gramzow, L., Weilandt, L. & Theißen, G. (2014). MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. *Annals of Botany*, 114(7), pp 1407–1429 Oxford University Press.

Guignard, L. (1899). Sur les anthérozoides et la double copulation sexualle chex les végétaux angiospermes. *Rev. Gén. Bot.*, 11, pp 129–135.

Guilfoyle, T. J. & Hagen, G. (2007). Auxin response factors. *Current Opinion in Plant Biology*, 10(5), pp 453–460 Elsevier Current Trends.

Guitton, A.-E. & Berger, F. (2005). Loss of Function of MULTICOPY SUPPRESSOR OF IRA 1 Produces Nonviable Parthenogenetic Embryos in Arabidopsis. *Current Biology*, 15(8), pp 750–754 Cell Press.

Gutierrez-Marcos, J. F., Pennington, P. D., Costa, L. M. & Dickinson, H. G. (2003). Imprinting in the endosperm: A possible role in preventing wide hybridization., 2003. pp 1105–1111.

Hafidh, S., Fíla, J. & Honys, D. (2016). Male gametophyte development and function in angiosperms: a general concept. *Plant Reproduction*.

Haig, D. (2004). Genomic Imprinting and Kinship: How Good is the Evidence? *Annual Review of Genetics*, 38(1), pp 553–585 Annual Reviews.

Haig, D. & Westoby, M. (1989). Parent-Specific Gene Expression and the Triploid Endosperm. *The American Naturalist*, 134(1), pp 147–155.

Haig, D. & Westoby, M. (1991). Genomic Imprinting in Endosperm: Its Effect on Seed Development in Crosses between Species, and between Different Ploidies of the Same Species, and Its Implications for the Evolution of Apomixis. *Philosophical Transactions of the Royal Society*, 333(1266), pp 1–13.

Håkansson, A. (1953). Endosperm formation after 2x, 4x crosses in certain cereals, especially in Hordeum vulgare. *Hereditas*, 39(1–2), pp 57–64.

Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M. M., Sasaki, N., Nakano, A., Berger, F. & Higashiyama, T. (2011). Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in Arabidopsis thaliana. *Current Biology*, 21(6), pp 497–502 Cell Press.

Hand, M. L. & Koltunow, A. M. G. (2014). The Genetic Control of Apomixis: Asexual Seed Formation. *Genetics*, 197(2), p 441 Genetics Society of America.

Haslam, T. M. & Yeung, E. C. (2011). Zygotic embryo culture: an overview. *Methods in Molecular biology*. pp 3–15. (Clifton, N.J.): Humana Press.

Hatorangan, M. R., Laenen, B., Steige, K. A., Slotte, T. & Köhler, C. (2016). Rapid Evolution of Genomic Imprinting in Two Species of the Brassicaceae. *The Plant Cell*, 28(8), pp 1815–1827.

Haughn, G. & Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. *Trends in Plant Science*, 10(10), pp 472–477 Elsevier Current Trends.

Hehenberger, E., Kradolfer, D. & Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development*, 139(11), pp 2031–9.

Hill, L. M., Morley-Smith, E. R. & Rawsthorne, S. (2003). Metabolism of sugars in the endosperm of developing seeds of oilseed rape. *Plant Physiology*, 131(1), pp 228–36 American Society of Plant Biologists.

Hirsch, C. D. & Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1860(1), pp 157–165 Elsevier.

Hsieh, T.-F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L. & Zilberman, D.

(2009). Genome-wide demethylation of Arabidopsis endosperm. *Science (New York, N.Y.)*, 324(5933), pp 1451–4 American Association for the Advancement of Science.

Hsieh, T. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., Hashimoto, M., Kirkbride, R. C., Harada, J. J., Zilberman, D. & Fischer, R. L. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. *PNAS*, 108(5), pp 1755–62 National Academy of Sciences.

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T.-F., Uzawa, R., Rodrigues, J. A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R. L., Tamaru, H. & Zilberman, D. (2012). Active DNA Demethylation in Plant Companion Cells Reinforces Transposon Methylation in Gametes. *Science*, 337(6100), pp 1360–1364.

Immink, R. G. H. & Angenent, G. C. (2002). Transcription factors do it together: The hows and whys of studying protein-protein interactions. *Trends in Plant Science*, 7(12), pp 531–534.

Ingram, G. C. (2010). Family life at close quarters: Communication and constraint in angiosperm seed development. *Protoplasma*.

Ingram, G. & Gutierrez-Marcos, J. (2015). Peptide signalling during angiosperm seed development. *Journal of Experimental Botany*, 66(17), pp 5151–5159 Oxford University Press.

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N. & Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant Journal*, 65(5), pp 798–806 John Wiley & Sons, Ltd (10.1111).

Jiang, H., Moreno-Romero, J., Santos-González, J., De Jaeger, G., Gevaert, K., Van De Slijke, E. & Köhler, C. (2017). Ectopic application of the repressive histone modification H3K9me2 establishes post-zygotic reproductive isolation in Arabidopsis thaliana. *Genes & Development*, 31(12), pp 1272–1287 Cold Spring Harbor Laboratory Press.

Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J. & Jacobsen, S. E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Current Biology*, 17(4), pp 379–384.

Johnston, S. A., den Nijs, T. P. M., Peloquin, S. J. & Hanneman Jr, R. E. (1980). The significance of genic balance to endosperm development. *Theoretical and Applied Genetics*, 57, pp 5--9.

Johnston, S. & Hanneman, R. (1982). Manipulations of Endosperm Balance Number Overcome Crossing Barriers Between Diploid Solanum Species. *Science*, 217(July), pp 446–449.

Josefsson, C., Dilkes, B. & Comai, L. (2006). Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. *Current Biology*, 16(13), pp 1322–1328 Cell Press.

Jullien, P. E., Kinoshita, T., Ohad, N. & Berger, F. (2006). Maintenance of DNA Methylation during the Arabidopsis Life Cycle Is Essential for Parental Imprinting. *The Plant Cell Online*, 18(6), pp 1360–1372.

Jürgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature*, 316(6024), pp 153–155 Nature Publishing Group.

Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A. & Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. *The Plant Cell*, 20(3), pp 635–47.

Kapitonov, V. V & Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *Trends in Genetics*, 23(10), pp 521–529.

Kasahara, H. (2016). Current aspects of auxin biosynthesis in plants. *Bioscience, Biotechnology and Biochemistry*.

Kaufmann, K., Melzer, R. & Theißen, G. (2005). MIKC-type MADS-domain proteins: Structural

modularity, protein interactions and network evolution in land plants. *Gene*.

Kawashima, T. & Goldberg, R. B. (2010). The suspensor: not just suspending the embryo. *Trends in Plant Science*, 15(1), pp 23–30 Elsevier Current Trends.

Kermicle, J. L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics*, 66(1), pp 69–85.

Khan, D., Millar, J. L., Girard, I. J. & Belmonte, M. F. (2014). Transcriptional circuitry underlying seed coat development in Arabidopsis. *Plant Science*, 219–220, pp 51–60 Elsevier.

Kim, M., Ohr, H., Lee, J. W., Hyun, Y., Fischer, R. L. & Choi, Y. (2008). Temporal and Spatial Downregulation of Arabidopsis MET1 Activity Results in Global DNA Hypomethylation and Developmental Defects. *Molecular Cell*, 26(6), pp 611–615 The Korean Society for Molecular and Cellular Biology.

Kinoshita, T. (2004). One-Way Control of FWA Imprinting in Arabidopsis Endosperm by DNA Methylation. *Science*, 303(5657), pp 521–523 American Association for the Advancement of Science.

Kirkbride, R. C., Yu, H. H., Nah, G., Zhang, C., Shi, X. & Chen, Z. J. (2015). An Epigenetic Role for Disrupted Paternal Gene Expression in Postzygotic Seed Abortion in Arabidopsis Interspecific Hybrids. *Molecular Plant*, 8(12), pp 1766–1775 Cell Press.

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (1999). Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp 4186–91 National Academy of Sciences.

Klosinska, M., Picard, C. L. & Gehring, M. (2016). Conserved imprinting associated with unique epigenetic signatures in the Arabidopsis genus. *Nature Plants*, 2(10), p 16145.

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. & Grossniklaus, U. (2003). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes and Development*, 17(12), pp 1540–1553.

Köhler, C., Page, D. R., Gagliardini, V. & Grossniklaus, U. (2005). The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nature Genetics*, 37(1), pp 28–30 Nature Publishing Group.

Köhler, C., Wolff, P. & Spillane, C. (2012). Epigenetic Mechanisms Underlying Genomic Imprinting in Plants. *Annual Review of Plant Biology*, 63(1), pp 331–352 Annual Reviews .

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A. & Köhler, C. (2013). An Imprinted Gene Underlies Postzygotic Reproductive Isolation in Arabidopsis thaliana. *Developmental Cell*, 26(5), pp 525–535 Cell Press.

Lafon-Placette, C., Hatorangan, M. R., Steige, K. A., Cornille, A., Lascoux, M., Slotte, T. & Köhler, C. (2018). Paternally expressed imprinted genes associate with hybridization barriers in Capsella. *Nature Plants*, 4(6), pp 352–357 Nature Publishing Group.

Lafon-Placette, C., Johannessen, I. M., Hornslien, K. S., Ali, M. F., Bjerkan, K. N., Bramsiepe, J., Glöckle, B. M., Rebernig, C. A., Brysting, A. K., Grini, P. E. & Köhler, C. (2017). Endosperm-based hybridization barriers explain the pattern of gene flow between Arabidopsis lyrata and Arabidopsis arenosa in Central Europe. *PNAS*, 114(6), pp E1027–E1035 National Academy of Sciences.

Lafon-Placette, C. & Köhler, C. (2014). Embryo and endosperm, partners in seed development. *Current Opinion in Plant Biology*, 17(1), pp 64–69.

Lafon-Placette, C. & Köhler, C. (2016). Endosperm-based postzygotic hybridization barriers: developmental mechanisms and evolutionary drivers. *Molecular Ecology*, 25(11), pp 2620–2629 John Wiley & Sons, Ltd (10.1111).

Laux, T., Würschum, T. & Breuninger, H. (2004). Genetic regulation of embryonic pattern formation. *The Plant cell*, 16 Suppl(suppl_1), pp S190-202.

Law, J. A. & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3), pp 204–220 Nature Publishing Group.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N. & Caboche, M. (2006). Genetics and Biochemistry of Seed Flavonoids. *Annual Review of Plant Biology*, 57(1), pp 405–430.

Leseberg, C. H., Li, A., Kang, H., Duvall, M. & Mao, L. (2006). Genome-wide analysis of the MADS-box gene family in Populus trichocarpa. *Gene*, 378(1–2), pp 84–94.

Leslie, A. B. & Kevin Boyce, C. (2012). Ovule Function and the Evolution of Angiosperm Reproductive Innovations. *International Journal of Plant Sciences*, 173(6), pp 640–648.

Lester, R. & Kang, J. H. (1998). Embryo and Endosperm Function and Failure in Solanum Species and Hybrids. *Annals of Botany*, 82(4), pp 445–453 Oxford University Press.

Li, J., Nie, X., Tan, J. L. H. & Berger, F. (2013). Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(38), pp 15479–84 National Academy of Sciences.

Lin, B. Y. (1984). Ploidy barrier to endosperm development in maize. *Genetics*, 107(1), pp 103–15.

Lisch, D. (2012). How important are transposons for plant evolution? *Nature Reviews Genetics*, 14(1), pp 49–61 Nature Publishing Group.

Lu, J., Zhang, C., Baulcombe, D. C. & Chen, Z. J. (2012). Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. *Proceedings of the National Academy of Sciences*,.

Luo, M., Bilodeau, P., Dennis, E. S., Peacock, W. J. & Chaudhury, A. (2000). Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proceedings of the National Academy of Sciences of the United States of America*, 97(19), pp 10637–42 National Academy of Sciences.

Luo, M., Dennis, E. S., Chaudhury, A., Peacock, W. J. & Berger, F. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *PNAS*, 102(48), pp 17531–17536.

Luo, M., Taylor, J. M., 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.(Kakutani, T., Ed) *PLoS Genetics*, 7(6), p e1002125 Public Library of Science.

Lur, H. S. & Setter, T. L. (2016). Role of Auxin in Maize Endosperm Development. *Plant Physiology*, 103(1), pp 273–280.

Maheshwari, P. (1950). *An introduction to the embryology of angiosperms.* 1st ed. New York,: McGraw-Hill,. ISBN 9788578110796.

Makarevich, G., Villar, C. B. R., Erilova, A. & Kohler, C. (2008). Mechanism of PHERES1 imprinting in Arabidopsis. *Journal of Cell Science*, 121(6), pp 906–912.

Mansfield, S. G. & Briarty, L. G. (1990a). Development of the free-nuclear endopserm in Arabidopsis thaliana (L.). *Arab. Inf. Serv.* pp 53–64.

Mansfield, S. G. & Briarty, L. G. (1990b). Endosperm cellularization in Arabidopsis thaliana (L.). *Arab. Inf. Serv.* pp 65–72.

Mansfield, S. G. & Briarty, L. G. (1991). Early embryogenesis in Arabidopsis. I. The mature embryo sac. *Can. J. Bot.*, 69, pp 461–476.

Marks, G. E. (1996). The Origin and Significance of Intraspecific Polyploidy: Experimental Evidence from Solanum chacoense. *Evolution*, 20(4), p 552.

Martínez, G., Panda, K., Köhler, C. & Slotkin, R. K. (2016). Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell. *Nature Plants*, 2(4), p 16030 Nature Publishing Group.

Martinez, G., Wolff, P., Wang, Z., Moreno-Romero, J., Santos-González, J., Conze, L. L., DeFraia, C., Slotkin, R. K. & Köhler, C. (2018). Paternal easiRNAs regulate parental genome dosage in Arabidopsis. *Nature Genetics*, 50(2), pp 193–198 Nature Publishing Group.

Mascarenhas, J. P. (1989). The male gametophyte of flowering plants. *Plant Cell*, 1(7), pp 657–664.

Mashiguchi, K., Tanaka, K., Sakai, T., Sugawara, S., Kawaide, H., Natsume, M., Hanada, A., Yaeno, T., Shirasu, K., Yao, H., McSteen, P., Zhao, Y., Hayashi, K., Kamiya, Y. & Kasahara, H. (2011). The main auxin biosynthesis pathway in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45), pp 18512–7.

Matzke, M. A. & Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, 15(6), pp 394–408 Nature Publishing Group.

Mayr, E. (1996). What Is a Species, and What Is Not? *Philosophy of Science*, 63(June), pp 262–277.

Messenguy, F. & Dubois, E. (2003). Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. *Gene*.

Millar, J. L., Khan, D., Becker, M. G., Chan, A., Dufresne, A., Sumner, M. & Belmonte, M. F. (2015). Chalazal seed coat development in Brassica napus. *Plant Science*, 241, pp 45–54 Elsevier.

Mizuta, Y. & Higashiyama, T. (2018). Chemical signaling for pollen tube guidance at a glance. *Journal of Cell Science*, 131(2), p jcs208447 The Company of Biologists Ltd.

Möller, B. & Weijers, D. (2009). Auxin control of embryo patterning. *Cold Spring Harbor Perspectives in Biology*, 1(5), p a001545.

Moreno-Romero, J., Del Toro-De León, G., Yadav, V. K., Santos-González, J. & Köhler, C. (2019). Epigenetic signatures associated with imprinted paternally expressed genes in the Arabidopsis endosperm. *Genome Biology*, 20(1), p 41 BioMed Central.

Moreno-Romero, J., Jiang, H., Santos-González, J. & Köhler, C. (2016). Parental epigenetic asymmetry of PRC2-mediated histone modifications in the Arabidopsis endosperm. *The EMBO Journal*, 35(12), pp 1298–1311 EMBO Press.

Morley-Smith, E. R., Pike, M. J., Findlay, K., Köckenberger, W., Hill, L. M., Smith, A. M. & Rawsthorne, S. (2008). The transport of sugars to developing embryos is not via the bulk endosperm in oilseed rape seeds. *Plant Physiology*, 147(4), pp 2121–30 American Society of Plant Biologists.

Mozgova, I. & Hennig, L. (2015). The Polycomb Group Protein Regulatory Network. *Annu. Rev. Plant Biol*, 66, pp 269–296.

Mozgova, I., Köhler, C. & Hennig, L. (2015). Keeping the gate closed: Functions of the polycomb repressive complex PRC2 in development. *Plant Journal*, 83(1), pp 121–132.

Muiño, J. M., de Bruijn, S., Pajoro, A., Geuten, K., Vingron, M., Angenent, G. C. & Kaufmann, K. (2016). Evolution of DNA-Binding Sites of a Floral Master Regulatory Transcription Factor. *Molecular Biology and Evolution*, 33(1), pp 185–200.

Nam, J., Kim, J., Lee, S., An, G., Ma, H. & Nei, M. (2004). Type I MADS-box genes have experienced faster birth-and-death evolution than type II MADS-box genes in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7), pp 1910–5.

Nawaschin, S. (1898). Resultate einer Revision der Befruchtungsvorgänge bei Lilium martagon und Fritillaria tenella. *Bull. Acad. Imp. Sci. St. Petersburg*, 9, pp 377–382.

Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. Do, Li, Q., Rohr, N. A., Rambani, A., Burke, J. M., Udall, J. A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S. A., Springer, N. M. & Schmitz, R. J. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*, 17(1), p 194 BioMed Central.

Noh, B., Murphy, A. S. & Spalding, E. P. (2001). Multidrug resistance–like genes of Arabidopsis required for auxin transport and auxin-mediated development. *The Plant Cell*, 13(11), p 2441 Am Soc Plant Biol.

Nonhebel, H. M. (2015). Tryptophan-Independent Indole-3-Acetic Acid Synthesis: Critical Evaluation of the Evidence. *Plant Physiology*, 169(2), pp 1001–1005.

Normanly, J., Cohen, J. D. & Fink, G. R. (1993). Arabidopsis thaliana auxotrophs reveal a tryptophan-independent biosynthetic pathway for indole-3-acetic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21), pp 10355–9 National Academy of Sciences.

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rosselló, J. A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P. K., Lundeberg, J. & Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), pp 579–584 Nature Publishing Group.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *The Plant cell*, 11(3), pp 407–16 American Society of Plant Biologists.

Ohto, M.-A., Fischer, R. L., Goldberg, R. B., Nakamura, K. & Harada, J. J. (2005). Control of seed mass by APETALA2. *Proceedings of the National Academy of Sciences*, 102(8), pp 3123–3128.

Olsen, O.-A. (2004). Nuclear endosperm development in cereals and Arabidopsis thaliana. *The Plant cell*, 16 Suppl(suppl_1), pp S214-27.

Ortega-Galisteo, A. P., Morales-Ruiz, T., Ariza, R. R. & Roldán-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Molecular Biology*, 67(6), pp 671–681 Springer Netherlands.

Otto, S. P. & Whitton, J. (2002). Polyploid Incidence and Evolution. *Annual Review of Genetics*, 34(1), pp 401–437.

Owen, H. A. & Makaroff, C. A. (1995). Ultrastructure of microsporogenesis and microgametogenesis inArabidopsis thaliana (L.) Heynh. ecotype Wassilewskija (Brassicaceae). *Protoplasma*, 185(1–2), pp 7–21 Springer-Verlag.

Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J.

F., Balazadeh, S., Arif, M., Ó'Maoiléidigh, D. S., Wellmer, F., Krajewski, P., Riechmann, J.-L., Angenent, G. C. & Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology*, 15(3), p R41 BioMed Central.

Palovaara, J., De Zeeuw, T. & Weijers, D. (2016). Tissue and Organ Initiation in the Plant Embryo: A First Time for Everything. *Annu. Rev. Cell Dev. Biol*, 32, pp 47–75.

Panda, K. & Slotkin, R. K. (2013). Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant Signaling and Behavior*, 8(8).

Pařenicová, L., De Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., Angenent, G. C. & Colombo, L. (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, pp 1538–1551.

Pask, A. J., Papenfuss, A. T., Ager, E. I., McColl, K. A., Speed, T. P. & Renfree, M. B. (2009). Analysis of the platypus genome suggests a transposon origin for mammalian imprinting. *Genome biology*, 10(1), p R1 BioMed Central.

Patten, M. M., Ross, L., Curley, J. P., Queller, D. C., Bonduriansky, R. & Wolf, J. B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity*, 113(2), pp 119–128 Nature Publishing Group.

Perrot-Rechenmann, C. (2010). Cellular responses to auxin: division versus expansion. *Cold Spring Harbor Perspectives in Biology* Cold Spring Harbor Laboratory Press.

Petrasek, J. & Friml, J. (2009). Auxin transport routes in plant development. *Development*, 136(16), pp 2675–2688.

Petrásek, J., Mravec, J., Bouchard, R., Blakeslee, J. J., Abas, M., Seifertová, D., Wisniewska, J., Tadele, Z., Kubes, M., Covanová, M., Dhonukshe, P., Skupa, P., Benková, E., Perry, L., Krecek, P., Lee, O. R., Fink, G. R., Geisler, M., Murphy, A. S., Luschnig, C., Zazímalová, E. & Friml, J. (2006). PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science (New York, N.Y.)*, 312(5775), pp 914–8 American Association for the Advancement of Science.

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W. & Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *eLife*, 4, pp 1–24.

Portereiko, M. F., Lloyd, A., Steffen, J. G., Punwani, J. A., Otsuga, D. & Drews, G. N. (2006). AGL80 Is Required for Central Cell and Endosperm Development in Arabidopsis. *THE PLANT CELL ONLINE*, 18(8), pp 1862–1872 American Society of Plant Biologists.

Raghavan, V. (2006). Double Fertilization — A Defining Feature of Flowering Plants. *Double Fertilization*. pp 1–28. Berlin/Heidelberg: Springer-Verlag.

Raven, J. A. (1975). Transport of indoleacetic acid in plant cells in relation to PH and electrical potential gradients, and its significance for polar IAA transport. *New Phytologist*, 74(2), pp 163–172.

Rebernig, C. A., Lafon-Placette, C., Hatorangan, M. R., Slotte, T. & Köhler, C. (2015). Non-reciprocal Interspecies Hybridization Barriers in the Capsella Genus Are Established in the Endosperm.(Bomblies, K., Ed) *PLoS Genetics*, 11(6), p e1005295.

Riechmann, J. L., Krizek, B. A. & Meyerowitz, E. M. (1996). Dimerization specificity of Arabidopsis MADS domain homeotic proteins APETALA1, APETALA3, PISTILLATA, and AGAMOUS. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), pp 4793–8 National Academy of Sciences.

Robert, H. S., Park, C., Gutièrrez, C. L., Wójcikowska, B., Pěnčík, A., Novák, O., Chen, J., Grunewald, W., Dresselhaus, T., Friml, J. & Laux, T. (2018). Maternal auxin supply contributes to early embryo patterning in Arabidopsis. *Nature Plants*, 4(8), pp 548–553 Nature Publishing Group.

Robinson-Beers, K., Pruitt, R. E. & Gasser, C. S. (1992). Ovule Development in Wild-Type Arabidopsis and Two Female-Sterile Mutants. *The Plant Cell*, 4(10), pp 1237–1249.

Rodrigues, J. A. & Zilberman, D. (2015). Evolution and function of genomic imprinting in plants. *Genes & development*, 29(24), pp 2517–31 Cold Spring Harbor Laboratory Press.

Ron, M., Saez, M. A., Williams, L. E., Fletcher, J. C. & McCormick, S. (2010). Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. *Genes and Development*, 24(10), pp 1010–1021.

Roszak, P. & Köhler, C. (2011). Polycomb group proteins are required to couple seed coat initiation to fertilization. *PNAS*, 108(51), pp 20826–31.

Roth, M., Florez-Rueda, A. M. & Städler, T. (2018). Differences in effective ploidy as drivers of genome-wide endosperm expression asymmetries and seed failure in wild tomato hybrids. *bioRxiv*, pp 1–38 Cold Spring Harbor Laboratory.

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Després, B., Drevensek, S., Barneche, F., Dèrozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M.-L., Robin, S., Caboche, M. & Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO journal*, 30(10), pp 1928–38 EMBO Press.

Rubery, P. H. & Sheldrake, A. R. (1974). Carrier-mediated auxin transport. *Planta*, 118(2), pp 101–121 Springer-Verlag.

Sargant, E. (1900). Recent work on the results of fertilization in angiosperms. *Annals of Botany*, os-14(4), pp 689–712.

Schaller, G. E., Bishopp, A. & Kieber, J. J. (2015). The yin-yang of hormones: cytokinin and auxin interactions in plant development. *The Plant cell*, 27(1), pp 44–63 American Society of Plant Biologists.

Schatlowski, N., Köhler, C. & Kohler, C. (2012). Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. *Journal of Experimental Botany*, 63(17), pp 6059–6067.

Schmid, M. W., Schmidt, A. & Grossniklaus, U. (2015). The female gametophyte: an emerging model for cell type-specific systems biology in plant development. *Frontiers in Plant Science*, 6, p 907 Frontiers.

Schneitz, K., Hülskamp, M. & Pruitt, R. E. (1995). Wild-type ovule development in Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue. *The Plant Journal*, 7(5), pp 731–749 John Wiley & Sons, Ltd (10.1111).

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.-S., Park, G. T., Fischer, R. L. & Tamaru, H. (2011). Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. *Proceedings of the National Academy of Sciences of the United States of America*, 108(19), pp 8042–7 National Academy of Sciences.

Scott, R. J., Spielman, M., Bailey, J. & Dickinson, H. G. (1998). Parent-of-origin effects on seed development in Arabidopsis thaliana. *Development*, 125(17), pp 3329–41.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F.,

Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Möst, M., Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M. & Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), pp 176–192.

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 Journal*, 76(5), pp 792–799 John Wiley & Sons, Ltd (10.1111).

Shirley, B. W. (2008). Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis. *Seed Science Research*, 8(04).

Shore, P. & Sharrocks, A. D. (1995). The MADS-Box Family of Transcription Factors. *European Journal of Biochemistry*, 229(1), pp 1–13 John Wiley & Sons, Ltd (10.1111).

Slotkin, R. K. & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 8(4), pp 272–285 Nature Publishing Group.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A. & Martienssen, R. A. (2009). Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell*, 136(3), pp 461–472 Cell Press.

Smaczniak, C., Immink, R. G. H., Muiño, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Liu, S., Westphal, A. H., Boeren, S., Parcy, F., Xu, L., Carles, C. C., Angenent, G. C., Kaufmann, K., Dinh, Q. D. P., Liu, S., Westphal, A. H., Boeren, S., Parcy, F., Xu, L., Carles, C. C., Angenent, G. C. & Kaufmann, K. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), pp 1560–5 National Academy of Sciences.

Smaczniak, C., Muiño, J. M., Chen, D., Kaufmann, K. & Angenent, G. C. (2017). Differences in DNA Binding Specificity of Floral Homeotic Protein Complexes Predict Organ-Specific Target Genes. *The Plant Cell*, 29(8), pp 1822–1835.

Song, X., Yuan, L. & Sundaresan, V. (2014). Antipodal cells persist through fertilization in the female gametophyte of Arabidopsis. *Plant reproduction*, 27(4), pp 197–203.

Spillane, C., Schmid, K. J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., Gagliardini, V., Page, D. R., Wolfe, K. H. & Grossniklaus, U. (2007). Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature*, 448(7151), pp 349–352 Nature Publishing Group.

Sridhar, V. V, Surendrarao, A. & Liu, Z. (2006). APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. *Development (Cambridge, England)*, 133(16), pp 3159–66 Oxford University Press for The Company of Biologists Limited.

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

Stoute, A. I., Varenko, V., King, G. J., Scott, R. J. & Kurup, S. (2012). Parental genome imbalance in Brassica oleracea causes asymmetric triploid block. *Plant Journal*, 71(3), pp 503–516 John Wiley & Sons, Ltd (10.1111).

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P. & Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*, 24(12), pp 1963–1976.

Susaki, D., Nagahara, S., Kasahara, R. D., Maruyama, D., Higashiyama, T., Notaguchi, M., Suzuki, T. & Honma, Y. (2016). Pollen tube contents initiate ovule enlargement and enhance seed coat development without fertilization. *Science Advances*, 2(10), p e1600554.

Swarup, K., Benková, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., Levesque, M. P., Carrier, D., James, N., Calvo, V., Ljung, K., Kramer, E., Roberts, R., Graham, N., Marillonnet, S., Patel, K., Jones, J. D. G., Taylor, C. G., Schachtman, D. P., May, S., Sandberg, G., Benfey, P., Friml, J., Kerr, I., Beeckman, T., Laplaze, L. & Bennett, M. J. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nature Cell Biology*, 10(8), pp 946–954 Nature Publishing Group.

Theissen, G. & Gramzow, L. (2016). Structure and Evolution of Plant MADS Domain Transcription Factors. *Plant Transcription Factors*. pp 127–138. ISBN 9780128008546.

Tiwari, S., Spielman, M., Schulz, R., Oakey, R. J., Kelsey, G., Salazar, A., Zhang, K., Pennell, R. & Scott, R. J. (2010). Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. *BMC Plant Biology*, 10(1), p 72 BioMed Central.

Tonosaki, K., Osabe, K., Kawanabe, T. & Fujimoto, R. (2016). The importance of reproductive barriers and the effect of allopolyploidization on crop breeding. *Breeding Science*, 66(3), pp 333–349 Japanese Society of Breeding.

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 Journal*, 93(3), pp 534–544.

Tuteja, R., McKeown, P. C., Ryan, P., Morgan, C. C., Donoghue, M. T. A., Downing, T., O'Connell, M. J. & Spillane, C. (2019). Paternally expressed imprinted genes under positive Darwinian selection in Arabidopsis thaliana.(de Meaux, J., Ed) *Molecular Biology and Evolution*,.

Twell, D. (2011). Male gametogenesis and germline specification in flowering plants. *Sexual Plant Reproduction*, 24(2), pp 149–160.

Villar, C. B. R., Erilova, A., Makarevich, G., Trösch, R. & Köhler, C. (2009). Control of PHERES1 Imprinting in Arabidopsis by Direct Tandem Repeats. *Molecular Plant*, 2(4), pp 654–660 Cell Press.

de Vries, S. C. & Weijers, D. (2017). Plant embryogenesis. *Current Biology*, 27(17), pp R870–R873 Cell Press.

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. *Current Biology*, 19(13), pp 1128–1132 Cell Press.

Wang, G., Jiang, H., Del Toro de León, G., Martinez, G. & Köhler, C. (2018). Sequestration of a Transposon-Derived siRNA by a Target Mimic Imprinted Gene Induces Postzygotic Reproductive Isolation in Arabidopsis. *Developmental Cell*, 46(6), pp 696-705.e4 Cell Press.

Waters, A. J., Bilinski, P., Eichten, S. R., Vaughn, M. W., Ross-Ibarra, J., Gehring, M. & Springer, N. M. (2013). Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proceedings of the National Academy of Sciences*, 110(48), pp 19639–19644 National Academy of Sciences.

Weijers, D. (2003). Diphtheria Toxin-Mediated Cell Ablation Reveals Interregional Communication during Arabidopsis Seed Development. *PLANT PHYSIOLOGY*, 133(4), pp 1882–1892.

Weijers, D. & Wagner, D. (2016). Transcriptional Responses to the Auxin Hormone. *Annual Review of Plant Biology*, 67(1), pp 539–574.

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L. & Köhler, C. (2010). H3K27me3 profiling of the

endosperm implies exclusion of polycomb group protein targeting by DNA methylation.(Kakutani, T., Ed) *PLoS Genetics*, 6(10), pp 1–14 Public Library of Science.

Windsor, J. B., Symonds, V. V., Mendenhall, J. & Lloyd, A. M. (2000). Arabidopsis seed coat development: morphological differentiation of the outer integument. *The Plant Journal*, 22(6), pp 483–493 John Wiley & Sons, Ltd (10.1111).

Wisniewska, J., Xu, J., Seifertová, D., Brewer, P. B., Ruzicka, K., Blilou, I., Rouquié, D., Benková, E., Scheres, B. & Friml, J. (2006). Polar PIN localization directs auxin flow in plants. *Science (New York, N.Y.)*, 312(5775), p 883 American Association for the Advancement of Science.

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(September 2015).

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M. T. A., Spillane, C., Nordborg, M., Rehmsmeier, M. & Köhler, C. (2011). High-Resolution analysis of parent-of-origin allelic expression in the Arabidopsis endosperm.(Kakutani, T., Ed) *PLoS Genetics*, 7(6), p e1002126 Public Library of Science.

Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., Kasahara, H., Kamiya, Y., Chory, J. & Zhao, Y. (2011). Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45), pp 18518–23.

Woo, H. R., Dittmer, T. A. & Richards, E. J. (2008). Three SRA-Domain Methylcytosine-Binding Proteins Cooperate to Maintain Global CpG Methylation and Epigenetic Silencing in Arabidopsis.(Kakutani, T., Ed) *PLoS Genetics*, 4(8), p e1000156 Public Library of Science.

Woodward, A. W. & Bartel, B. (2005). Auxin: regulation, action, and interaction. *Annals of Botany*, 95(5), pp 707–35.

Wright, A. D., Sampson, M. B., Neuffer, M. G., Michalczuk, L., Slovin, J. P. & Cohen, J. D. (1991). Indole-3-acetic acid biosynthesis in the mutant maize orange pericarp, a tryptophan auxotroph. *Science*, 254(5034), pp 998–1000 American Association for the Advancement of Science.

Yadegari, R. & Drews, G. N. (2004). Female Gametophyte Development. *Plant Cell*, 16(suppl_1), pp S133–S141.

Yamamuro, C., Zhu, J.-K. & Yang, Z. (2016). Epigenetic Modifications and Plant Hormone Action. *Molecular Plant*, 9(1), pp 57–70 Cell Press.

Yan, D., Duermeyer, L., Leoveanu, C. & Nambara, E. (2014). The Functions of the Endosperm During Seed Germination. *Plant and Cell Physiology*, 55(9), pp 1521–1533 Oxford University Press.

Yan, W., Chen, D. & Kaufmann, K. (2016). Molecular mechanisms of floral organ specification by MADS domain proteins. *Current Opinion in Plant Biology*, 29, pp 154–162 Elsevier Current Trends.

Yang, Y., Hammes, U. Z., Taylor, C. G., Schachtman, D. P. & Nielsen, E. (2006). High-affinity auxin transport by the AUX1 influx carrier protein. *Current Biology*, 16(11), pp 1123–7.

Zemach, A., Kim, M. Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L. & Zilberman, D. (2013). The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell*, 153(1), pp 193–205 Cell Press.

Zhang, H., Lang, Z. & Zhu, J.-K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, 19(8), pp 489–506 Nature Publishing Group.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W.-L., Chen, H., Henderson, I. R., Shinn, P.,

Pellegrini, M., Jacobsen, S. E. & Ecker, J. R. (2006). Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. *Cell*, 126(6), pp 1189–1201 Cell Press.

Zhao, Y. (2018). Essential Roles of Local Auxin Biosynthesis in Plant Development and in Adaptation to Environmental Changes.

Zhou, K., Gaullier, G. & Luger, K. (2019). Nucleosome structure and dynamics are coming of age. *Nature Structural and Molecular Biology* Nature Publishing Group.

Zhou, L. zi & Dresselhaus, T. (2018). Friend or foe: Signaling mechanisms during double fertilization in flowering seed plants. *Current Topics in Developmental Biology* pp 453–496.

Zhou, Y., Zhao, X., Ni, M., Zhang, X. & Kang, X. (2009). SHORT HYPOCOTYL UNDER BLUE1 Associates with MINISEED3 and HAIKU2 Promoters in Vivo to Regulate Arabidopsis Seed Development. *THE PLANT CELL ONLINE*, 21(1), pp 106–117.

Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology*, 18(1), p 87.

Popular science summary

Most plants on the Earth's surface produce seeds, and the global plant biodiversity is tightly linked to the existence of these structures. Seeds are a "starter-kit" where all the resources to form a new plant are neatly packed: the embryo, which will become the new plant; the endosperm, which provides energy to the embryo; and the seed coat, which surrounds and protects the embryo and the endosperm. The embryo and the endosperm are derived from fertilisation, meaning they are formed through the union of the maternal and paternal gametes. On the other hand, the seed coat is formed from a maternal tissue, which is not fertilised. For a seed to develop successfully, the embryo, the endosperm, and the seed coat need to synchronise their growth, so that they reach maturity at the same time. If this fails, the seed is not functional, and will abort before it can generate a new plant. In this study, we found that the hormone auxin is the switch that turns on seed development after fertilisation. This hormone is produced in the endosperm, and promotes its growth. Simultaneously, auxin is transported to the seed coat, initiating its development. Interestingly, the genes responsible for auxin production are active in the paternal genome, but switched off in the maternal genome. Consequently, the presence of the paternal genome is required for seed growth, thus explaining how fertilisation triggers seed development. Many cereals and vegetables used today are hybrid plants, meaning they derived from crosses between two plants of different species or varieties. Hybrids are very useful due to their superior yield and high stress resistance. However, obtaining them is not always easy, since hybrid seeds often abort. We found that this seed abortion is explained by abnormally high levels of auxin in the endosperm and seed coat, which trigger uncontrolled growth. Furthermore, by decreasing auxin levels in these seeds, we could prevent their abortion, thus allowing the production of viable hybrid plants. Together, these results show that auxin is a central regulator of endosperm and seed coat growth, and as such, it has the unique ability to control the fate of seed development.

Populärvetenskaplig sammanfattning

Frön produceras av de flesta växter på jordens yta och den globala växtbiodiversiteten är tätt förknippad med dess existens. Frön utgör ett ”startpaket” där alla resurser för att skapa en ny växt är snyggt förpackade: embryot som ska bli den nya plantan; endospermet som förser embryot med energi; och fröskalet som omger och skyddar embryo och endosperm. Embryo och endosperm kommer av befruktning, vilket betyder att de bildas genom förening av honliga och hanliga könsceller. Fröskalet bildas dock av honlig vävnad som inte befruktas. För att ett frö ska utvecklas framgångsrikt, måste embryo, endosperm och fröskal synkronisera sin tillväxt så de mognar samtidigt. Om detta misslyckas fungerar inte fröet och kommer att aborteras innan det kan generera en ny planta. I denna studie fann vi att ett växthormon, ”auxin”, sätter igång fröutveckling efter befruktning. Auxin bildas i endospermet efter befruktning där det stimulerar tillväxt. Samtidigt transporteras auxin till fröhöljet för att initiera fröskalets utveckling. Intressant är att de gener som är ansvariga för auxinproduktion är aktiva i det hanliga genomet men avstängt i det honliga. Följaktligen kräver frötillväxt närvaron av hanligt genom, vilket kopplar befruktning till fröutveckling. Många av våra nutida grödor är hybrider, men att få fram dem är ofta en utmaning inom jordbruket. Hybrider är resultatet av korsningar mellan olika arter eller sorter, och är mycket användbara tack vare hög avkastning och hög stresstålighet. Korsningarna bildar dock ofta aborterade frön, vilket hindrar en ny generation av hybridplantor. Vi fann att abortering av hybridfrön förklaras av onormalt höga nivåer av auxin i endosperm och fröskal, vilket stimulerar okontrollerad tillväxt. Vidare kunde vi förhindra abort av dessa frön genom att minska auxinnivåerna och därmed få vitala hybridplantor. Tillsammans vissar dessa resultat att auxin är viktig för endosperm- och fröskaltillväxt och har en unik förmåga att påverka fröutveckling.

Sumário de divulgação científica

A maior parte das plantas existentes à face da Terra produzem sementes, e a biodiversidade observada hoje em dia é em parte explicada pela existência destas estruturas. As sementes são um kit básico onde estão presentes todos os recursos necessários à formação de uma nova planta: o embrião, que dará origem à nova planta; o endosperma, responsável por nutrir o embrião; e a capa da semente, que envolve e protege o embrião e o endosperma. A fertilização, ou união entre o gâmeta feminino e masculino, formam o embrião e o endosperma. Por sua vez, a capa da semente é uma estrutura materna que não é fertilizada. A correcta formação de uma semente requer o desenvolvimento coordenado do embrião, do endosperma, e da capa da semente. Se isto falhar, a semente abortará antes de poder gerar uma nova planta. Neste estudo descobrimos que a hormona auxina é responsável por activar o desenvolvimento da semente. Esta hormona é produzida no endosperma, transportada para a capa da semente, e tem a capacidade de estimular o crescimento destas duas estruturas. Curiosamente, os genes responsáveis pela produção de auxina estão activos no genoma paterno, mas desligados no genoma materno. Desta forma, explica-se o porquê de o genoma paterno ser necessário para o crescimento da semente. Muitos cereais e vegetais consumidos hoje em dia são híbridos que provêm do cruzamento entre duas plantas de espécies ou variedades diferentes. Geralmente estas plantas têm um rendimento superior e são mais resistentes a stresses ambientais. Contudo, nem sempre é fácil obtê-las, já que os cruzamentos que as originam frequentemente resultam em sementes abortadas. Neste estudo descobrimos que as sementes híbridas abortam devido a níveis elevados de auxina no endosperma, o que causa o seu crescimento descontrolado. Observámos também que a redução dos níveis de auxina nestas sementes viabiliza a sua formação, sendo assim possível obter plantas híbridas. Desta forma, estes resultados demonstram que a auxina tem um papel central na regulação do desenvolvimento da capa da semente e do endosperma e, como tal, tem a capacidade única de controlar o destino de uma semente.

Acknowledgements

This PhD was a life-changing voyage. One that made me grow as a scientist, but most of all, as a person. During this time, I had the opportunity to experience life in a different country, get to know a new culture, travel around the world, meet fun and interesting people, and make friends for life. At times, this voyage was very challenging, but I had the great luck of being surrounded by so many fantastic people who, each in their own way, helped me reach the destination.

To everyone – a big thank you!

First and foremost, I would like to thank my supervisor Claudia Köhler, for giving me the opportunity to do this PhD. Your motivation, passion for science, and immense knowledge are exemplary, and contributed a lot for this thesis, as well as for my development as a scientist.

I am grateful to Eva Sundberg and Mattias Thelander for being part of my PhD evaluation committee throughout these years. Thank you for your advice and diligence.

I would like to express my gratitude to all the previous and current members of the Plant Biology department for making this workplace so enjoyable and stimulating. A special thank you to all the current and former members of the Köhler and Hennig groups for the exciting discussions, and fruitful cooperation.

I would like to thank the administrative personal at the Plant Biology Department, especially to Marie for all the help and the support in the lab, and to Björn for always handling my IT mishaps with a smile.

I am forever grateful to Cecilia for her outstanding help in the lab, and for her humorous and kind personality. I truly enjoyed all our conversations,

especially those that made the never-ending hours at the phytotron go by a little faster. An additional thank you for the translation of the Swedish texts that feature in this thesis. Och ett stort tack för din hjälp med att förbättra min svenska! Jag hopas att språkpolisen godkänner dessa meningar… ☺

I would also like to acknowledge and thank all the co-authors involved in the papers generated during this PhD. A special thanks to Juan for the helpful bioinformatics discussions, and to Jordi, whose tremendous expertise in ChIP-seq was determinant for this thesis.

During my PhD I had the great pleasure to teach and supervise several students. I would like thank them for this learning experience, and for helping me discover a passion for teaching and supervising. I am especially grateful to Monica, Jan, and Charlotte, whose dedicated work helped move this thesis forward.

I am grateful for the company of all the people with whom I shared an office throughout these years. A special thanks to Christina, Vicky, Suayib, and Stefan for the interesting and fun conversations, as well as for all the support.

A huge thank you to all the people who have made life outside work so pleasant and interesting: Adrien, Agneta, Agustín, Alyona, Bart, Bengt, Carolin, Catarina, Chen, Eric, German, Hanneke, Iva, Jenny, Jordi, Martin, Matheus, Minerva, Nici, Stefan, Stefanie, Suayib, Tomas, and Ulrike. I truly enjoyed all the times we spent together, and hope there will be many more of these in the future!

Anis, life becomes more colourful, fun, and exciting whenever you are around! I am very happy that our lives crossed, and very grateful for our interesting discussions, for your support, and for all the moments we shared together.

Miyuki, you always have wise and insightful words to offer. Thank you for your friendship, for your support, and for all the fun times we shared throughout these years. I look forward to meeting you again in Japan! (ˆ ω ˆ)

Rocky, you are an endless source of fantastic stories and discussions. I really enjoyed our crazy conversations, and all the fun times we spent together, inside and outside of the lab. Thank you for always making our brains spin, and our bellies hurt from laughter!

Duarte, you had many different contributions for my PhD – as a supervisor, as a co-author, and most of all, as a friend. Thank you for fulfilling each of these roles perfectly throughout these years. A special thank you for the helpful comments during the writing of this thesis. Much of the work featured here is yours, and much of my development as a scientist is also your work. I am truly grateful for all your support, your friendship, and all the things I learned from you!

Clément, one of the best things that came from this PhD was to have met you. Thank you for your scientific input during all these years - our discussions helped shape many of my PhD projects. I am also grateful for the helpful comments during the writing of this thesis. Outside work, your friendship and support held me together during the hardest times, and made the good times even better. You challenged me to grow and develop as a person, and I am deeply grateful for that. I will forever cherish the moments we spent together, and look forward to enjoying our companionship for many more years to come!

E finalmente, à minha família, o maior obrigado! Sem o vosso apoio incondicional tudo isto seria impossível.

Aos meus avós um obrigado por me receberem sempre de braços abertos, por todo o suporte e carinho, por todos os preciosos ensinamentos, e por todos os momentos que partilhámos.

Mana, mesmo quando a vida é difícil, tu és um farol de alegria e carinho. Obrigada por todas a gargalhadas, por todos os abracinhos, e por me animares sempre com o teu jeito especial.

Mãe, a cada surpresa que a vida te atira tu respondes com determinação e astúcia. Admiro muito a tua força, gentileza e bondade. Obrigada por estares sempre presente para ouvir e ajudar.

Pai, és um exemplo de determinação, competência e resiliência. Desde nova aprendi contigo a lutar por aquilo em que acredito, e a nunca desistir perante as dificuldades da vida. Obrigada por estas preciosas lições.

Todos vocês me inspiram a ser e a fazer melhor. Estimarei sempre tudo o que aprendi convosco, e agradeço muito por me tornarem na pessoa que sou hoje. Obrigado pela vossa amizade e amor!