UNIVERSITÀ DEGLI STUDI DI MILANO

Scuola di Dottorato in Scienze Biologiche e Molecolari

XXVI Ciclo

MOLECULAR CONTROL OF SEED SIZE

Dario Paolo

PhD Thesis

Scientific tutor: Prof.ssa Lucia Colombo

Academic year: 2013-2014

SSD: BIO/01; BIO/18

Thesis performed at Dipartimento di BioScienze – Università degli Studi di Milano

Contents

Part I

Abstract		3
State of the art		4
1	The importance of seeds	4
1.1	Seeds and reproduction	4
1.2	Why study seed size regulation?	5
2	Seed development	9
2.1	The life cycle of flowering plants	9
2.2	Ovule formation	10
	Primordia formation and megasporogenesis	10
	Megagametogenesis	12
	Integuments formation	13
2.3	Double fertilization	16
2.4	Embryogenesis and endosperm development	17
2.5	Seed coat development	19
2.6	The role of *SEEDSTICK* during ovule and seed development	22
3	Molecular control of proanthocyanidins accumulation	27
4	Molecular control of seed size in Arabidopsis	29
Aim of the project		31
Main results		31
Conclusions and future prospects		32
References		34

Part II

“SEEDSTICK is a master regulator of development and metabolism in the Arabidopsis seed coat” 45

“Networks controlling seed size in Arabidopsis” 99

Part III

“The genetic network involving *SEEDSTICK* and *AUXIN RESPONSE FACTOR 2* in the control of seed size in *Arabidopsis thaliana*” 153

Eadem est, inquam,
praeceptorum condicio quae seminum:
multum efficiunt, et angusta sunt.

I precetti, a mio parere,
sono come i semi:
danno grossi risultati, eppure sono piccola cosa.

Seneca, *Lettere a Lucilio*, IV-38

Part I

Abstract

Seed derive from fertilized ovules and they constitute complex structures formed by three different parts: the seed coat, and the fertilization products, the endosperm and the embryo. During my Ph.D. I analyzed the role of MADS-domain transcription factor *SEEDSTICK (STK)* during the development of the seed coat in the model species *Arabidopsis thaliana*.

Part II of this thesis contains the description of the role of *STK* in the regulation of several metabolic processes in the seed coat, providing evidences for a connection between cell fate determination, development and metabolism. In particular, the focus of the work is on the characterization of the differences, between the wild type and the *stk* mutant, in the synthesis of the proanthocyanidins (PAs) stored in the seed coat. PAs are a large group of compounds that have been previously reported for their role in different aspects of plant physiology, including defence against pathogens, protection of the embryo against UV-radiation and regulation of seed size.

Part III of this thesis is dedicated to the analysis of seed size in *Arabidopsis thaliana*. We analysed the role of *STK* and of *AUXIN RESPONSE FACTOR 2 (ARF2)* in the control of seed coat development and in response to the hormonal cascade of brassinosteroids, known regulators of seed size. We provide evidences that both *STK* and *ARF2* are regulators of cell proliferation and of cell expansion; we show that these two genes act antagonistically to determine seed size and data obtained from the detailed characterization of *stk*, *arf2* and *stk arf2* mutants are used to discuss a novel genetic network, pivotal in the maternal control of seed yield.

State of the art

1. The importance of seeds

1.1 Seeds and reproduction

Inside of the Plant kingdom, Spermatophytes comprehend two of the main groups of plants: the Gymnosperms and the Angiosperms. The word Spermatophytes derives from the Greek term Σπερματόφυτα, meaning seed plants, and indicate the plants that rely on seeds for their reproduction, a characteristic that strongly determined their evolutionary success since their appearance on the planet approximately 320 millions of years ago. The presence of seeds marks in fact the main difference of Spermatophytes to more primitive plants such as mosses, liverworts and ferns, representing a crucial reproductive innovation that has been associated with adaptive radiation and colonization of new environments and ecological niches on land, from forests to grasslands both in hot and cold climates. The evolutionary success of Spermatophytes is well explained by the fact that they comprehend an estimated 89% of total plant species present on earth, a number that is mainly influenced by the total number of Angiosperms species (between 300.000 and 400.000), where the seed dispersal is due to the presence of the fruit (Crepet, 2000).

Among these plants the seed is required for the development of the offspring, the embryo, and it acts as a functional unit of nutrition, with a progressively lower metabolism during maturation, of protection against physical and chemical stresses, of dispersion and of germination, an event that occur only under favorable condition and determines the arising of the new plant (Nowack et al., 2010)

1.2 Why study seed size regulation?

Of all the fascinating aspects regarding the study of seed, one should mention the analysis of the different mechanisms that determine its size, and in general of what is referred as seed yield of the plant.

It is remarkable and well documented, how seed size can tremendously vary between different species; this happens not only among taxa (seed weight in the orchid *Corallorhiza maculate* is approximately 1µg, while it is usually around 10 kg in *Lodoicea maldivica*) (Moles et al., 2005), but also within tax, as shown for example for the genus Solanum, with a 12-times increase in seed weight between *Solanum pennellii* and *Solanum lycopersicum*, the cultivated tomato (Orsi and Tanksley, 2009) (Figure 1).

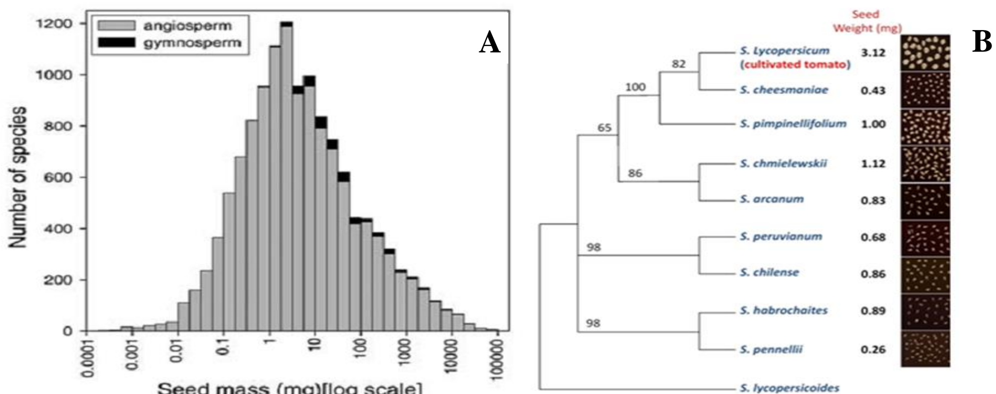


Figure 1: A) Seed mass variation between Spermatophytes (from Moles et al., 2005); B) Seed mass variation within the genus Solanum (adapted from Orsi and Tanksley, 2009).

One of the reasons that could better explain why this morphological trait has undergone such a wide evolutionary radiation in plants is related to the direct effect that seed size plays on plant fitness. Several studies have in fact

previously suggested that larger seeds have a higher probability of germination (Jacobson et al., 1998; Westoby et al., 2002) and that the seedlings of large-seed plants are better able to tolerate many of the stresses encountered during seedling establishment, with a greater survival and improved performance under a wide range of environmental conditions (Albertine and Manning, 2009; Westoby et al., 1996). On the contrary, as the resources from the mother plant are limited, plants with smaller seeds have higher colonization abilities; this relates to the possibility of producing a large number of seeds, as it happens for example in the model organism *Arabidopsis thaliana* (Krannitz et al., 1991; Moles et al., 2005; Westoby et al., 2002). Taken together this considerations confirm that seed size deeply influence plant fitness and that, therefore, it is a trait subjected to a strong selection depending on the environment; this produced a vast range of variation, whose origin is yet to be completely understood and that results in a puzzling and stimulating scenario to investigate on (Silvertown, 1989).

Another reason of interest in the study of seed size regulation is the strong dependence that humanity has on plants in terms of nutrition. Our diet has always been based on the photosynthetic activity of plants, that convert sunlight into usable chemical energy. Although roots, tubers, leaves and fruits are used for human nutrition, in every continent human and animal diet is mainly based on calories derived from harvested mature seeds, from cultivated species that are commonly known as grain-crops (Berger, 2003). The two main types of commercial grain-crops are cereals, such as wheat and rice, and legumes, such as beans and soybeans. The nutritional quality of these seeds is represented by their rich content of carbohydrates, oil and proteins, stored as nutritional resource for the embryo in a tissue called the endosperm (Focks and Benning, 1998). Moreover, grain-crops also present agronomic and commercial

advantages as their harvested and dried seeds are durable and stable in time, a characteristic that made them well suited to industrial agriculture: they can be mechanically harvested, transported over long distances, stored for long periods in silos, milled for flour or pressed for oil.

For this reason, since the beginning of agriculture, food grains have been subjected to selection and breeding for size, as well as for other qualities, with the result that nowadays crops have usually larger and heavier seeds in comparison to wild-type relatives. (Fan et al., 2006; Orsi and Tanksley, 2009; Song et al., 2007; Sundaresan, 2005; Weng et al., 2008). (Figure 2)

Figure 2: Domestication of wheat led to changes in grain size, shape and to a wide range of phenotypic variation. Grains shown are representative of modern elite varieties (top row) and ancestral wheat species (bottom row) (adapted from Gegas et al., 2010).

Although the increasing number of indications that seed size is controlled by different mechanisms and the fact that this trait has been used and modified by

breeders over the last centuries, only in the last few years we have at last begun to identify molecular regulators of seed size in plants, mainly through studies on model plant Arabidopsis.

Given the importance of seeds for our diet, deciphering the mechanisms behind their development could help the improvement of yield and of food production, a necessity that is becoming increasingly urgent nowadays. Already at the moment in fact, an estimated one in six people cannot receive enough food to conduct a normal life, a situation that makes hunger and malnutrition serious threats to health in many part of the world (WFP, 2009). This dramatic situation is destined to get worse, as world population is predicted to increase to 9,1 billions of people by 2050 (FAO, 2014). The result is that an enormous increase in agricultural yields is required to sustain the growth of population, similarly to what happened with the Green Revolution in the past (Shah and Strong, 2002).

The term Green Revolution refers to a series of research outbreaks, improved crop varieties, and development of new farming technologies that dramatically increased agricultural output in many developing countries between the ‘40s and the ‘70s of last century. This very wide scenario of agricultural events was highly successful at meeting its objective of increasing crop yields and food supplies, so that without it nowadays it would not be possible to sustain, albeit only partially, global population (Evans, 1996; FAO, 2014).

Advances in biotechnology and a deeper understanding of developmental biology of plants might target directly the mechanisms controlling seed yield. Achieving this goal requires a deeper understanding of the molecular basis that control seed development and that determine final seed size.

2. Seed development

2.1 The life cycle of flowering plants

In plants, sexual reproduction requires that the diploid generation, known as the *sporophyte*, produces a special lineage leading to meiosis. The haploid cells that result from the meiosis of cells belonging to this lineage are called *spores* and they mitotically divide and then develop as what are known as *gametophytes*, specifically the male gametophyte and the female gametophyte. The gametophytes will then differentiate a germ-line and produce the male and female *gametes* (Spielman et al., 2001).

In Spermatophytes, the female gametophyte (also called *embryo sac*) differentiates and develops always protected by the sporophyte, in a structure called the *ovule*. In particular in Angiosperms (or flowering plants, such as *Arabidopsis thaliana*), the ovule is further enclosed in a part of the flower named the *ovary*, which develops into the fruit after fertilization.

Double fertilization (discussed below) is the unique and complex fertilization mechanism of flowering plants. This process involves the joining of a female gametophyte with two male gametes (or sperm cells) produced by the male gametophyte, the *pollen grain*. The result is the formation of a *seed*, where the diploid *embryo* and a nourishing tissue called the *endosperm* remain enclosed by the sporophytic tissues, named the *seed coat* (Figure 3).

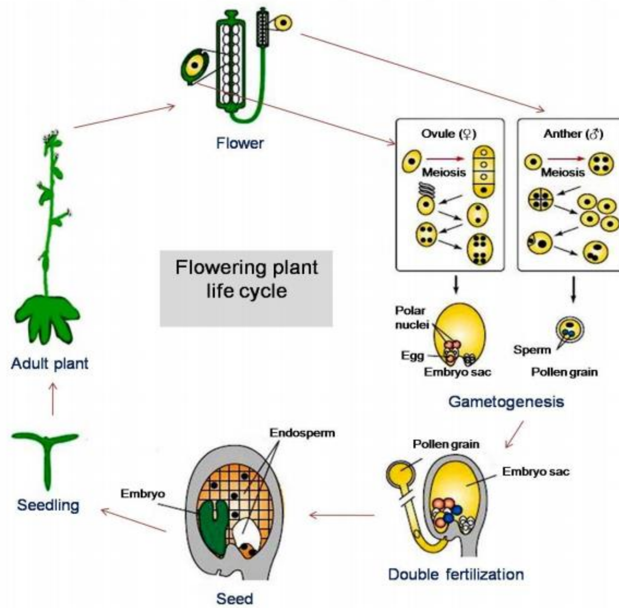


Figure 3: Schematic representation of *Arabidopsis thaliana* life cycle (adapted from Spielman et al., 2001).

2.2 Ovule development

Primordia formation and megasporogenesis

As seeds derive from ovules, the ovule development has to be taken in consideration when describing the control and improvement of seed yield. The ovule represents the structural support and functional unit for female gametophyte development, thus the structure where the sporophytic and the female gametophytic generations coexist.

Ovule development consists mainly of three different stages: primordia formation and differentiation, megasporogenesis and megagametogeneis (Schneitz et al., 1995).

Ovule primordia develop as result of periclinal divisions of the meristematic placental cells inside the ovary. The resulting finger-like primordium differentiates three zones: the *funiculus*, that connects the ovule to the placenta, the *chalaza*, from which the integuments are formed and the *nucellus*, where the Megaspore Mother Cell (MMC) differentiates (Schneitz et al., 1995) (Figure 4). Megasporogenesis terminates when the nucellar MMC completes meiosis; the result of this process is a tetrad of haploid cells: three of them will degenerate, while the remaining one, the *megaspore*, will subsequently originate the female gametophyte (Figure 4 A).

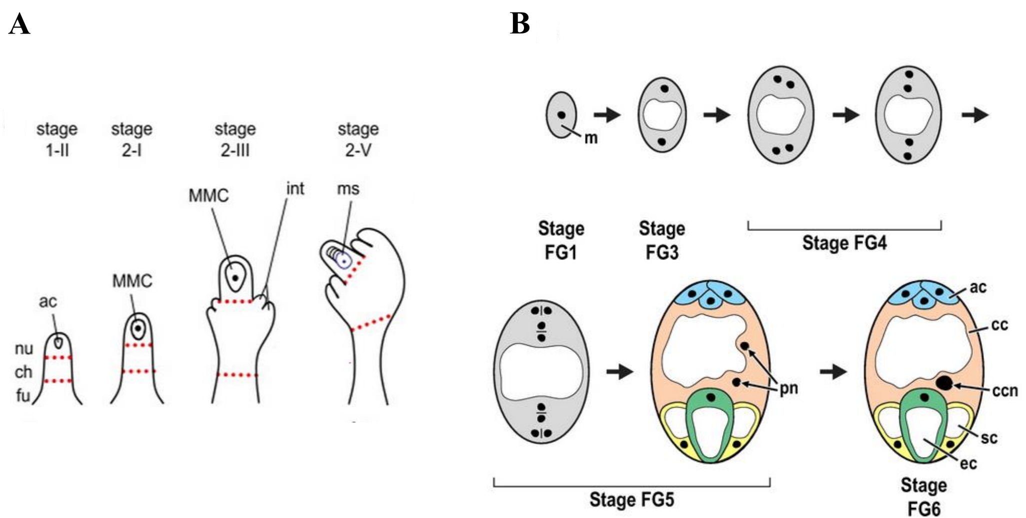


Figure 4: Schematic representation of ovule formation. (adapted from Lieber et al., 2011 and Drews and Koltunow, 2011). A) Ovule primordia formation and megasporogenesis in Arabidopis. fu: funiculus; ch: chalaza; nu: nucellus; ac: archesporial cell; MMC: megaspore mother cell; int: inner and outer integuments; ms: megaspore B) megagametogenesis and embryo sac formation. m: megaspore; pn: polar nuclei; ec: egg cell; sc: synergid cells; cc: central cell; ccn: central cell nucleus; ac: antipodal cells.

The understanding of mechanisms for controlling the correct positioning and number of ovule primordia along the placenta represents a topic of great interest in relation to seed yield; nevertheless, only a few genes describing these processes have been reported. This reduced knowledge on the subject might be explained by the fact that many genes that control ovule development are also pleiotropic developmental regulators of other floral organs, masking their effects on ovules. This is particularly true in the case of a mutations that affect in general pistil / gynoecium and ovary growth. The current state of art on the different pathways controlling ovule number in Arabidopsis was covered in a recent publication (Cucinotta et al., 2014); the authors review the discovery of several molecular mechanisms controlling gynoecium and ovary development, determining an alteration in ovule number. Among them one should consider genetic factors controlling the formation of carpel margin meristem (CMM), that will give rise to the placenta, the differentiation of the boundary regions between the different primordia along the placenta, factors controlling cell proliferation in the pistil, the role of different phytohormones (most notably auxin and cytokinins) and the new perspectives on how ovule number might also be under epigenetic regulation, as it emerges from the phenotypic description of *Arabidopsis thaliana* triploids obtained by crossing a tetraploid *Ler-0* line (used as a male or female parent) with different diploid accessions (Duszynska et al., 2013).

Megagametogenesis

A common pattern of embryo sac development, typical for example of Arabidopsis ovules, is the *Polygonum*-type maturation pattern, where one single cell of the nucellar tetrad, the megaspore, survives after meiosis, originating the female gametophyte (Reise and Fisher, 1993). The megaspore enlarges and

undergoes two rounds of syncytial mitosis: the result of this process is a four-nucleate coenocyte with two nuclei at each pole separated by a large central vacuole (Figure 4B, stage FG4). During the third mitosis, the cellularization process begins and the female gametophyte cells become completely surrounded by cell walls (Figure 4B, stage FG5). At this point, one nucleus from each pole (where *chalazal pole* indicates the proximal one to the placenta, and *micropylar pole* the distal one) migrates towards the center of the developing female gametophyte and fuse together forming the diploid central cell (Christensen et al., 1998; Schneitz et al., 1995). These events result in a seven-celled structure consisting of three haploid antipodal cells, one diploid central cell, two haploid synergid cells, and one haploid egg cell, that is the gamete (Figures 4B, stage FG6). The egg cell and central cell are polarized and their nuclei lie very close to each other (Figures 4B, stage FG6). It is important to mention that the region between the cells of the female gametophyte is devoid of the cell walls or that the cell wall is discontinuous such that the plasma membranes of these cells are in direct contact with each other (Kasahara et al., 2005; Mansfield and Briarty, 1991). This is arrangement is functional for the success of the subsequent double fertilization between the female gametophyte and the two sperm cells carried by the pollen grain.

Integuments formation

An integument is a protective cell layer surrounding the ovule. Gymnosperms ovules typically have one integument (*unitegmic ovules*) while Angiosperms ones typically have two (*bitegmic ovules*).

From a functional point of view, apart from the protection of the developing female gametophyte (and later on of the embryo) against stresses, previous data support the possibility of an interaction between the ovule integuments and the

developing gametophyte, as well as between the seed coat and the fertilization products (Mizzotti et al., 2011; Nowack et al., 2010; reviewed Bencivenga et al., 2011;). The existence of an interaction between ovule integuments and the developing embryo sac has been proposed on the basis of the characterization of sporophytic ovule mutants, which show defects in embryo sac development, in gametophytic cell specification or present altered fertilization (Bencivenga et al., 2011; Mizzotti et al., 2011). This communication between the two generations is likely to involve hormonal signaling of auxin and cytokinins pathways, determining a strong control of the sporophyte over female gametophyte development (Bencivenga et al., 2012; Ceccato et al., 2013), although further studies are required to unveil this crosstalk (Bencivenga et al., 2011).

As mentioned above, the ovule integuments derive from the chalazal region of the primordium, and the morphological description of their development in Arabidopsis has been presented in detail in the past (Robinson-Beers et al., 1992).

The inner integument is initiated through a series of cell divisions in the dermal layer at the chalazal region of the primordium, resulting in the formation of a ring-like belt around the nucellus (Figure 5B). Shortly after that, the outer integument starts developing, through a series of similar cell divisions (Figure 5 C). As the growing ovule enlarges and the funiculus elongates, the ovule begins to exhibit the effects of asymmetric growth: the larger number of cells on the convex side, relative to the concave side, determines that the nucellus and integuments curve forward and the outer integument overtakes the inner integument (Figure 5D-F). At complete integuments growth, although the developing gametophyte is fully protected, a small distal cleft surrounded by elongated cells of the outer integument is left in the distal part: it is called the

micropyle, and it will allow the pollen tube to reach the female gametophyte during fertilization (Figure 5 G-H).

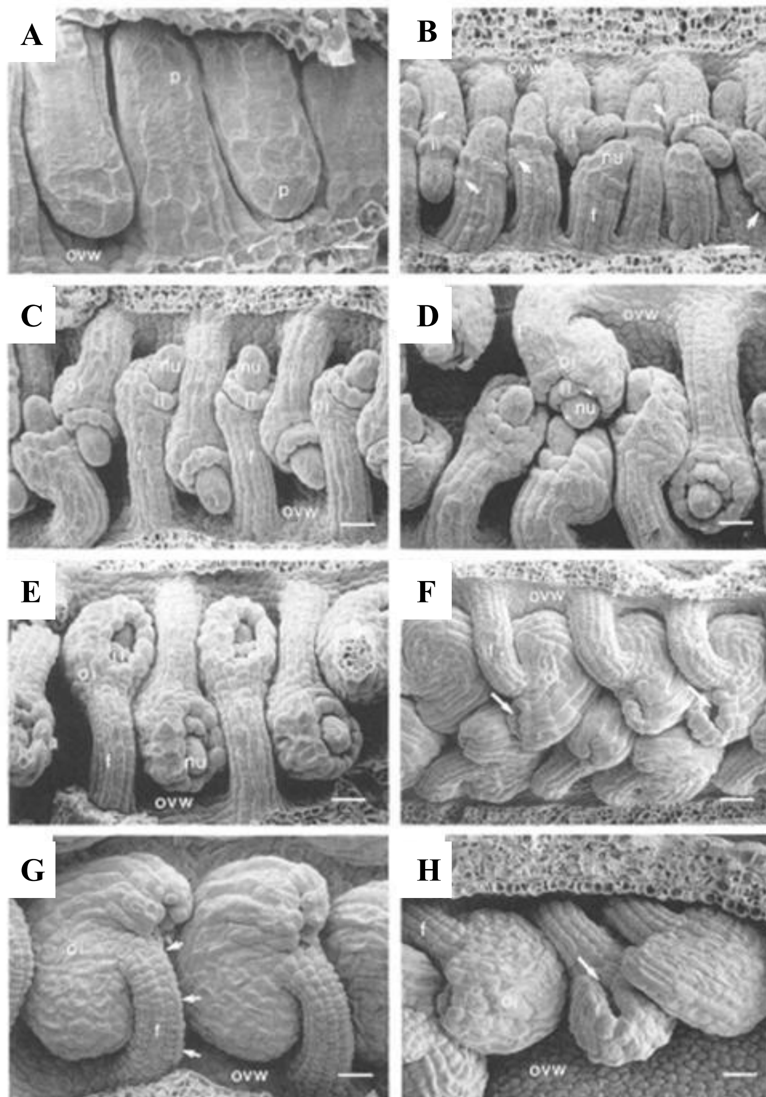


Figure 5: Integuments development in Arabidopsis. A) Elongated ovule primordial and B) ring-belt structure of the forming inner integument; C) initiation of the outer integument; D-F) asimmetric growth and bending of the ovule; G-H) mature ovules, with micropyle. p: ovule

primordium; f: funiculus; ii: inner integument; nu: nucellus; oi: outer integument; owv: ovary wall (modified from Robinson-Beers et al., 1992).

2.3 Double fertilization process

In a flower, the carpel represents the female reproductive organ. A carpel consists of different parts: an apical *stigma*, a *style* and an *ovary*, inside of which the differentiation of the meristematic placenta occurs and, therefore, the place of ovule development.

The function of the stigma is to determine the right condition for the hydration and germination of desiccated pollen grain, each of which will subsequently emit a pollen tube to deliver the two sperm cells inside the female gametophyte in the ovary. In plants with a dry stigma surface, like Arabidopsis, hydration of foreign pollen is often inhibited (Preuss, 1995). The first steps of pollen tube growth through the style require its interaction with an extra-cellular matrix secreted by transmitting tract cells (Alvarez and Smyth, 1999; Smyth et al., 1990). Moreover, the attraction and guidance of the pollen tube involves several mechanisms, including chemo-attraction, mechanical guidance, growth stimulation, adhesion, re-orientation and competence control that enables the pollen tube to respond to the attraction signal, that is strongly controlled by the female gametophyte itself, mainly by the activity of the two micropylar synergid cells (Dresselhaus, 2006; Higashiyama et al., 1998; Higashiyama et al., 2001; Johnson and Lord, 2006).

After entering the micropyle, the pollen tube stops elongating and bursts, releasing the two sperm cells inside the female gametophyte. The penetration of the female gametophyte occurs at the level of one of the synergid cells, that undergoes apoptosis before or upon pollen tube arrival. Immediately after these

events, one of the two male gametes fuses with the egg cell, while the other fuses with the diploid central cell. The result of the double fertilization, after *karyogamy* (fusion of the male and female nuclei) is the formation of a diploid *zygote*, that will originate the embryo, and of a *triploid cell*, that will form the endosperm (van Went and Willemse, 1984; Berger et al., 2008; Russell, 1992; Russell, 1996).

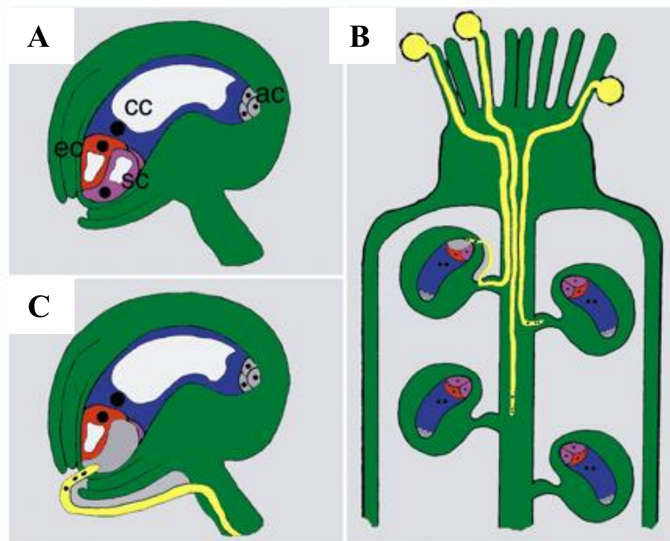


Figure 6: Schematic representation of double fertilization in Arabidopsis. A) Mature ovule with highlighted positions of female gametophyte cells; B) pollen germination on the stigma and attraction of the pollen tubes towards the ovules; C) the pollen tube enters via the embryo sac via the micropyle, degeneration of one synergid cell and release of the male gametes. ec: egg cell; sc: synergid cells; cc: central cell; ac: antipodal cells (adapted from Huck et al., 2003).

2.4 Embryogenesis and endosperm development

One of the product of double fertilization is the diploid zygote, obtained by the fusion of the male gamete (a sperm cell) with the egg cell of the embryo sac. The zygote will proliferate and differentiate the different parts of the embryo, representing the new generation and the conclusion of the plant life cycle. The

processes relative to the formation of the embryo are collectively known as embryogenesis, that itself can be divided in an earlier morphogenesis phase followed by maturation (West and Harada, 1993).

Following fertilization, the zygote divides asymmetrically, originating two cells: a smaller *apical* one, and a larger *basal* cell. This asymmetric division is crucial in the morphogenesis of the embryo, as the smaller apical cell and its daughters will undergo sequential mitosis and differentiation processes that will characterize the apical part of the growing embryo. This cell proliferation phase establishes a sequence of precise embryo morphologies that are used as time points to describe embryo development: among them, two/four celled embryo, octant, globular stage, heart stage (where cotyledons start to appear) and torpedo.

At the same time the large basal cell expands longitudinally and divides transversally to originate the *hypophysis* (that will develop the root system), while the other daughters form the *suspensor*, a transient structure that ensures communication of the embryo with the maternal tissues (Schwartz et al., 1994) (Figure 7).

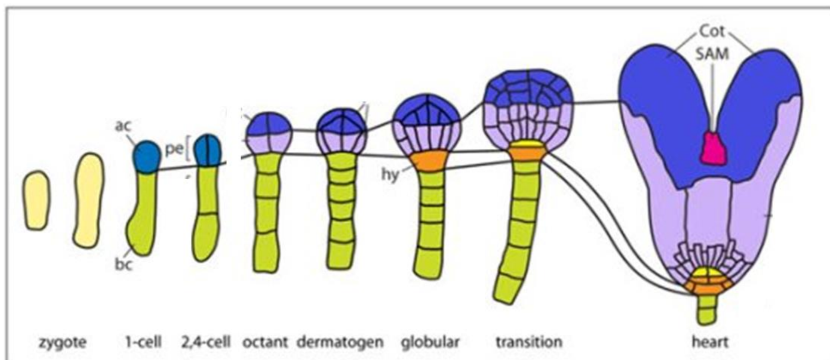


Figure 7: Different stages of embryogenesis in Arabidopsis. ac: apical cell; bc: basal cell; hy: hypophysis; cot: cotyledons (from Peris et al., 2010).

The second fertilization product is the triploid cell that gives rise to the endosperm, the nourishing and energy accumulation tissue characterized by a misbalance of genome dosage in on the maternal side (Lopes and Larkins, 1993).

In Arabidopsis, as in the majority of Angiosperms, the endosperm is initially nuclear, meaning that it derives from a series of syncytial divisions of the first triploid cell, with the consequent fast expansion of the new tissue inside the seed cavity. In Arabidopsis this developmental plan enables the endosperm to grow more rapidly than the embryo, that will only later replace the space occupied by the endosperm, during its maturation.

The other phases that characterize endosperm development are the cellularization, differentiation and death (Berger, 1999). Cellularization starts in the peripheral region previously occupied by the central cell and continues towards the center, so that it is possible to identify three main endospermal regions: peripheral, chalazal and micropylar (Figure 8). In Arabidopsis the mature embryo occupies most of the seed cavity, progressively replacing the differentiated cells of the endosperm (Berger, 1999; Berger, 2003).

2.5 Seed coat development

Fertilization triggers a series of events that lead to differentiation of the ovule integuments into the seed coat (Beeckman et al., 2000; Western et al., 2000).

In Arabidopsis, integuments of a mature ovule consist in five layers of cells, two deriving from the outer integument, and three from the inner one (Haughn and Chaudhury, 2005): these five layers delimit the seed cavity, and they experience a deep morphological and functional changes over the course of

seed maturation, for a period of approximately two weeks from the day of fertilization (Western et al., 2000). In particular, cell proliferation in the integuments peaks after fertilization and an intense mitotic activity is followed by a sharp decline that lead to a complete absence of cell division 4 days after pollination, determining a time window where the size of the seed significantly increases.

The five cell layers (Figure 8) have different functions and they undergo consequent modifications. A common nomenclature of the layers takes in consideration their origin and position in the seed coat (Beeckman et al., 2000): ii1 or *endothelium*, is the inner layer of the inner integument; ii1' and ii2 are the other two layers deriving from the inner integument; from the outer integument of the ovule there is the differentiation of layer oi1 and oi2, with this second being the outmost of the seed coat (Figure 8, Beeckman et al., 2000).

Cells of the endothelium, typically characterized by bigger size in comparison to the other layers (Haughn and Chaudhury, 2005), synthesize flavonoid compounds called proanthocyanidins (PAs, or condensed tannins). PAs are accumulated in the central vacuole of endothelium cells during the first week after fertilization, and their later oxidation is responsible for the brown pigment that characterize Arabidopsis seed coat (Debeaujon et al., 2003). On the other hand, the cells of ii1' and ii2 do not appear to differentiate further and are crushed together as the seed develops.

Finally the outer layers, oi1 and oi2, accumulate starch-containing amyloplasts (Windsor et al., 2000). The oi1 then produces a thickened cell wall and cells of the oi2, also called the *epidermis*, synthesizes and secretes mucilage, that accumulates in a donut-shaped depression around a cytoplasmatic column of these cells. Later on, when secondary cell wall is deposited, it completely fills

the space occupied by the cytoplasmic column, forming the *columella*. At maturity, the mucilage dehydrates into a thin, compressed layer.

Different roles for the seed coat have been previously described; among them, one should mention the obvious protective function for the growing embryo against mechanical damage or pathogen attack (Mohamed-Yasseen et al., 1994), UV-protection guaranteed by the endothelial pigments (Debeaujon et al., 2003), the maintenance of dormancy until favourable conditions for the seedling, as well as water uptake during germination (Debeaujon et al., 2000).

The seed coat is also the only part of the seed exclusively of maternal origin and it plays a crucial role in the determination of final seed size, a role that is examined in detail in the other sections of this work.

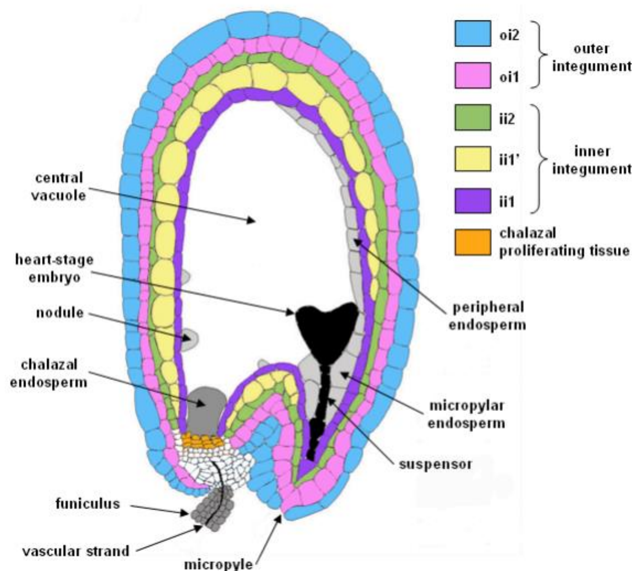


Figure 8: The structure of the Arabidopsis seed. (adapted from Debeaujon et al., 2003, seed coat nomenclature according to Beeckman et al., 2000).

2.6 The role of *SEEDSTICK* during ovule and seed development

The genetic regulation of flower development has been intensively studied over the course of years, with considerable breakthrough discoveries coming from the mutants characterization performed on model organisms such as *Arabidopsis thaliana*, *Petunia hybrida, Antirrhinum majus* and rice (*Oryza sativa*). The determination of the organ identity for flower structure depends on the regulatory network involving numerous genes, mostly belonging to MADS-domain family of transcription factors, therefore playing a pivotal regulatory role throughout plant development. Their combinatory action in identity determination of different flower organs has been explained with the description of the ABCDE model (Theissen, 2001) (Figure 9A).

Inside the carpel, ovule identity depends on the correct expression of the genes belonging to the D-class of MADS-domain genes. This class was first identified in Petunia by the characterization of the role the two genes *FLORAL BINDING PROTEIN7 (FBP7)* and *FBP11*, whose co-suppression results in the formation of carpelloid structures arising from the placenta, instead of ovules (Colombo et al., 1995). The ovule identity determination of the D-class genes is conserved also in other species, as it was observed by the study of the orthologs of *FBP7* and *FBP11* in other species: *SEEDSTICK (STK)* in Arabidopsis and *OsMADS13* in rice (Dreni et al., 2007; Favaro et al., 2003; Pinyopich et al., 2003). In particular, the Arabidopsis triple mutant *stk shp1shp2* (*SHATTERPROOF1* and *SHP2* are orthologs of Petunia *PLENA* and *FPB6*) produces only a limited number of seeds, as most of the ovules lose their identity and are converted in carpelloid structures (Pinyopich et al., 2003) (Figure 9B). Biochemical studies showed that STK, SHP1, SHP2 proteins can form multimeric complexes, an interaction established together with the the SEPALLATA proteins (SEP3), encoded by the genes of E-class genes, known

for their genetic interaction with the A, B and C-class genes (Honma and Goto, 2001; Pelaz et al., 2001).

Beside its role in the ovule identity complex, *STK* was shown also to have other later function in ovule and seed development. The *stk* single mutant does not present homeotic conversions during ovule development, but instead presents fully viable ovules with elongated and ticker funiculus, demonstrating that STK is required for correct seed dispersal, as the seeds in the *stk* mutant tend to remain attached to the fruit septum upon dehiscence of the silique (Pinyopich et al., 2003) (Figure 9 C).

The first reported direct target of MADS-domain ovule identity complex, *VERDANDI (VDD)* is a putative transcription factor that belongs to the plant-specific B3 superfamily. The *vdd* mutant shows defects in the identity of female gametophyte cells and during the fertilization process, resulting in semisterility (Figure 9 E). This demonstrates a role for STK also in gametophyte development and fertilization process (Matias-Hernandez et al., 2010; Mendes et al., 2013).

The *STK* gene also controls, together with *ARABIDOPSIS B-SISTER (ABS)*, the formation of the endothelium, as shown by lack of endothelium of the *stk abs* double mutant. This phenotype is associated with a high sterility, due to both ovule and seed abortions, probably as a result of the enormously increased starch accumulation in *stk abs* mutant in comparison to wild-type, both before and after fertilization (Mizzotti et al., 2011) (Figure 9 F).

Finally, although seed development does not appear altered in the *stk* single mutant, it was already reported that *stk* seeds are smaller than wild type (Pinyopich et al., 2003) (Figure 9 D). In this thesis, a detailed characterization of this phenotype is presented, showing that *STK* acts maternally in determining seed size. This is determined by the regulatory action of *STK* on cell cyle

progression and its positive control of cell expansion in the maternal seed coat (see Part III). Interestingly, the only other member of the B-Sister subclade apart from *ABS* is *GORDITA (GOA)*, a gene for which it has been proposed a negative role in fruit and seed expansion (Erdmann et al., 2010; Prasad and Ambrose, 2010; Prasad et al., 2010).

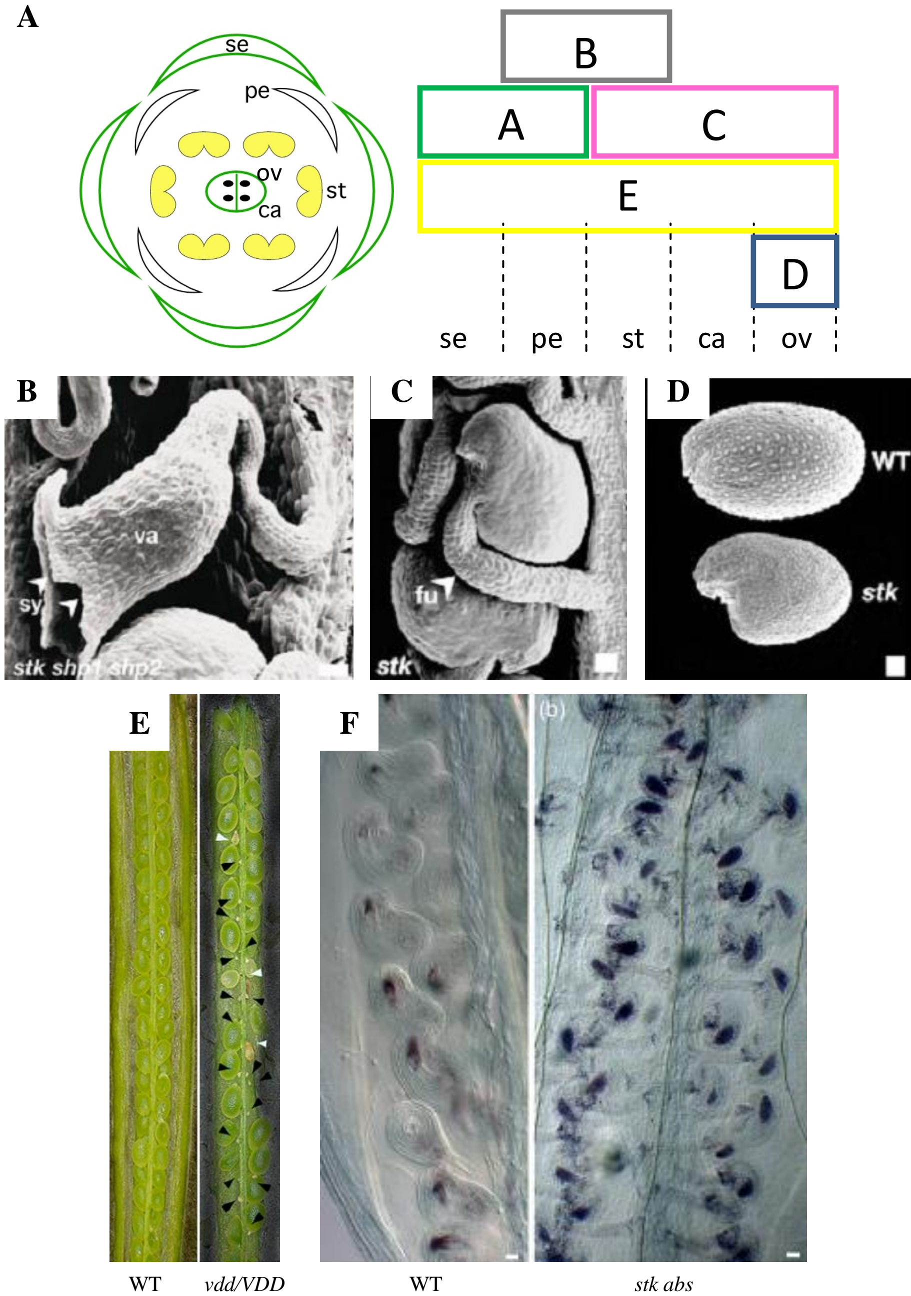


Figure 9: A) Schematic representation of the Arabidopsis flowers and of the ABCDE model. se: sepals; pe: petals; st: stamens; ca: carpel; ov: ovules. B-F) *STK* is required for different steps of ovule and seed development B) *stk shp1 shp2* carpelloid structures; C) enlarged funiculus in

the *stk* mutant; D) comparison between wild-type and *stk* seeds; E) ovule and seed abortions in the *vdd* mutant; F) starch accumulation phenotype of the *stk abs* mutant. (adapted from Matias-Hernandez et al., 2010; Mizzotti et al., 2011; Pinyopich et al., 2003).

3. Molecular control of proanthocyanidins accumulation

As briefly mentioned above (Paragraph 2.5), one of the function of the seed coat is the cellular accumulation of proanthocyanidins (PAs or condensed tannins).

These molecules constitute an astringent, bitter plant polyphenolic compound that binds to and precipitates proteins and various other organic compounds including amino acids and alkaloids; in Arabidopsis seed coat, PAs consist of oligomers of the flavan-3-ol 2,3-cis-epicatechin (Abrahams et al., 2003). PAs are synthesized as colourless polymers in vesicles (known as tanninosomes) that fuse into the central vacuole. The function of PAs in seeds has been previously described for their involvement in different processes from defense against predators and pathogens (Dixon et al., 2005) to seed dormancy increase (Debeaujon et al., 2000) and protection against UV-radiation (Winkel-Shirley, 2002).

Accumulation of PAs starts specifically after fertilization, and it involves the innermost integument protecting the seeds, the endothelium, that will undergo significant changes in morphology for acquiring its specific function (Debeaujon et al., 2003). PAs biosynthesis starts around 1 to 2 days after fertilization (DAF) at the micropyle, and the deposition progresses in the endothelium towards the chalaza until 5 to 6 DAF (Debeaujon et al., 2003; Lepiniec et al., 2006).

The PAs biosynthetic pathway has been described thanks to the characterization of Arabidopsis mutants: *transparent testa (tt)* and *tannin-deficient mutants* produce seeds that lack endothelium pigmentation or present alteration in colours (Abrahams et al., 2003; Winkel-Shirley, 2002).

The PAs biosynthesis and accumulation is under the regulatory action of two different group of genes, the Early (EBGs) and Late (LBGs) Biosynthetic

Genes (Lepiniec et al., 2006; Pelletier et al., 1999; Xu et al., 2014). The EBGs comprise four genes: *CHALCONE SYNTHASE (CHS), CHALCONE ISOMERASE (CHI), FLAVONOL 3-HYDROXYLASE (F3H)* and *FLAVONOL 3'-HYDROXYLASE (F3'H)*. EBGs are involved in the biosynthesis of precursors for PAs and other classes of flavonols.

The LBGs include *DIHYDROFLAVONOL-4-REDUCTASE (DFR), LEUCOANTHOCYANIDIN DIOXYGENASE (LDOX) and BANYULS/ANTHOCYANIDIN REDUCTASE (BAN/ANR). BAN* is considered to be a branching-point enzyme because it is channelling the anthocyanin pathway to flavonoloids.

An additional third group of genes, recently proposed as members of LBGs (Xu et al., 2014) has been identified in Arabidopsis and it comprises *TRANSPARENT TESTA 12* (*TT12)*, *TT10, TT19* and *AHA10 (H+-ATPase)*. These genes are involved in flavan-3-ols modification, transport and oxidation (Baxter et al., 2005; Kitamura et al., 2004; Marinova et al., 2007; Pourcel et al., 2005). The regulation of LBGs occurs via a ternary protein complex, called MBW (MYB-bHLH-WDR), formed by a specific R2R3-MYB, a bHLH transcription factor and the WD repeat protein TRANSPARENT TESTA GLABRA 1 (TTG1) (Lepiniec et al., 2006; Zhao et al., 2013). Moreover, the regulation of LBGs requires the presence of other transcription factors belonging to different families, such as Zinc finger (*TT1/WIP1), MADS (TT16/ABS/AGL32)* and WRKY *(TTG2/DSL1/WRKY44*).

Part II of the present thesis proposes a novel model for the molecular regulation of PAs production in the seed: we show that MADS-domain ovule identity transcription factor *SEEDSTICK (STK)* directly regulates a network of PAs metabolic genes, a regulation that likely involves modification of chromatin landscape on STK targets.

4. Molecular control of seed size in Arabidopsis

Plant growth is determined by a succession of two moments: the proliferative phase, where the cell number largely increases but the cell size remains constant, and a second elongation phase where cells substantially increase their dimension to reach the final organ size. This has been described also in the case of seed development, with the strong indication that seed size regulation is subjected to an intrinsic control due to numerous molecular pathways (Egli, 2006).

The subject of organ size regulation is of outstanding interest in the case of seeds for the peculiarity that they represent; during their development in fact, coordinated growth and interplay of their three components occurs, a scenario which is further complicated by the fact that the seed coat, the endosperm and the embryo are characterized by different origins and fate, different ploidy level and different genotype (Garcia et al., 2005). Additionally, a deep understanding of mechanisms regulating seed size in model organism such as *Arabidopsis thaliana* is of great interest in the frame of grain-crops yield improvement, an inevitable scenario that we should face given the current worldwide demographic situation (see Paragraph 1.2).

For these reasons the study of seed development is of key importance to understand the whole mechanism governing seed size. While major regulators of seed size have been described, mostly in Arabidopsis (Garcia et al., 2005; Sun et al., 2010), the knowledge on the subject is still mostly fragmented and remains to be assembled into a global and coherent picture.

Part III of the present thesis proposes a novel pathway for the molecular regulation of seed size in *Arabidopsis thaliana* that involves the MADS-domain transcription factor *STK* and *AUXIN RESPONSE FACTOR 2 (ARF2)*: these data

confirm the pivotal role that the seed coat, thus the maternal tissues, have determining seed size.

Finally, for a more organic description of the subject, we recently reviewed the state of the art on seed size regulation knowledge in Arabidopsis, covering the description of the genomic imprinting and parent-of-origin effects, the analysis size regulatory pathways affecting developmental processes of the endosperm and of the integuments and describing the hormonal regulation of this important trait. Such information can be found in Part II of the this thesis (Networks controlling seed size in Arabidopsis).

Aim of the project

During my Ph.D. I have characterized the network controlling seed coat during seed development in *Arabidopsis thaliana*. In particular I have analyzed the role of the transcription factors STK and ARF2 in the determination of seed size and seed quality. Furthermore, I collaborated to unveiling *STK* function in proanthocyanidins synthesis in the seed.

Main results

It was previously proposed that the seed coat has a fundamental role in the determination of seed size, both for its role in physical delimitation of the inner seed cavity, both for evidences indicating a strict control of the maternal tissues on the development of the endosperm and of the embryo (Doughty et al., 2014; Sundaresan, 2005).
We characterized the phenotype of the *stk* mutant, that produces seeds smaller than the wild type (Pinyopich et al., 2003), proving that it mainly acts as a positive regulator of cell expansion in the seed coat and showing how it is also involved in the progression of the cell cycle, in a brassinosteroid-indipendent pathway. We also characterized the role of a known repressor of organ size, *ARF2* (Okushima et al., 2005), unveiling how it acts antagonistically of *STK* in determining seed size.
We also proved how *STK* is a master regulator of seed coat metabolism, unveiling its role in PAs biosynthesis and accumulation in the seed coat. Furthermore, we proposed that a common molecular mechanisms in the

regulation of the STK target might involve local remodeling of the chromatin state and, thus, of the accessibility of the DNA.

Conclusions and future prospects

By a detailed genetic and morphological analysis of loss-of-function mutants *stk*, *arf2* and *stk arf2* we were able to unveil a novel genetic network involving the role of *STK* and *ARF2* in the development of the seed coat and, thus, in the determination of seed size. We proved that *STK* acts in the seed coat, and its function involves a positive regulation of cell expansion, as the reduced cell length in the *stk* seeds suggests. The *stk* phenotype is also explained by the fact that *STK* negatively regulates *ARF2*, which itself acts as repressor of expansion and cell proliferation. We concluded that the molecular pathways downstream of *ARF2* are regulated by the brassinosteroids cascade, that on the contrary does not control *STK* activity on seed size. These two reciprocally influenced pathways are both fundamental in the determination and they represent master regulators of seed yield, confirming the importance of maternal control in seed size determination. In order to further investigate on these regulators, it would be of great interest the possibility of performing ChIP-seq analysis to identify the targets of STK and ARF2 during seed development. This approach would help identifying master regulator of seed size, and the knowledge acquired through the studies in Arabidopsis could eventually help the improvement in seed yield of agronomic relevant crop.

Our work on the role of *STK* in PAs production demonstrates that this key homeotic transcription factor not only determines the identity of ovules but also controls metabolic processes that occur later after initial identity determination

process, thus suggesting a link between identity determination and cell-specific (metabolic) processes.

Finally, we have proved for the first time that the regulation of the STK targets (such as *BANYULS* and *E2Fa*) in seeds could involve local remodeling of the chromatin landscape on their genomic regulatory regions: further studies on STK interactors could help understanding whether this molecular mechanism is conserved in the control of other metabolic and developmental pathways and if epigenetic regulation of target genes is a common feature among MADS-domain transcription factors.

References

Abrahams, S., Lee, E., Walker, A. R., Tanner, G. J., Larkin, P. J. and Ashton, A. R. (2003). The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant J.* **35**, 624–36.

Albertine, J. M. and Manning, W. J. (2009). Elevated night soil temperatures result in earlier incidence and increased extent of foliar ozone injury to common bean (Phaseolus vulgaris L.). *Environ. Pollut.* **157**, 711–3.

Alvarez, J. and Smyth, D. R. (1999). CRABS CLAW and SPATULA, two Arabidopsis genes that control carpel development in parallel with AGAMOUS. *Development* **126**, 2377–86.

Baxter, I. R., Young, J. C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., Peer, W. A., Hazen, S. P., Murphy, A. S. and Harper, J. F. (2005). A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2649–54.

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

Bencivenga, S., Colombo, L. and Masiero, S. (2011). Cross talk between the sporophyte and the megagametophyte during ovule development. *Sex. Plant Reprod.* **24**, 113–21.

Bencivenga, S., Simonini, S., Benková, E. and Colombo, L. (2012). The transcription factors BEL1 and SPL are required for cytokinin and auxin signaling during ovule development in Arabidopsis. *Plant Cell* **24**, 2886–97.

Berger, F. (1999). Endosperm development. *Curr. Opin. Plant Biol.* **2**, 28–32.

Berger, F. (2003). Endosperm: the crossroad of seed development. *Curr. Opin. Plant Biol.* **6**, 42–50.

Berger, F., Hamamura, Y., Ingouff, M. and Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci.* **13**, 437–43.

Ceccato, L., Masiero, S., Sinha Roy, D., Bencivenga, S., Roig-Villanova, I., Ditengou, F. A., Palme, K., Simon, R. and Colombo, L. (2013). Maternal Control of PIN1 Is Required for Female Gametophyte Development in Arabidopsis. *PLoS One* **8**, e66148.

Christensen, C. A., Subramanian, S. and Drews, G. N. (1998). Identification of gametophytic mutations affecting female gametophyte development in Arabidopsis. *Dev. Biol.* **202**, 136–51.

Colombo, L., Franken, J. and Koetje, E. (1995). The petunia MADS box gene FBP11 determines ovule identity. *Plant Cell ...* **7**, 1859–1868.

Crepet, W. L. (2000). Progress in understanding angiosperm history, success, and relationships: Darwin's abominably "perplexing phenomenon". *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12939–41.

Cucinotta, M., Colombo, L. and Roig-Villanova, I. (2014). Ovule development, a new model for lateral organ formation. *Front. Plant Sci.* **5**, 117.

Debeaujon, I., Léon-Kloosterziel, K. M. and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol.* **122**, 403–14.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M. and Lepiniec, L. (2003). Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* **15**, 2514–31.

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

Doughty, J., Aljabri, M. and Scott, R. J. (2014). Flavonoids and the regulation of seed size in Arabidopsis. *Biochem. Soc. Trans.* **42**, 364–9.

Dreni, L., Jacchia, S., Fornara, F., Fornari, M., Ouwerkerk, P. B. F., An, G., Colombo, L., Kater, M. M. and Celoria, V. (2007). The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. 690–699.

Dresselhaus, T. (2006). Cell–cell communication during double fertilization. *Curr. Opin. Plant Biol.* 41–47.

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

Duszynska, D., McKeown, P. C., Juenger, T. E., Pietraszewska-Bogiel, A., Geelen, D. and Spillane, C. (2013). Gamete fertility and ovule number variation in selfed reciprocal F1 hybrid triploid plants are heritable and display epigenetic parent-of-origin effects. *New Phytol.* **198**, 71–81.

Egli, D. B. (2006). The role of seed in the determination of yield of grain crops. *Aust. J. Agric. Res.* **57**, 1237.

Erdmann, R., Gramzow, L., Melzer, R., Theissen, G. and Becker, A. (2010). GORDITA (AGL63) is a young paralog of the Arabidopsis thaliana B(sister) MADS box gene ABS (TT16) that has undergone neofunctionalization. *Plant J.* **63**, 914–24.

Evans, L. T. (1996). *Crop Evolution, Adaptation and Yield.*

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X. and Zhang, Q. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–71.

FAO (2014). The State of Food Insecurity in the World (SOFI).

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M. and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **15**, 2603–11.

Focks, N. and Benning, C. (1998). wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* **118**, 91–101.

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

Gegas, V. C., Nazari, A., Griffiths, S., Simmonds, J., Fish, L., Orford, S., Sayers, L., Doonan, J. H. and Snape, J. W. (2010). A genetic framework for grain size and shape variation in wheat. *Plant Cell* **22**, 1046–56.

Haughn, G. and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci.* **10**, 472–7.

Higashiyama, T., Kuroiwa, H., Kawano, S. and Kuroiwa, T. (1998). Guidance in vitro of the pollen tube to the naked embryo sac of torenia fournieri. *Plant Cell* **10**, 2019–32.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima S, Kuroiwa, H. and Kuroiwa, T. (2001). Pollen tube attraction by the synergid cell. *Science* **293**, 1480–3.

Honma, T. and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525–9.

Huck, N., Moore, J. M., Federer, M. and Grossniklaus, U. (2003). The Arabidopsis mutant feronia disrupts the female gametophytic control of pollen tube reception. *Development* **130**, 2149–59.

Jacobson, D., LeFebvre, S., Ojerio, R., Berwald, N. and Heikkinen, E. (1998). Persistent, systemic, asymptomatic infections of Albugo candida, an oomycete parasite, detected in three wild crucifer species. *Canada J. Bot.* 739–750.

Johnson, M. and Lord, E. (2006). Extracellular Guidance Cues and Intracellular Signaling Pathways that Direct Pollen Tube Growth. In *The pollen tube: a molecular and cellular perspective* (ed. Malhó, R.), pp. 223–242. Berlin/Heidelberg: Springer-Verlag.

Kasahara, R. D., Portereiko, M. F., Sandaklie-Nikolova, L., Rabiger, D. S. and Drews, G. N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. *Plant Cell* **17**, 2981–92.

Kitamura, S., Shikazono, N. and Tanaka, A. (2004). TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant J.* **37**, 104–14.

Krannitz, P., Aaeresen, L. and Dow, J. (1991). The effect of genetically based differences in seed size on seedling survival in Arabidopsis thaliana (Brassicaceae). *Am. J. Bot.* 446–450.

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N. and Caboche, M. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–30.

Lieber, D., Lora, J., Schrempp, S., Lenhard, M. and Laux, T. (2011). Arabidopsis WIH1 and WIH2 genes act in the transition from somatic to reproductive cell fate. *Curr. Biol.* **21**, 1009–17.

Lopes, M. A. and Larkins, B. A. (1993). Endosperm origin, development, and function. *Plant Cell* **5**, 1383–99.

Mansfield, S. G. and Briarty, L. G. (1991). Early embryogenesis in Arabidopsis thaliana . II. The developing embryo. *Can. J. Bot.* **69**, 461–476.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.-M., Debeaujon, I. and Klein, M. (2007). The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* **19**, 2023–38.

Matias-Hernandez, L., Battaglia, R., Galbiati, F., Rubes, M., Eichenberger, C., Grossniklaus, U., Kater, M. M. and Colombo, L. (2010). VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *Plant Cell* **22**, 1702–15.

Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., Kater, M. M. and Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. *Plant Cell* **25**, 2560–72.

Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R. and Colombo, L. (2011). The MADS box genes SEEDSTICK and ARABIDOPSIS B(sister) play a maternal role in fertilization and seed development. *Plant J.* 1–12.

Mohamed-Yasseen, Y., Barringer, S. A., Splittstoesser, W. E. and Costanza, S. (1994). The role of seed coats in seed viability. *Bot. Rev.* **60**, 426–439.

Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B. and Westoby, M. (2005). A brief history of seed size. *Science* **307**, 576–80.

Nowack, M. K., Ungru, A., Bjerkan, K. N., Grini, P. E. and Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* **38**, 604–12.

Okushima, Y., Mitina, I., Quach, H. L. and Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. *Plant J.* **43**, 29–46.

Orsi, C. H. and Tanksley, S. D. (2009). Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. *PLoS Genet.* **5**, e1000347.

Pelaz, S., Gustafson-Brown, C., Kohalmi, S. E., Crosby, W. L. and Yanofsky, M. F. (2001). APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J.* **26**, 385–94.

Pelletier, M. K., Burbulis, I. E. and Winkel-Shirley, B. (1999). Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and end-products in Arabidopsis seedlings. *Plant Mol. Biol.* **40**, 45–54.

Peris, C. I. L., Rademacher, E. H. and Weijers, D. (2010). Green beginnings - pattern formation in the early plant embryo. *Curr. Top. Dev. Biol.* **91**, 1–27.

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E. and Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85–8.

Pourcel, L., Routaboul, J.-M., Kerhoas, L., Caboche, M., Lepiniec, L. and Debeaujon, I. (2005). TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell* **17**, 2966–80.

Prasad, K. and Ambrose, B. A. (2010). Shaping up the fruit. 899–902.

Prasad, K., Zhang, X., Tobón, E. and Ambrose, B. a (2010). The Arabidopsis B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. *Plant J.* **62**, 203–14.

Preuss, D. (1995). Being fruitful: genetics of reproduction in Arabidopsis. *Trends Genet.* **11**, 147–53.

Reise, L. and Fisher, R. (1993). The ovule and embryo sac. *Plant Cell* **5**, 1291–1301.

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

Russell, S. (1992). Double fertlization. *Int. Rev. Cytol.* **140**, 357–388.

Russell, S. D. (1996). Attraction and transport of male gametes for fertilization. *Sex. Plant Reprod.* **9**, 337–342.

Schneitz, K., Hulskamp, M. and Pruitt, R. (1995). Wild-type ovule development in Arabidopsis thaliana: a light microscope analysis of cleared whole-mount tissue. *Plant J.* **7**, 331–335.

Schwartz, B., Yeung, E. and DW, M. (1994). Disruption of morphogenesis and transformation of the suspensor in abnormal suspensor mutants of Arabidopsis. *Development* 3235–3245.

Shah, M. and Strong, M. F. (2002). Food in the 21st Century: from Science to Sustainable Agriculture. *World Bannk* http://www.worldbank.org/html/cgiar/publications/s.

Silvertown, J. (1989). The paradox of seed size adaptation. *Trends Ecol. Evol.* 24–26.

Smyth, D. R., Bowman, J. L. and Meyerowitz, E. M. (1990). Early flower development in Arabidopsis. *Plant Cell* **2**, 755–67.

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. and Lin, H.-X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**, 623–30.

Spielman, M., Vinkenoog, R., Dickinson, H. G. and Scott, R. J. (2001). The epigenetic basis of gender in flowering plants and mammals. *Trends Genet.* **17**, 705–11.

Sun, X., Shantharaj, D., Kang, X. and Ni, M. (2010). Transcriptional and hormonal signaling control of Arabidopsis seed development. *Curr. Opin. Plant Biol.* **13**, 611–20.

Sundaresan, V. (2005). Control of seed size in plants. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17887–8.

Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* **4**, 75–85.

Van Went, J. and Willemse, M. (1984). Ferilization in Angiosperms. In *Embryology of Angiosperm*, pp. 273–318.

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., Lei, C., Zhang, X., Cheng, Z., Guo, X., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18**, 1199–209.

West, M. and Harada, J. J. (1993). Embryogenesis in Higher Plants: An Overview. *Plant Cell* **5**, 1361–1369.

Western, T. L., Skinner, D. J. and Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant Physiol.* **122**, 345–56.

Westoby, M., Leishman, M., Lord, J., Poorter, H. and Schoen, D. J. (1996). Comparative Ecology of Seed Size and Dispersal [and Discussion]. *Philos. Trans. R. Soc. B Biol. Sci.* **351**, 1309–1318.

Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. and Wright, I. J. (2002). PLANT ECOLOGICAL STRATEGIES: Some Leading Dimensions of Variation Between Species. *Annu. Rev. Ecol. Syst.* **33**, 125–159.

WFP (2009). World hunger. http://www.wfp.org/hunger.

Windsor, J. B., Symonds, V. V, Mendenhall, J. and Lloyd, A. M. (2000). Arabidopsis seed coat development: morphological differentiation of the outer integument. *Plant J.* **22**, 483–93.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* **5**, 218–23.

Xu, W., Grain, D., Bobet, S., Le Gourrierec, J., Thévenin, J., Kelemen, Z., Lepiniec, L. and Dubos, C. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. *New Phytol.* **202**, 132–44.

Zhao, L., Gao, L., Wang, H., Chen, X., Wang, Y., Yang, H., Wei, C., Wan, X. and Xia, T. (2013). The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Funct. Integr. Genomics* **13**, 75–98.

Part II

Contents

"SEEDSTICK Is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat."

PLOS Genetics, *article in press*

"Networks controlling seed size in Arabidopsis"
Plant Reproduction, invited review for the special issue *From gametes to seeds*.

SEEDSTICK Is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat

Short title: STK Controls Metabolite Accumulation in Seeds

Authors: Chiara Mizzotti[1,#], Ignacio Ezquer[1,#], Dario Paolo[1], Paloma Rueda-Romero[2,a], Rosalinda Fiorella Guerra[1,b], Raffaella Battaglia[1,c], Ilana Rogachev[3], Asaph Aharoni[3] , Martin M. Kater[1], Elisabetta Caporali[1] and Lucia Colombo[1,4,*].

[#] These authors contributed equally to this work.

Author affiliation:

[1] Dipartimento di BioScienze, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy

[2] Centro de Biotecnología y Genómica de Plantas-UPM-INIA, ETSI Agrónomos, Universidad Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

[3] Department of Plant Sciences, Weizmann Institute of Science, P.O. Box 26 Rehovot 76100, Israel

[4] Consiglio Nazionale delle Ricerche, Istituto di Biofisica, 20133 Milan, Italy

[a] Present address: Department of Plant and Microbial Biology, University of California, 111 Koshland Hall, Berkeley CA 94720-3102

[b] Present address: Department of Craniofacial Development & Stem Cell Biology, King's College London, London SE1 9RT

Dario Paolo

[c] Present address: Istituto di Agronomia, Genetica e Coltivazioni Erbacee, Università Cattolica del Sacro Cuore, Via Emilia Parmense 84, 29122 Piacenza, Italy

[*]Corresponding author e-mail: lucia.colombo@unimi.it

ABSTRACT

The role of secondary metabolites in the determination of cell identity has been an area of particular interest over recent years, and studies strongly indicate a connection between cell fate and the regulation of enzymes involved in secondary metabolism. In *Arabidopsis thaliana* the maternally derived seed coat plays pivotal roles in both the protection of the developing embryo and the first steps of germination. In this regard a characteristic feature of seed coat development is the accumulation of proanthocyanidins (PAs - a class of phenylpropanoid metabolites) in the innermost layer of the seed coat. Our genome-wide transcriptomic analysis suggests that the ovule identity factor SEEDSTICK (STK) is involved in the regulation of several metabolic processes, providing a strong basis for a connection between cell fate determination, development and metabolism. Using phenotypic, genetic, biochemical and transcriptomic approaches, we have focussed specifically on the role of *STK* in PA biosynthesis. Our results indicate that STK exerts its effect by direct regulation of the gene encoding *BANYULS*/*ANTHOCYANIDIN REDUCTASE* (*BAN*/*ANR*), which converts anthocyanidins into their corresponding 2,3-cis-flavan-3-ols. Our study also demonstrates that the levels of H3K9ac chromatin modification directly correlate with the active state of *BAN* in an STK-dependent way. This is consistent with the idea that MADS-domain proteins control the expression of their target genes through the modification of chromatin states. STK might thus recruit or regulate histone modifying factors to control their activity. In addition, we show that STK is able to regulate other *BAN* regulators. Our study demonstrates for the first time how a floral homeotic gene controls organ identity through the regulation of a wide range of processes including the accumulation of secondary metabolites.

AUTHOR SUMMARY

Plant secondary metabolites accumulate in seeds to protect the developing embryo. Using an RNA sequencing approach in conjunction with enrichment analyses we identified the homeotic MADS-domain gene *SEEDSTICK* (*STK*) as a regulator of metabolic processes during seed development. We analyzed the role of *STK* as a key regulator of the production of proanthocyanidins, compounds which are important for the pigmentation of the seed. STK directly regulates a network of metabolic genes and is also implicated in changes occurring in the chromatin landscape. Our work demonstrates that a key homeotic transcription factor not only determines the identity of ovules but also controls metabolic processes that occur subsequent to the initial identity determination process, thus suggesting a link between identity determination and cell-specific (metabolic) processes.

INTRODUCTION

Seeds are essential units for plant propagation. Their development is an intricate process that requires the coordinated development of the embryo, the endosperm and the seed coat. The seed coat is derived from the maternal integuments and surrounds the embryo providing the latter protection against both mechanical damage as well as that inflicted by UV radiation. In addition, it facilitates the efficient dispersion of offspring and mediates initial water uptake during germination [1]. Upon double fertilization, the first phase of seed development is characterized by several morphological changes followed by the accumulation of secondary metabolites in specialized seed coat cells which mainly act in defence responses [2]. Although some interconnections between secondary metabolism and cell differentiation have been reported, the molecular mechanisms involved have not yet been elucidated. New approaches to genome-wide target identification indicate the existence of a correlation between cell identity specification and the regulation of secondary metabolism. For example, direct targets of the MADS-domain transcription factor SEPALLATA3 (SEP3) have been shown to include genes involved in lipid biosynthesis, hormone production and the biosynthesis of sterol and wax [3]; SHORT VEGETATIVE PHASE (SVP), another MADS-domain protein, binds to genes involved in the hormone stimulus response, suggesting its involvement in the cytokinin, auxin and jasmonate signalling pathways [4]. MADS-domain transcription factors have been demonstrated to be important regulators of floral organ specification, and SEEDSTICK (STK) in particular has been shown to play a pivotal role in ovule ontogeny [5]. The *STK* gene controls ovule identity redundantly with *SHATTERPROOF1* (*SHP1*) and *SHP2* [6–8]. Furthermore, *STK* is required for normal seed shedding and, together with another MADS-

domain gene *ARABIDOPSIS B SISTER* (*ABS*), for proper formation of the endothelium, the innermost layer of the seed coat [7,9]. Transcriptome analysis of developing ovules and seeds has suggested the involvement of STK in the control of secondary metabolism (see later). Based on the data obtained and the morphological characterization of the *stk* mutant we have focused on the regulation of the flavonoid metabolic pathway.

Fertilized ovules accumulate proanthocyanidins (PAs) in the endothelium. These molecules are flavan-3-ols and in Arabidopsis they are composed of epicatechin monomers and polymers [10]. PAs are important compounds as they provide protection against light and predation by herbivores, they have antimicrobial and antioxidant activities, and in addition they limit the growth of neighbouring plants [11–16]. Epicatechins are responsible for the brown pigmentation of the Arabidopsis seed [16]. They are synthesized in the cytoplasm and then transported to the vacuoles where finally they are polymerized [17]. In Arabidopsis, PA biosynthesis begins in the micropylar region of the endothelium around 1 to 2 days after fertilization and then progressively extends to include the rest of the endothelium up to 5 to 6 days after fertilization [2,16]. The genetics of PA biosynthesis and accumulation has been well-studied in Arabidopsis revealing the participation of a complex network of several groups of genes. PA biosynthetic structural genes are divided into two groups: the Early (EBGs) and the Late (LBGs) Biosynthetic Genes [16,18,19]. The EBGs comprise four genes: *CHALCONE SYNTHASE* (*CHS*), *CHALCONE ISOMERASE* (*CHI*), *FLAVONOL 3-HYDROXYLASE* (*F3H*) and *FLAVONOL 3'-HYDROXYLASE* (*F3'H*), and are involved in the biosynthesis of precursors for PAs and other classes of Arabidopsis flavonoids. The LBGs include *DIHYDROFLAVONOL-4-REDUCTASE* (*DFR*), *LEUCOANTHOCYANIDIN DIOXYGENASE* (*LDOX*) and

BANYULS/ANTHOCYANIDIN REDUCTASE (*BAN/ANR*). *BAN* is considered to be a branch point in the phenylpropanoid biosynthetic pathway. It encodes an anthocyanidin reductase which converts anthocyanidins to their corresponding 2,3-cis-flavan-3-ols [20]. A third group of structural genes has also been identified in Arabidopsis that comprises *TRANSPARENT TESTA 12* (*TT12*, *MATE transporter*), *TT10* (*laccase 15*), *TT19* (*glutathione-S-transferase*) and *AHA10* (*H^+-ATPase*). These genes are involved in flavan-3-ol modification, transport and oxidation [21–24] and have recently been proposed to be LBG members [19]. The regulation of LBGs occurs via a ternary protein complex called MBW (MYB-bHLH-WDR), formed by a specific R2R3-MYB, a bHLH transcription factor and the WD repeat protein TRANSPARENT TESTA GLABRA 1 (TTG1; [16,25,26]). In addition to this complex, other transcription factors belonging to different families such as Zinc finger (*TT1*/*WIP1*), MADS (*ABS*/*TT16*/*AGL32*) and WRKY (*TTG2*/*DSL1*/*WRKY44*) also participate in their regulation [16,27–29].

Studies in several fields indicate a correlation between cell identity determination and metabolism. In unicellular organisms, like some species of yeast and *Streptomyces*, cellular differentiation is influenced by nutrients and metabolism [30,31]. Some years ago it was proposed that enzymes involved in carbon metabolism also regulate myoblast differentiation [32,33], while stem cell differentiation has recently been proposed to be regulated by the metabolisms of both lipids and methionine [34,35].

The results presented in this manuscript demonstrate that STK directly prevents ectopic accumulation of PAs in the seed coat. STK both directly and indirectly regulates *BAN* and this involves STK-dependent changes in histone modification. The discovery of STK as a master regulator of genes involved in the anthocyanin biosynthetic pathway provides an interesting link between the

determination of organ identity determination and more downstream cell-specific metabolic processes. Furthermore it opens up new possibilities to increase the levels of bioactive natural products which constitute a rich source of novel therapeutic compounds, or to modify pigments in plant tissues.

RESULTS

STK is involved in the regulation of metabolic pathways

As a first step to understand which processes are controlled by STK, a high-throughput RNA-Seq analysis was performed comparing wild-type plants with the *stk* mutant. Based on the *STK* expression pattern, RNA was extracted from flowers starting from early stages of development (stage 9) until maturity and after fertilization until 5 Days After Pollination (DAP). Analysis of the raw data was performed on the commercially available CLC Genomics Workbench v.4.7.1 (http://www.clcbio.com/genomics/). A total of 102,278,242 reads passed a quality filter and 85% were mapped back to the Arabidopsis TAIR10 genome. Approximately 90% of these mapped uniquely to single locations and each could thus be assigned to a single annotated TAIR10 gene. Normalization of expression was performed using RPKM values [36]. All other parameters were kept at default levels. The CLC Genomic Workbench was further used to identify and assess the levels of all the differentially expressed transcripts found in each cDNA library. Baggerley's test and False Discovery Rate (FDR) correction were used for the statistical evaluation of samples [37]. Our analysis revealed that 156 genes were up-regulated (Table S1) in the *stk* mutant compared to wild type, whereas 90 were found to be down-regulated (Table S2). To obtain an initial insight into the potential functions of *STK* downstream genes, a global view of function and the underlying biology of the differentially expressed genes was obtained by examining their gene ontology using agriGO (Figure 1; [38]). For up-regulated genes in the *stk* mutant, the biological process category showed enrichment for lipid localization and secondary metabolic processes (Figure 1A; Table S3). There was also notable enrichment for terms related to the phenylpropanoid metabolic process as well as flavonoid

biosynthesis. Analysis of the molecular functions affected revealed enrichment for genes encoding pectinesterase, enzyme inhibitor and lipid binding activities (Figure 1B; Table S3). Analysis of enriched cellular component terms in the list of up-regulated genes included categories related to the endomembrane system and also cell and cell part groups (Figure 1C; Table S3). Among the group of genes down-regulated in *stk*, we only found significant enrichment of terms in the molecular function category related to DNA binding (Figure 1D; Table S3). The transcriptome picture emerging provided a global view of the downstream networks regulated by STK. Interestingly, we found genes involved in flavonoid biosynthesis to be significantly represented in the up-regulated genes. That this group included key enzymatic players involved in PA synthesis constituted the basis for the work we report here on deciphering the role of STK in this seed coat process.

The *stk* mutant accumulates PAs ectopically

The transcriptomic analysis of *stk* mutant ovules and seeds highlighted a significant increase in the abundance of transcripts involved in secondary metabolic processes, including those of genes involved in flavonoid and phenylpropanoid biosynthesis. Interestingly, in this group we found *DFR*, *LDOX* and *BAN*, three key enzyme encoding genes controlling the synthesis of catechin and epicatechin, the precursors of PAs. We therefore decided to study the role of STK in the regulation of the PA biosynthetic pathway in more detail since this is considered to be a key metabolic pathway linked to seed development (for review see [16]). Furthermore, recent discoveries have shown that flavonoids play a fundamental role in regulating communication between the seed coat and the endosperm [39]. PAs in wild-type Arabidopsis seeds are

accumulated in the endothelium. To investigate and compare the accumulation of PAs in wild-type and *stk* mutant seeds we made seed sections at the heart stage of embryo development (when PA accumulation in the endothelium is completed) and stained these with toluidine blue O. This staining provided a general view of all seed coat cells and revealed the presence of phenolic compounds in the endothelium of wild-type and *stk* mutant seeds as highlighted by the blue staining of their vacuoles (Figure 2A - C). However, the toluidine blue also evidenced considerable accumulation of phenolic compounds in the outermost layer of the inner integument (ii2) in the *stk* mutant seed coat (Figure 2C, asterisk). In order to confirm the nature of the latter we used the vanillin assay that specifically detects flavan-3-ols and their proanthocyanidin polymers [2]. This analysis confirmed that PAs are accumulated in the endothelium in wild-type seeds (Figure 2D) whereas in the *stk* mutant they are additionally observed in the ii2 layer (Figure 2E, asterisk).

To better understand the role of STK in PA synthesis, soluble and insoluble extracts from mature and immature (6 DAP) seeds were analyzed by Liquid Chromatography-Mass Spectrometry (LC-MS) and the complete metabolic profiles obtained are shown in Figure 3. Peaks that could be identified as known compounds were selected and analyzed (Table S4). No differences were observed in the metabolic profiles of insoluble PAs between mature wild-type and *stk* seeds (Figure S1). At 6 DAP, insoluble PAs could not be detected which concords with the solvent-soluble nature of PAs at the immature stage [10]. Soluble PAs, however, showed some differences: at 6 DAP, both wild-type and *stk* mutant seeds contained the same levels of PA oligomers (n = 2–9) but the level of epicatechin monomers was higher in the *stk* mutant compared to the wild type (Figure 3A). In mature seeds only six soluble PA metabolites were detected. The levels of dimers, trimers and tetramers were the same in the wild

type and in the mutant, but the levels of epicatechin monomers, pentamers and hexamers were different (Figure 3B). In particular, the levels of pentamers and hexamers in the wild type were higher compared to the mutant. By contrast the level of epicatechin monomers was higher in the *stk* mutant, as already detected in the metabolic profiles of immature seeds. The total amount of PAs in the *stk* mutant was greater than in wild-type seeds. These data support the morphological analysis and suggest that *stk* mutant seeds have a higher level of PAs than wild-type seeds. Furthermore, it implies that STK is predominantly involved in epicatechin monomer metabolism and only slightly affects oligomer production.

STK negatively controls key regulators of the PA biosynthetic pathway

The biosynthesis of PAs is dependent on structural genes that can be divided into EBGs and LBGs (Figure 4), the EBGs being expressed prior to the LBGs [16]. Based on the transcriptome data, we focused our attention on the set of genes controlling the transformation of anthocyanidin into epicatechin (Figure 4). The RNA-Seq data revealed that the expression of genes belonging to the EBGs is unaltered in the *stk* mutant background (Table 1); however, the levels of all the LBGs were found to be increased in the mutant (Table 1). This suggests that STK acts as a repressor of the expression of all the LBGs. Among these, *BAN* codes for the core enzyme of PA production [25,40–42]. We performed quantitative Real Time-PCR (qRT-PCR) experiments on siliques from 0 to 6 DAP and confirmed the results obtained by the RNA-Seq experiment showing *BAN* to be up-regulated in the *stk* mutant (Figure S2).

To investigate in which seed tissues *BAN* is expressed we performed *in situ* hybridization experiments using an antisense *BAN* probe. In wild-type seeds *BAN* was expressed only in the endothelium layer (Figure 5A; [20]). In contrast,

BAN expression was observed in the endothelium layer and ectopically in the ii2 layer in the *stk* mutant (Figure 5B - C). These data confirm that STK controls the spatial and temporal expression of *BAN*.

The SEEDSTICK-GFP reporter line reveals a dynamic fluorescence pattern in developing ovules and seed integuments.

To investigate how STK regulates *BAN* expression we examined the expression pattern of the STK protein. We cloned the entire *STK* genomic region (a DNA fragment comprising 3.5 kb of sequence upstream of the ATG codon plus all the coding region) as a translational fusion to a GFP reporter gene, and the final *pSTK::STK-GFP* construct was introduced into the *stk* mutant background. The resulting plants produced seeds that were indistinguishable from wild-type and were able to abscise from the fruit upon maturity, demonstrating that the STK-GFP fusion protein was biologically active and able to fully complement the absence of the endogenous STK protein. STK-GFP was uniformly detected in the nuclei of the placenta and the early ovule primordia (Figure 6A - B) which is consistent with previous data showing *STK* mRNA expression from stage 9 of flower development [7]. Later, as the developing ovules initiate the inner and outer integuments, GFP expression was found to be restricted to the nucellus and the funiculus (Figure 6B). However, during subsequent stages the fluorescence signal appeared throughout the outer and inner integuments (Figure 6C - D), and as development proceeded it could be seen to be present strongly in the growing funiculus and also in its contact region with the placenta. This pattern of protein localization during ovule development is consistent with previous *in situ* hybridization data [7]. During early embryogenesis (Figure 6E) the STK-GFP protein remained extended throughout the outer integuments. Cell wall specific staining with propidium

iodide (PI) was also carried out and allowed us to better define the sites of STK-GFP localization as the outer integuments and the second layer of the inner integument (Figure 6F).
Our observations thus show that the distribution of the STK protein changes over the course of ovule and seed development. During early stages it is detected uniformly distributed in the placenta and the different cell types of the ovule primordia, whereas later it becomes restricted to the maternal seed coat and funiculus. No STK-GFP was detected in the ii1 and ii1' layers suggesting that here STK does not repress *BAN* in these tissues. These data are consistent with the observation that STK is able to repress *BAN* expression only in the ii2 layer of the inner integument.

STK binds to the *BAN* promoter affecting histone H3K9 acetylation levels
RNA-Seq, expression analysis and *in situ* data all indicate that *BAN* expression is regulated by STK. In order to investigate whether this regulation involved direct interaction between the STK protein and the *BAN* gene we performed a ChIP (Chromatin Immunoprecipitation) assay on wild-type inflorescences and siliques (up to 6 DAP) using an antibody specific against the STK protein. Chromatin extracted from wild-type leaves was used as a negative control since *STK* is not expressed in this tissue, and the binding of STK to the *VERDANDI* (*VDD*) promoter [43] was used as the positive control. Bearing in mind that MADS-domain proteins recognize and bind CArG boxes [44], the *BAN* gene genomic sequences comprising the 3 Kb upstream of the ATG start codon, the structural gene and 1 kb downstream of the STOP codon were analysed for the presence of such consensus motifs (allowing up to one base mismatch; Figure 5D). Six CArG boxes were found in the region of the *BAN* promoter sequence [2] extending up to 355 bp upstream from the ATG. We detected significant

enrichment for the region immediately upstream of the *BAN* translational start site (primer set spanning positions -140 to +1, including three CArG boxes (Figure 5E)) thus indicating the presence of STK bound to this region.

To investigate a possible mechanism of regulation of the expression of *BAN* by STK, we examined epigenetic marks at the *BAN* locus. Modifications such as the hyperacetylation of histones H3 and H4 have diverse impacts on gene transcriptional activity and chromatin organization [45]. H3K9ac is one of the most well-characterized epigenetic marks associated with active transcription and has been shown to influence numerous developmental and biological processes in higher plants [46–48]. To address the question of whether the differential expression of *BAN* observed between *stk* and wild-type tissues correlates with alterations in this epigenetic marker we analyzed wild-type and *stk* mutant siliques at 3-4 DAP for H3K9 acetylation at the *BAN* locus. ChIP experiments were performed using an antibody specific to H3K9ac and were analyzed by qRT-PCR (Figure 5F). *IAA8* was used as a reference as it carries the H3K9ac mark and is equally expressed in both wild-type and *stk* mutant plants [49]. Interestingly, in wild-type material we found considerable enrichment of DNA sequences corresponding to the region around the translational start site within which STK binding sites are located. When we assayed the same region in the *stk* mutant we observed a dramatic increase in enrichment compared to wild type. These results are consistent with the presence of elevated levels of H3K9ac (compared to H3) in the region of the wild-type *BAN* promoter where STK interacts (transcriptionally active chromatin), and in addition evidences a very considerable enrichment of H3K9ac in this region in the *stk* mutant which correlates with the increased transcriptional activity of the *BAN* gene observed in the mutant background.

STK controls *BAN* transcriptional regulators

Our data suggest that STK might regulate *BAN* expression directly binding to its promoter. Previous studies have identified other key regulators of BAN expression, including ABS, TT8 and ENHANCER OF GLABRA3 (EGL3). Whereas TT8 and EGL3 act redundantly in a protein complex that promotes the expression of *BAN*, ABS is necessary for PA biosynthesis and normal endothelium cell morphology [9,28,41,42,50]. This raised the question as to whether STK might act as master regulator also controlling the genes encoding transcription factors that regulate *BAN*. We therefore investigated the expression of *ABS*, *EGL3* and *TT8* in both wild-type and *stk* mutant siliques (3-4 DAP) by qRT-PCR, and also analyzed the expression of *ABS* in developing flowers in order to study the relationship between *STK* and *ABS* in their roles in endothelium formation [9]. This experiment revealed up-regulation of *TT8*, *EGL3* and *ABS* in *stk* mutant siliques and hence that STK represses these genes in the wild type at this specific stage of development (Figure 7A); that the level of *ABS* in un-pollinated inflorescences was unaffected indicates that STK does not play a role in regulating the expression of *ABS* prior to pollination. These data therefore suggest that STK regulates these genes during a narrow stage of development, at 3-4 DAP.

In order to clarify whether *STK* directly controls the expression of these regulatory genes, we performed a ChIP assay. Bioinformatics analysis of the gene loci revealed the presence of two CArG boxes in the putative *EGL3* promoter region at -2213 and -2177 bp; three CArG boxes in the putative *ABS* promoter region between -1692 and -1599 bp, and the *TT8* gene presented two CArG boxes in the last exon at +3805 and +3815 bp from the ATG (Figure 7B). The ChIP assays showed that STK indeed binds to the CArG-containing regions of the *EGL3* and *ABS* genes but not to the selected region of *TT8*

(Figure 7C). This analysis suggests that the control of PA biosynthesis by STK might occur via two mechanisms: a direct interaction with the *BAN* locus, and in addition by direct and indirect regulation of the expression of genes that encode transcriptional regulators of *BAN*.

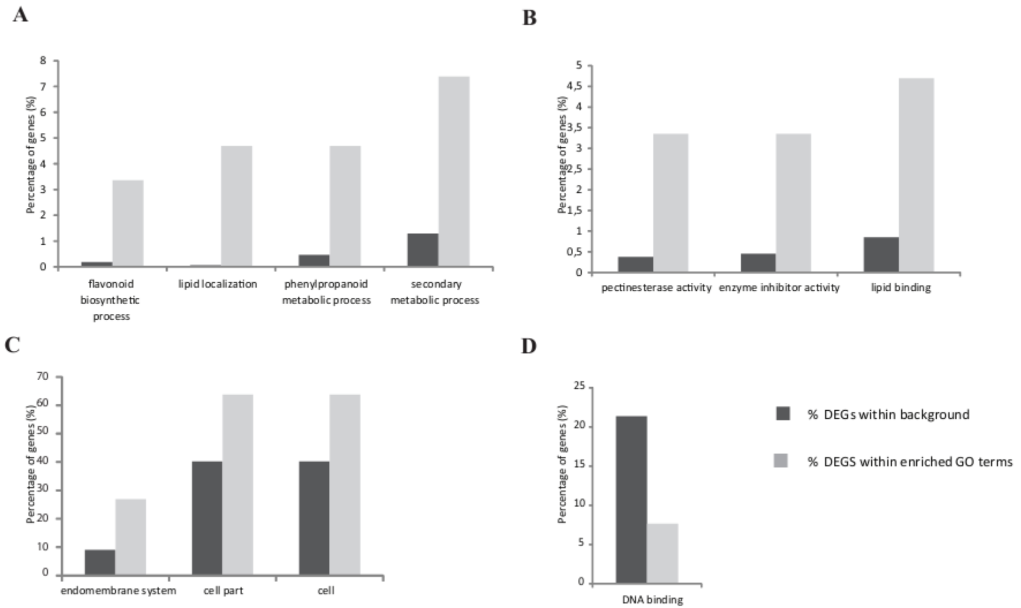


Figure 1: **Histogram of functional Gene Ontology analysis of differentially expressed genes. Slim Plant term enrichment - up and down-regulated genes.**

Genes with higher expression in *stk* in the Biological Process (*A*), Molecular Function (*B*) and Cellular Component (*C*) categories. Genes with lower expression in *stk* Molecular Function category (*D*).

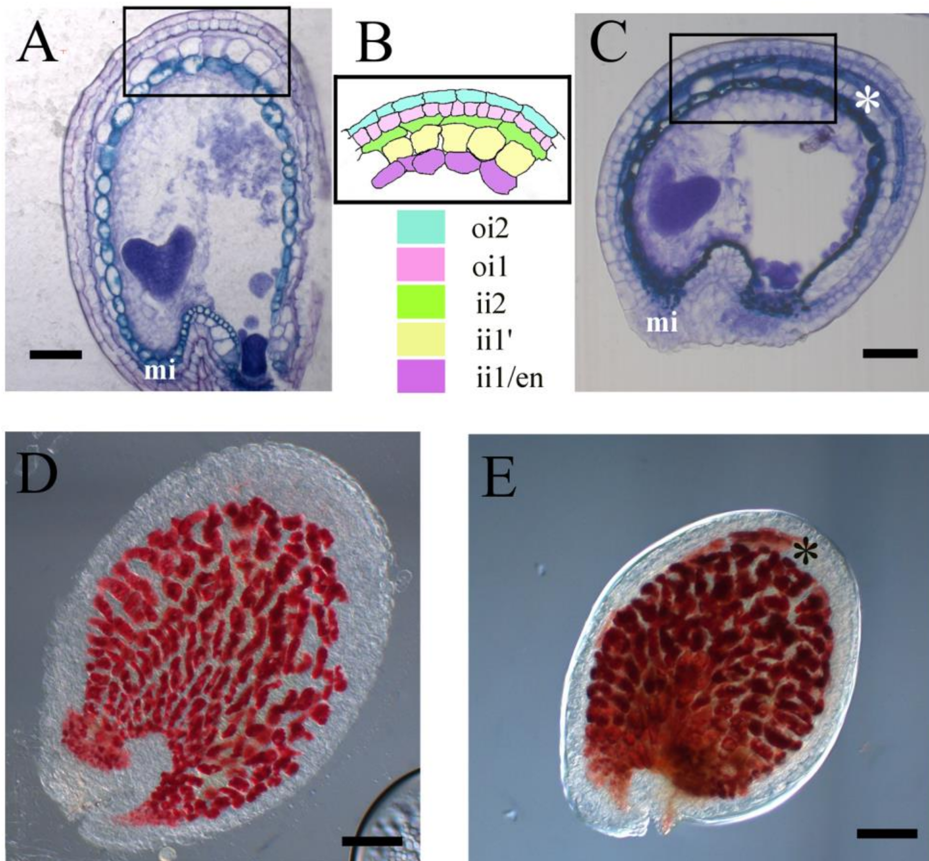


Figure 2: *stk* mutant seeds present defects in seed coat PA accumulation.
(*A*) Sections of wild-type seeds stained with the toluidine blue O revealed the presence of phenolic compounds in the endothelium (ii1). (*B*) Scheme of Arabidopsis seed coat anatomy. (*C*) In the *stk* mutant phenolic compounds are accumulated in the endothelium (ii1) and also in the second layer of the inner integument (ii2, asterisk). (*D*) Whole-mount vanillin staining confirmed the presence of PAs in the wild-type and (*E*) in the *stk* mutant endothelium. In the *stk* mutant PAs are also accumulated outside the endothelium in the second layer of the inner integument (asterisk).
mi, micropyle; en, endothelium. Scale bars = 30 µm (*A*-*E*).

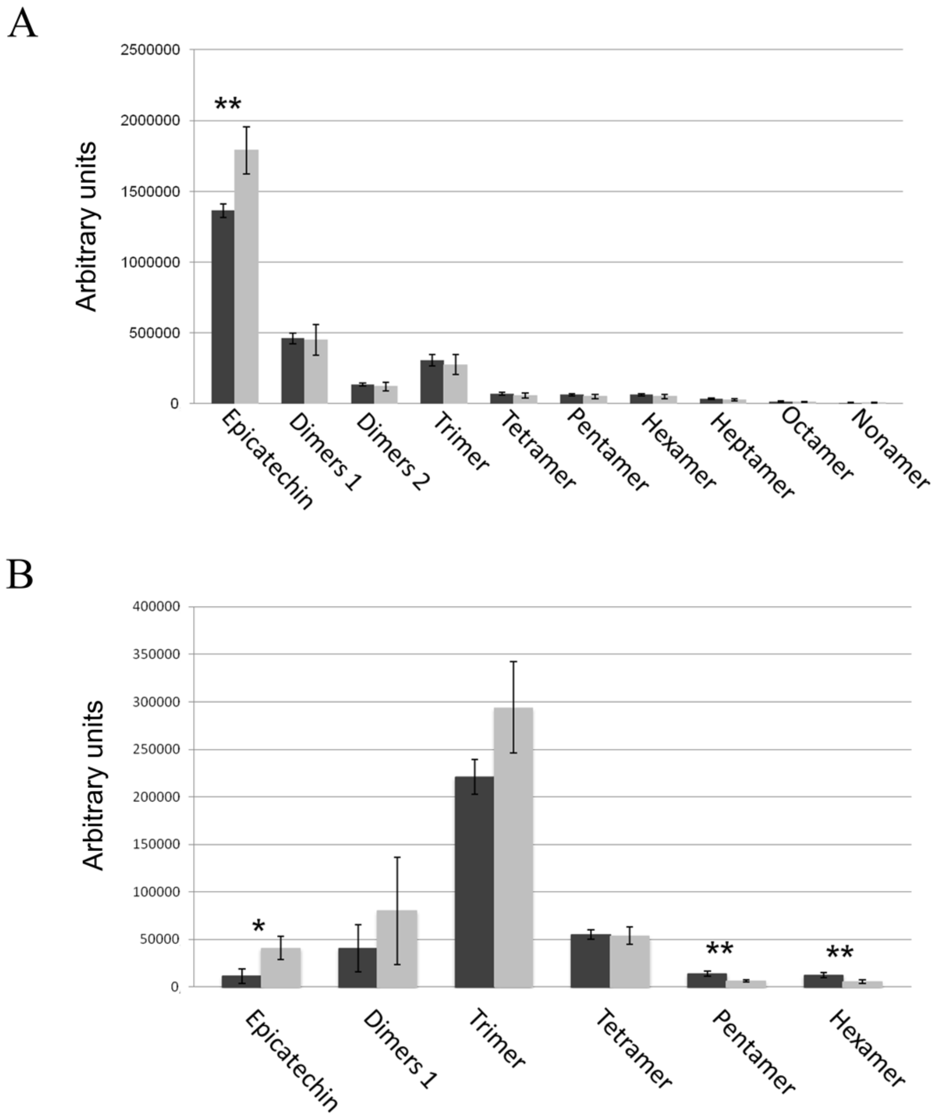


Figure 3: Solvent soluble PAs analysed by LC-MS.

Soluble PAs were detected in wild type (black bars) and the *stk* mutant (grey bars) at the immature (6 DAP; *A*) and mature (*B*) stages of seed development. Error bars represent SD of three independent measurements. Asterisks indicate statistically significant differences as determined by Student's *t* test (* $P < 0.05$, ** $P < 0.01$).

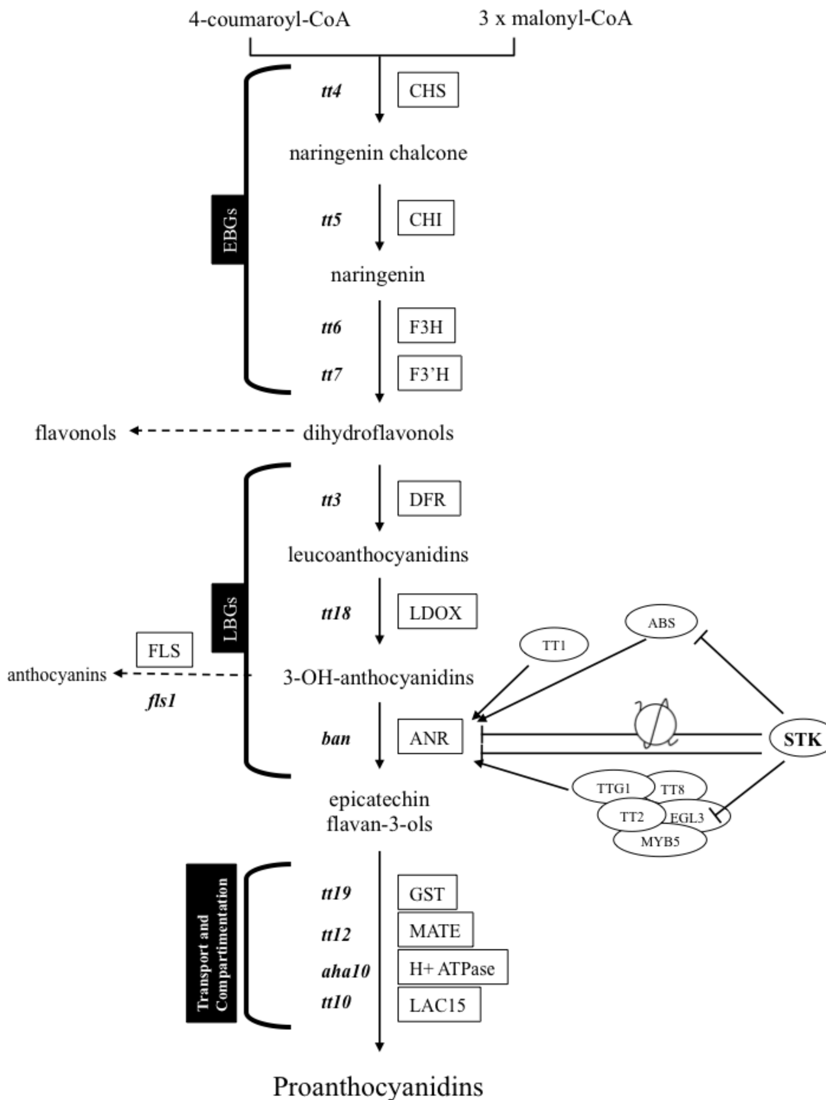
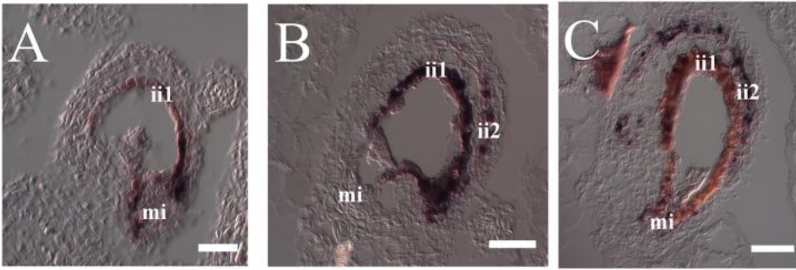


Figure 4: *STK* has a pivotal role in the control of PA production.

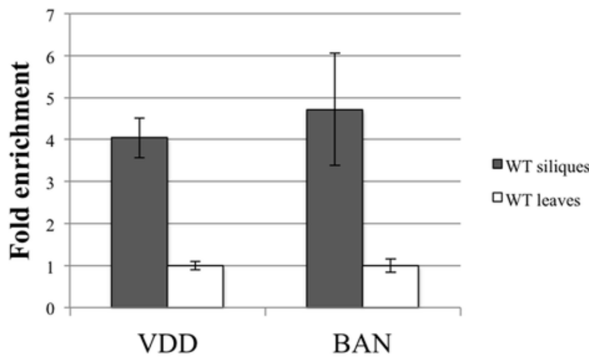
Schematic representation of the pathways for PA production. The genes involved in these pathways are shown in boxes. *BAN*, that was found to be up-regulated in RNA-Seq, has been analysed by *in situ* hybridization, qRT-PCR and ChIP assay. The ChIP assay demonstrated that STK directly regulates *BAN*, *ABS* and *EGL3* (solid line). The ChIP assay revealed that STK function

negatively correlates with the level of the H3K9ac mark on the *BAN* promoter (solid line with nucleosome).



D

E



F

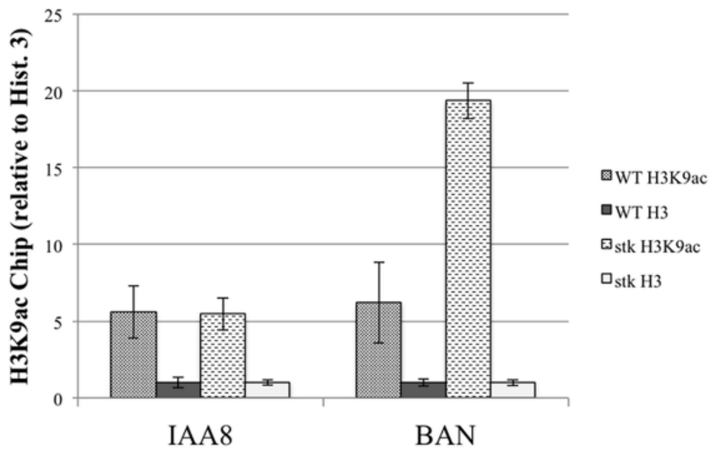


Figure 5: STK directly regulates *BAN* expression through the modification of chromatin state.

(*A*) *In situ* hybridization experiments illustrating the expression of the *BAN* transcript: in wild-type seeds at 1 DAP, *BAN* is expressed only in the endothelium layer. (*B*) *BAN* expression in the *stk* mutant background at 1 DAP and (*C*) at 4 DAP. In the *stk* mutant background, *BAN* expression is affected: the *BAN* transcript was detected not only in the endothelium layer but also in the ii2 layer.

(*D*) Schematic representation of CArG box positions indicating the regions analysed by the ChIP experiment (black bars). Black boxes: exons; white boxes: promoter, introns, 3' and 5′ UTRs. Asterisks indicate CArG boxes.

(*E*) ChIP enrichment tests by qRT-PCR show that STK binds to the selected region of *BAN*. Fold enrichment was calculated over the negative controls. Error bars represent the propagated error value using three replicates.

(*F*) ChIP enrichment tests by qRT-PCR show that STK negatively correlates with the H3K9ac acetylation mark at the *BAN* translational start site. qRT-PCR quantification of *BAN* sequences in precipitated chromatin was used to infer the acetylated histone H3 and total histone H3 representation at the STK-binding site. Levels of histone modification were normalized to total histone H3. Ct values were used to calculate the IP/IN signal. ChIP enrichments are presented as the percentage (%) of bound/input signal. ChIP enrichments for H3K9ac were normalized to histone H3 density. We tested the efficiency of IP by quantifying the presence of the H3K9ac mark in IAA8 [49] which was shown to be strongly and equally expressed in both samples and yielded equal enrichment ratios. mi, micropyle; ii1, endothelium; ii2, internal layer of inner integument. Scale bars = 40 μm (*A*-*C*).

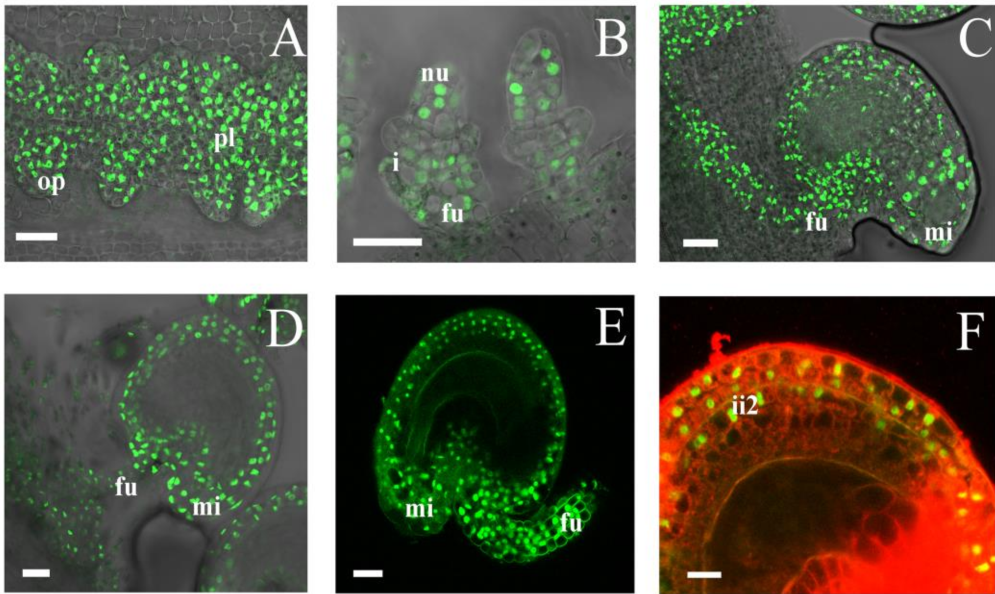


Figure 6: Confocal laser-scanning images of *pSTK::STK-GFP* expression patterns during ovule and seed development.

(*A*) Early ovule development: STK-GFP nuclear protein is expressed in the placenta and in the ovule primordia. (*B*) When integuments arise STK-GFP signal is localized in the nucellus and in the funiculus. (*C*, *D*) Mature ovule development: GFP can be detected throughout the integuments, funiculus and the adjacent placental region. (*E*) After fertilization the STK-GFP signal is present in the outer integuments and funiculus of developing seeds. (*F*) Magnification of figure E with an overlay projection images of specific PI staining to determine the presence of the GFP signal in the integuments. Protein can be detected in the two layers of the outer integument and also in the more external layer of the inner integument.

op, ovule primordia; pl, placenta; nu, nucellus; i, integuments; fu, funiculus; mi, micropyle; ii2, internal layer of inner integument. Scale bars = 50 µm (*A* and *B*), 40 µm (*C*, *D* and *E*) and 20 µm (*F*).

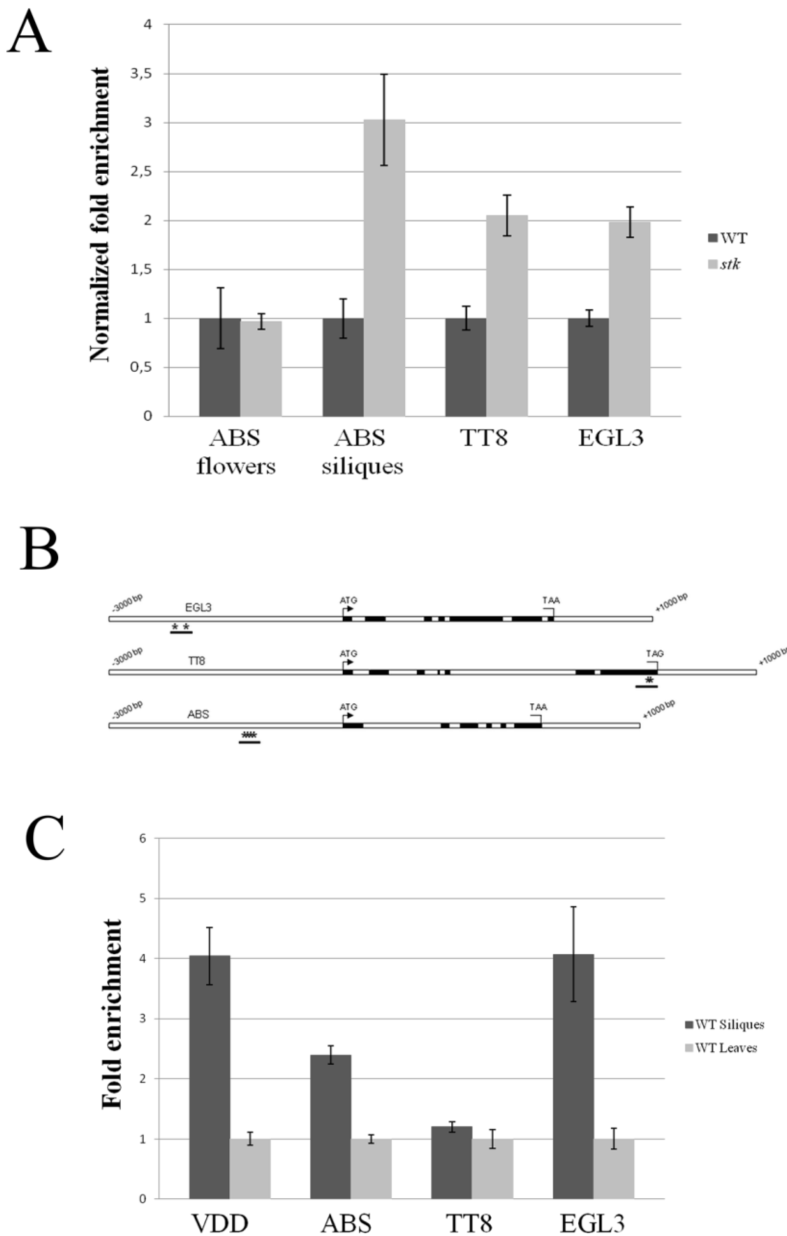


Figure 7: *BAN* regulators are direct targets of STK.

(*A*) qRT-PCR performed on cDNA obtained from siliques from 3 to 4 DAP and from unpollinated flowers for *ABS*. Relative mRNA levels indicate that the

expression of all the genes is up-regulated in the absence of the STK protein at 3-4 DAP; these differences are statistically significant as determined by Statistical Student's *t* test ($P < 0.01$). The expression level of *ABS* is not affected in unpollinated flowers in the absence of the STK protein. Error bars represent the propagated error value using three replicates.

(*B*) Schematic representation of CArG box positions.

Schematic diagrams of EGL3, ABS and TT8 loci indicating the regions analysed by the ChIP experiment (black bars). Black boxes: exons; white boxes: promoters, introns, 3' and 5' UTRs. Asterisks indicate CArG boxes.

(*C*) ChIP enrichment tests by qRT-PCR show that STK binds to the selected regions of *ABS* and *EGL3*. Fold enrichment was calculated over the negative controls. Error bars represent the propagated error value using three replicates.

Table 1. Summary of the RNA sequencing results for the genes involved in the PA pathway.

	Difference	***Fold Change***
Early Biosynthetic Genes (EBGs)		
CHALCONE SYNTHASE (CHS)	-40.86	-1.05 *
CHALCONE ISOMERASE (CHI)	-3.35	-1.04
FLAVANONE-3-HYDROXYLASE (F3H)	-5.97	-1.05
FLAVANONE-3'-HYDROXYLASE (F3'H)	+2.67	+1.04
FLAVONOL SYNTHASE (FLS)	-37.15	-1.18
Late Biosynthetic Genes (LBGs)		
DIHYDROFLAVONOL REDUCTASE (DFR)	+26.13	+2.2 **
LEUCOCYANIDIN DIOXYGENASE (LDOX)	+29.71	+2.35 **
ANTHOCYANIDIN REDUCTASE (BAN)	+45.21	+2.18 **
Regulatory Genes		
ARABIDOPSIS Bsister (ABS)	+0.33	+1.64
TRANSPARENT TESTA GLABRA 1 (TTG1)	-0.27	-1.01
TRANSPARENT TESTA GLABRA 2 (TTG2)	+0.74	+1.27
TRANSPARENT TESTA 1 (TT1)	+1.64	+2.03
TRANSPARENT TESTA 2 (TT2)	+3.9	+2.43 *
TRANSPARENT TESTA 8 (TT8)	+2.30	+1.56
Transport and Compartmentation Genes		
TRANSPARENT TESTA 12 (TT12)	+8.08	+2.35 *
TRANSPARENT TESTA 19 (TT19)	+13.98	+1.78 **
AUTOINHIBITED H+-ATPASE ISOFORM 10 (AHA10)	+4.32	+1.58

Asterisks indicate P-value * < 0.05, ** < 0.01

Table 1: Summary of the RNA sequencing results for the genes involved in the PA pathway.

SUPPLEMENTAL MATERIALS

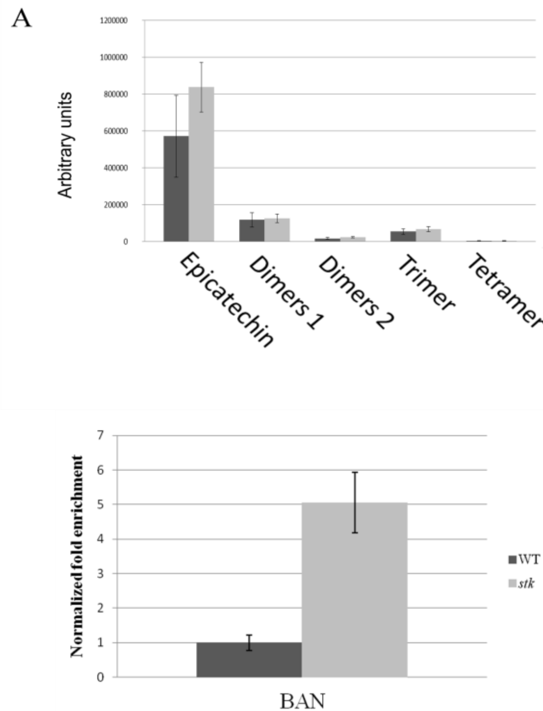


Figure S1: Insoluble PAs analysed by LC-MS.

No differences were detected in the metabolic profiles of insoluble PAs between mature wild-type (black bars) and *stk* mutant seeds (grey bars). Error bars represent SD of three independent measurements.

Figure S2: *BAN* expression is up-regulated in the *stk* mutant.

qRT-PCR performed on cDNA obtained from siliques from 0 to 6 DAP. The relative mRNA levels confirmed the result obtained by the RNA-Seq experiment indicating that the expression of BAN is up-regulated in the absence of the STK protein. Error bars represent the propagated error value using three replicates.

Table S1: List of 156 genes up-regulated in the *stk* mutant background according to the criteria described in the text. (available on-line)

Table S2: List of 90 genes down-regulated in the *stk* mutant background according to the criteria described in the text. (available on-line)

Table S3: Gene ontology analysis.
Singular enrichment analysis of GO annotation terms over-represented in the list of genes up- and down-regulated in the Arabidopsis *stk* mutant with respect to the wild-type. (available on-line)

Table S4: List of the annotated metabolites shown in Figure 3. (available on-line)

Primers for ChIP experiment	
ACT7	5'-CGTTTCGCTTTCCTTAGTGTTAGCT-3' 5'-AGCGAACGGATCTAGAGACTCACCTTG-3'
VDD	5'-GGAAATATGACGCTTGTCTTTTTAG-3' 5'-CAGAAACAGCAATATGCTCGTG-3'
BAN	5'-GATCACGTGCTTACCTTC-3' 5'-TGTGTGTAAGAGTCTGGT-3'
ABS	5'-GAAGTTTAGACCCGTTCAAG-3' 5'-ATGCTGAAACTCGAACACTG-3'
EGL3	5'-TCCGTGCAATTAATGCGAGC-3' 5'-CTAACGCCGCAGATGATGATG-3'
TT8	5'-AAAGATAAGAGGCTACCGCG-3' 5'-ATTCTCAAGCTCATGGACCC-3'
IAA8	5'-GAACGTAACTACTTGGGTCT-3' 5'-GTCTTTAGAAGGTAGCAAC-3'
MU-like	5'-AATTCCGCCTTCGAACAACTTCTCG-3' 5'-GGCAACCCTTGCTGTTGCATTTG-3'
Primers for expression analysis experiment	
ACT8	5'-CTCAGGTATTGCAGACCGTATGAG-3' 5'-CTGGACCTGCTTCATCATACTCTG-3'
UBQ	5'-CTGTTCACGGAACCCAATTC-3' 5'-GGAAAAAGGTCTGACCGACA-3'
SAND	5'-CAGACAAGGCGATGGCGATA-3' 5'-GCTTTCTCTCAAGGGTTTCTGGGT-3'
PP2a	5'-CAGCAACGAATTGTGTTTGG-3' 5'-AAATACGCCCAACGAACAAA-3'
BAN	5'- ATCTTCCATGTCGCAACTCC -3' 5'- CACACGCTTGACTGATTTCG -3'
ABS	5'-GCAGCAACAGTTGGAGAATC-3' 5'-TCCCAGCTTGTTGAAACTCC-3'
TT8	5'-ATGAAGAAGCCGAAGACGAA-3' 5'-CTTGTGGGGTGTGACATGAG-3'
EGL3	5'-AACCAGGAGTGTTGGAGTGG-3' 5'-CGGAAGCTCGTCTAGTGACC-3'

Table S5: List of primer pairs used in the ChIP and expression analysis experiments.

DISCUSSION

Seed development is a highly complex process that includes the formation of the zygote, storage tissues and a protective seed coat. Differentiation of the various structures is evidenced at the morphological level but is also reflected by the spatial distribution of metabolites. *STK* is involved in seed development in several ways, for example *stk* mutant seeds are smaller than those of wild type and do not detach from the mother plant. Furthermore, combining the *stk* mutant with the *arabidopsis b-sister* (*abs*) mutant resulted in ovules that failed to develop the endothelium layer that forms the innermost component of the seed coat [9]. To gain a deeper insight into the role of *STK* during ovule and seed development we performed RNA-Seq on material extracted from wild-type and *stk* mutant inflorescences and seeds. The resulting transcriptomic analysis yielded a list of 246 genes identified as being differentially expressed between *stk* and wild type. The majority of these (156 genes) were up-regulated in the *stk* background whilst 90 were down-regulated, which suggests that *STK* may act primarily as a repressor. Of all the deregulated transcripts, we found 18 that were expressed in the *stk* mutant but not in wild type, and 11 that were specifically expressed in the wild type but not in *stk*. GO analysis of the genes that were down-regulated in the *stk* mutant background revealed that they encoded proteins involved in DNA binding, including group II WRKY transcription factors (TFs) such as WRKY39 [51].

The importance of TTG2 (WRKY group I) as a regulator of the LBG biosynthetic pathway has been described previously [16]. Our RNA-Seq data revealed that expression levels of TTG2 are unaltered between wild type and *stk*. WRKY TFs are global regulators acting at various levels, including the direct modulation of immediately downstream target genes, but they also appear

to interact with key chromatin-remodelling factors [52]. The latter could be of special interest since we found that the STK-dependent *BAN* regulatory mechanism involves a chromatin remodelling activity which may imply the action of downstream STK targets acting as remodelers that have yet to be defined. In this regard, the nucleosome assembly/disassembly protein NAP1-RELATED PROTEIN 1 (NRP1) which interacts with chromatin remodelling factors [53] was also found in this list. In addition we identified NAC transcription factor-like 9 (NTL9), a calmodulin-regulated NAC transcriptional repressor in Arabidopsis [54], and SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 2 (SPL2), a member of the SQUAMOSA PROMOTER BINDING PROTEIN (SBP)-box family of transcription factors [55]. Future analysis will be directed to determine the possible mechanism(s) of action of STK.

GO analysis of genes that are up-regulated in the *stk* mutant showed an abundance of transcripts involved in secondary metabolic processes, including those involved in flavonoid and phenylpropanoid biosynthesis. We hypothesize that STK may act on the PA pathway at two different levels: i) impacting directly on the elements comprising the enzymatic pathway of PA biosynthesis, and/or ii) influencing key regulatory elements. In the first case, all the genes of the Late Biosynthetic step were found to be up-regulated in the *stk* mutant. However, no changes were found for the group of genes involved in the metabolic steps corresponding to early synthesis. This may indicate that STK regulation does not involve major alterations at the level of production of dihydroflavonols, the precursors to PAs, but impact instead on the synthesis of anthocyandins and proanthocyanidins. We also found that two out of the three genes involved in the pathway controlling transport and compartmentation were up-regulated: TT12 and TT19. Since anthocyanins are transported from the

cytosol to the vacuole as part of the PA biosynthetic pathway, it may be reasonable to expect coordinated regulation of transporter-mediated and vesicle-mediated mechanisms by STK. As commented above, STK may also act on key regulatory elements. In this regard we observed a strong influence on the expression levels of the members of the bHLH transcription factor family (*TT8* and *EGL3*), the Zinc-Finger transcription factor (*TT1*) and the R2R3 MYB domain putative transcription factor (*TT2*) in the *stk* mutant background. These regulatory elements together with other key MADS-domain transcription factors like *ABS* were all found to be up-regulated.

Based on observation from the RNA-Seq data we investigated the role of STK in PA biosynthesis in more detail. Histological analysis revealed that PAs accumulated ectopically in the outer-layer of the inner integument in *stk* mutant seeds, and analysis by LC-MS analysis demonstrated that the level of soluble epicatechin monomers was greater in the *stk* mutant in both immature and mature seeds. The latter data confirm the morphological analysis (Figure 2) and support the idea that STK is involved in PA accumulation and in particular the accumulation of epicatechin monomer.

Regulation of the production of flavonoids and proanthocyanidins has been extensively studied (for review see [16]). Several transcription factors are known to be involved in the regulation of *BAN* expression, and in particular a R2R3-MYB/bHLH/WDR complex is responsible for *BAN* activation in the endothelium [25,40–42]. Not all the members of this complex exhibit the same hierarchy, for instance it has been suggested that TT8 is necessary for activation of the LBGs whereas a predominant role of TT2 controlling via a positive feedback loop is required to maintain the transcript levels of both *TT8* and *BAN* [2,25,56]. Interestingly our RNA-Seq dataset highlighted a repressive role for STK on the regulation of *TT2*. Another gene required for the activation of *BAN*

in the endothelium is *ABS*. This gene is a member of the MADS-domain gene family and to date it is the only MADS-domain gene identified to be involved in the control of PA accumulation. *abs* mutants show significant reductions in epicatechin and procyanidins accumulation [10]; moreover, in the *abs* background both GUS reporter gene expression driven off the *BAN* promoter and qPCR analysis of *BAN* expression itself showed significant down-regulation [28,57]. In the complex regulatory network that governs PA accumulation in the endothelium layer, STK plays a key role controlling *BAN* expression, and our results suggest that this occurs through the binding of STK to the *BAN* regulatory region. Recently, Dean and collaborators performed genome-wide expression profiling using microarrays to identify those genes differentially expressed in the wild-type and *abs* mutant seed coats [57]. This showed that *STK* expression is unaffected in the *abs* mutant at 3 DAP but is up-regulated at 7 DAP. Furthermore, Nesi and colleagues demonstrated that ectopic expression of *ABS* causes altered PA accumulation in a manner very similar to that which we have observed in the *stk* mutant, since these plants present PAs not only in the endothelium but also in the more external layer of the inner integuments [28]. Integration of previous data with our results indicates that STK is a master regulator of inner seed coat differentiation. Whilst several transcription factors involved in seed coat determination or secondary metabolite accumulation have been already characterized [58], no gene connecting these processes has been described previously. In our study we provide evidence of a role for chromatin modification, specifically H3K9 acetylation, in the transcriptional regulation of *BAN*. We have shown that the region of the *BAN* promoter proximal to the translational start site is heavily covered by this epigenetic marker of transcriptional activity. H3K9ac enrichment is greater when *BAN* is ectopically expressed due to the lack of STK

protein. This suggests that STK somehow represses the activity of histone deacetylases (HDACs) at the *BAN* locus in cells of the outer layer of the inner integument. It will be interesting to determine whether STK is able to recruit a chromatin remodelling partner, yet to be identified, that would form part of a hypothetical STK complex. It is known that histone acetyltransferases (HATs) and HDACs participate in the genome-wide turnover of acetyl groups, and that besides histones some also modify other factors. Future progress will therefore be focused on determining their availability for interaction with specific transcription factors like STK and other protein complex partners.

PA biosynthesis and its spatial accumulation are under the control of a complex regulatory network. *BAN* is one of the key genes of the LBG group and its expression falls under the influence of several transcription factors, including ABS, EGL3 and TT8. Our work adds STK to this list. Our data also suggest STK to be a master regulator of PA biosynthesis and accumulation since we observed that it also controls the expression of *ABS* and *EGL3*. This shows that STK not only acts as a regulator of ovule identity but also orchestrates important aspects of seed development, adding new evidence for the importance of MADS-domain genes in the control of plant developmental processes.

The role of *STK* in the regulation of epicatechin accumulation could have relevance for certain agricultural applications: PAs are important in several aspects of plant protection and their significance in the flavour and astringency of foods and beverages is already known (for review see [59]). Indeed, it has been demonstrated that avocado fruits that contain higher levels of epicatechin exhibit stronger resistance to fungal attack [60]. In this regard it will be interesting to study the fungal resistance of *stk* mutant seeds since regulation of STK levels might provide a tool to make plant seeds more resistant to fungi.

MATERIAL AND METHODS

Plant Material and Growth Conditions

Arabidopsis thaliana wild-type (ecotype Columbia) and *stk* mutant plants were grown at 22°C under short-day (8 h light/16 h dark) or long-day (16 h light/8 h dark) conditions. The Arabidopsis *stk* mutant was kindly provided by M. Yanofsky [7]. The *stk-2* allele contains a 74 nucleotide insertion near the splice site of the third intron.

PCR-Based Genotyping

Identification of *STK* wild-type and mutant alleles was performed by PCR analysis using oligonucleotides AtP_204 (5'-GCTTGTTCTGATAGCACCAACACTAGCA-3') and AtP_561 (5'-GGAACTCAAAGAGTCTCCCATCAG-3'). The mutant allele yields a 399 bp DNA fragment whilst the wild-type allele produces a 325 bp fragment.

RNA Extraction, cDNA Library Preparation, and Sequencing for RNA-Seq

Total RNA was extracted from two biological replicates (1 gr) from both wild-type and *stk* mutant inflorescences and siliques until 5 DAP using the Qiagen 'RNeasy miniKit' according to the manufacturer's instructions. DNA contamination was removed using PROMEGA RQ1 RNase-Free DNase according to the manufacturer's instructions. RNA quality and integrity were analyzed by gel electrophoresis and validated on a Bioanalyzer 2100 (Aligent, Santa Clara, CA); RNA Integrity Number (RIN) values were greater than 7 for all samples. In order to confirm that the *stk* mutant was a knock-out line, *STK* expression was checked by qRT-PCR with primers RT_780 (5’-

TGCGATGCAGAAGTTGCGCTC-3') and RT_781 (5'-AGTACGCGGCATTGATTTCTTG-3'). Sequencing libraries were prepared according to the manufacturer's instructions using the TruSeq RNA Sample Prep kit (Illumina Inc.) and sequenced on an Illumina HiSeq2000 (50 bp single-read). The processing of fluorescent images into sequences, base-calling and quality value calculations were performed using the Illumina data processing pipeline (version 1.8). Raw reads were filtered to obtain high-quality reads by removing low-quality reads containing more than 30% bases with $Q < 20$. Finally, quality control of the raw sequence data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Mapping of Short Reads, Quality Analysis and Assessment of Gene Expression Analysis for RNA-Seq

Evaluation and treatment of raw data was performed on the commercially available CLC Genomics Workbench v.4.7.1 (http://www.clcbio.com/genomics/). After trimming, the resulting high-quality reads were mapped onto the Arabidopsis genome (TAIR10). Approximately 25M reads of each sample that mapped with ≤2 mismatches were used for further analyses. The read number of each gene model was computed based on the coordinates of the mapped reads. A read was counted if any portion of that read's coordinates were included within a gene model. As CLC Genomics Workbench v.4.7.1 distributes multireads at similar loci in proportion to the number of unique reads recorded and normalized by transcript length, we included both unique reads and reads that occur up to 10 times in the analysis to avoid undercount for genes that have closely related paralogs [36]. Gene expression values were based on reads per kilobase of exon model per million mapped read (RPKM) values [36]. The fold change and differential expression

values between wild type and the *stk* mutant was calculated in terms of RPKM of the corresponding transcripts. Statistical analysis of biological replicates was assessed using heat map visualization of Euclidean distances. This clustered the biological replicates in the wild-type and mutant groups as expected. We further checked whether the overall variability of the samples reflected their grouping by performing Principal Component Analysis (PCA). This confirmed that the replicates were relatively homogenous and distinguishable from the samples of the other group. Finally, the overall distribution of expression values between the different samples confirmed that none of the samples stood out from the rest. To obtain statistical confirmation of the differences in gene expression, P and FDR values were computed using Baggerley's test on expression proportions. We applied a threshold value of P=0,05 to ensure that differential gene expression was maintained at a significant level (5%) for the individual statistical tests. Transcripts that exhibited an estimated absolute Fold Change ≥ 2 (*i.e.* 2 mapped reads per kilobase of mRNA) were determined to be significantly differentially expressed. To gain insight into the biological processes associated with the regulated genes, we determined which GO annotation terms were over-represented, in both the up-regulated (Table S1) and down-regulated (Table S2) lists. Gene set enrichment analysis was performed with the agriGO database [38] using the Singular Enrichment Analysis (SEA).

Histological Analysis

For the morphological analysis of integuments, *Arabidopsis thaliana* wild-type (ecotype Columbia) and *stk* mutant plants were fixed for Technovit 7100 embedding (Heraeus Kulzer) following the manufacturer's instructions. Sections of plant tissue (0.8 μm) were stained in 0.5% (w/v) toluidine blue O. Samples were observed using a Zeiss Axiophot D1 microscope

(http://zeiss.com/) equipped with differential interface contrast (DIC) optics. Images were recorded with an Axiocam MRc5 camera (Zeiss) using the Axiovision program (version 4.1). The whole-mount vanillin assay for PA detection was performed as described previously [61]. Vanillin (vanilaldehyde) condenses specifically with PAs and flavan-3-ol precursors to yield a bright-red product under acidic conditions. Microscopic observations were performed as detailed above.

Preparation of Extracts

Three biological replicates of immature and mature seeds (30 mg) were frozen in liquid nitrogen and ground to a fine powder using an analytical mill (IKA; A11 basic). The soluble PA fraction was extracted with 75% methanol:water (v/v) containing 0.1% formic acid. The mixture was vortexed, sonicated for 30 min at room temperature, vortexed again and then centrifuged (13000 rpm, 10 min). Soluble PAs (contained in the supernatant) were analyzed by LC-MS. The pellet was washed with 50% methanol:water, centrifuged (13000 rpm, 10 min), washed again with 100% methanol and centrifuged (13000 rpm, 10 min). Samples were dried in a speed vacuum for 1 hour at 30°C and then hydrolyzed with NaOH (2N) for 15 min at 60°C. The mixture was vortexed and HCl (4N) was added. To remove lipids hexane was added to the mixture and after centrifugation (13000 rpm, 10 min) the upper phase was removed. The insoluble PAs were extracted three times with ethyl acetate, and the three extractions were combined and dried in a speed vacuum for 1 hour at 30°C. The pellet was dissolved in acetone:water:acetic acid (70:29.5:0.5 v/v/v), vortexed, sonicated for 15 min and centrifuged (13000 rpm, 10 min). This PA fraction was analyzed by LC-MS.

Targeted Profiling of Proanthocyanidins

The profiling of PAs in extracts of seeds was performed by MS analysis using the UPLC-qTOF instrument (Waters High Definition MS System; Synapt) with the UPLC column connected online to a photodiode array detector (Waters, Acquity) and then to the MS detector equipped with an electrospray probe. The separation of metabolites and detection of the eluted compound masses was performed as previously described [62–66].

Construction of Binary Vectors and Plant Transformation

For the construction of *pSTK::STK-GFP*, DNA fragments containing the 3,5 kb promoter region and the complete *STK* genomic region without the stop codon were amplified from wild-type genomic DNA and cloned into a pGreen II binary vector containing a *GFP* reporter gene cassette. For amplification we used the following primers containing attB1 and attB2 recombination sequences: AtP_3066 5’-GGGGACAAGTTTGTACAAAAAAGCAGGCTCCAACCAATATCACACCCTAAATAC-3’ and AtP_3067 5’-GGGGACCACTTTGTACAAGAAAGCTGGGTCGTCCGAGATGAAGAATTTTCTTGTC-3’.

The resulting binary vectors were transformed into *Agrobacterium tumefaciens* by electroporation prior to stable transformation of plants using the floral dip method. Transformant lines were obtained using BASTA as a selection agent. Resistant transgenic plants showing strong GFP fluorescence were genotyped and those that complemented the *stk* phenotype were selected. Protein expression patterns were analyzed by Confocal Laser Scanning Microscopy (CLSM). Fresh material was collected, mounted in 10 mg/ml of propidium iodide (Sigma P-4170) in water and immediately analyzed. CLSM analysis was

performed using a Leica TCS SPE with a 488 nm argon laser line for excitation of GFP fluorescence; emissions were detected between 505 and 580 nm. For PI fluorescence a 543 nm laser line was used and emissions were detected between 600 and 640 nm. Confocal scans were performed with the pinhole at 1 airy unit. Images were collected in the multi-channel mode and the overlay images were generated using the Leica analysis software LAS AF 2.2.0.

Expression Analysis by *in situ* Hybridization

DIG-labelled RNA probes for detection and hybridization of *BAN* were prepared as previously described [67]. Sections of plant tissue were hybridized with digoxigenin-labelled *BAN* antisense probe amplified using primers AtP_4331 (5'-CGAGTAGC TTATCTCTCTCG -3') and AtP_4332 (5'-TCAATCCTTTTGACTCGAAG -3').

ChIP and qRT-PCR Analysis

The genomic regions located 3 kb upstream of the ATG, 1 kb downstream of the stop codon, and in the exons and introns of the selected genes were analyzed to identify CArG box sequences with up to one base mismatch. ChIP experiments were performed in a modified version of a previously reported protocol [68]. The qRT-PCR assay was conducted in triplicate on three different biological replicates, with three technical replicates for each sample, and was performed in a Bio-Rad iCycler iQ optical system (software version 3.0a). ChIP efficiency was determined using the third CArG box of the *VDD* gene as a positive control [43]. Fold enrichment was calculated using the formulae of a previously reported protocol [43].

For ChIP-based analysis of histone modifications, the following antibodies were used for immunoprecipitation: rabbit anti-histone H3 (Sigma-Aldrich

H0164) and rabbit anti-H3 acetyl K9 (Upstate 07-352) and were handled in parallel to samples lacking antibody. qRT-PCRs were performed on input and immunoprecipitated samples and % of input was calculated. The signal obtained after precipitation with anti-H3K9ac antibody (as indicated in the figure) was normalized to the signal obtained by precipitation with an antibody to an invariant domain of histone H3. For H3K9ac analyses, *IAA8* (*At2g22670*) was used as a reference as it carries the H3K9ac mark and is equally expressed in both samples [49]. Relative enrichment of *Mu-like* transposons was included as negative control. Sequences of oligonucleotides used for ChIP analyses are listed in Table S5.

Expression Analysis

qRT-PCR experiments were performed on cDNA obtained from siliques from 3 to 4 DAP and unpollinated flowers. Total RNA was extracted using the LiCl method [69]. DNA contamination was removed using the Ambion TURBO DNA-free DNase kit according to the manufacturer's instructions. The treated RNA was reverse transcribed using the ImProm-II reverse transcription system (Promega). cDNAs were used as templates in the qRT-PCR reactions containing the iQ SYBR Green Supermix (Bio-Rad). The qRT-PCR assay was conducted in triplicate on three different biological replicates, with three technical replicates for each sample, and was performed in a Bio-Rad iCycler iQ Optical System (software version 3.0a). Relative transcript enrichment of genes of interest was calculated normalizing the amount of mRNA against different endogenous control fragments (*UBQ*, *ACT*, *PPa2* and *SAND* [70]). The difference between the cycle threshold (Ct) of the gene and that of the reference gene (ΔCt = CtGENE – CtREFERENCE) was used to obtain the

normalized expression of that gene, which corresponds to $2^{-\Delta\Delta Ct}$. The primers used for this analysis are listed in Table S5.

Accession Number

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession number: *SEEDSTICK*, *At4g09960*; *BANYULS*, *At1g61720*.

ACKNOWLEDGMENTS

We would like to thank Martin Yanofsky (Section of Cell and Developmental Biology, University of California at San Diego, CA, USA) for providing the *stk-2* mutant; Marta A. Mendes for technical help and Andrew MacCabe for editing the paper.

REFERENCES

1. Haughn G, Chaudhury A (2005) Genetic analysis of seed coat development in Arabidopsis. Trends Plant Sci 10: 472–477. doi:10.1016/j.tplants.2005.08.005.

2. Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, et al. (2003) Proanthocyanidin-Accumulating Cells in Arabidopsis Testa: regulation of differentiation and role in seed development. Plant Cell 15: 2514–2531. doi:10.1105/tpc.014043.1.

3. Kaufmann K, Muiño JM, Jauregui R, Airoldi C, Smaczniak C, et al. (2009) Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biol 7: e1000090. doi:10.1371/journal.pbio.1000090.

4. Gregis V, Andrés F, Sessa A, Guerra RF, Simonini S, et al. (2013) Identification of pathways directly regulated by SHORT VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. Genome biol 14: R56. doi:10.1186/gb-2013-14-6-r56.

5. Colombo L, Battaglia R, Kater MM (2008) Arabidopsis ovule development and its evolutionary conservation. Trends Plant Sci 13: 444–450. doi:http://dx.doi.org/10.1016/j.tplants.2008.04.011.

6. Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, et al. (2003) MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. Plant Cell 15: 2603–2611. doi:10.1105/tpc.015123.2.

7. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, et al. (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature 424: 85–88. doi:10.1038/nature01741.

8. Brambilla V, Battaglia R, Colombo M, Masiero S, Bencivenga S, et al. (2007) Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in Arabidopsis. Plant Cell 19: 2544–2556. doi:10.1105/tpc.107.051797.

9. Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, et al. (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. Plant J 70: 409–420. doi:10.1111/j.1365-313X.2011.04878.x.

10. Routaboul J-M, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, et al. (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta 224: 96–107. doi:10.1007/s00425-005-0197-5.

11. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126: 485–493. doi:10.1104/pp.126.2.485.

12. Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol 5: 218–223. doi:10.1016/S1369-5266(02)00256-X.

13. Bais HP, Vepachedu R, Gilroy S, Callaway RM, Vivanco JM (2003) Allelopathy and exotic plant invasion: from molecules and genes to species interactions. Science 301: 1377–1380. doi:10.1126/science.1083245.

14. D'Auria JC, Gershenzon J (2005) The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. Curr Opin Plant Biol 8: 308–316. doi:10.1016/j.pbi.2005.03.012.

15. Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10: 236–242. doi:10.1016/j.tplants.2005.03.002.

16. Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, et al. (2006) Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol 57: 405–430. doi:10.1146/annurev.arplant.57.032905.105252.

17. Abrahams S, Lee E, Walker AR, Tanner GJ, Larkin PJ, et al. (2003) The Arabidopsis TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. Plant J 35: 624–636. doi:10.1046/j.1365-313X.2003.01834.x.

18. Pelletier MK, Burbulis IE, Winkel-Shirley B (1999) Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and end-products in Arabidopsis seedlings. Plant Mol Biol 40: 45–54. doi:10.1023/A:1026414301100.

19. Xu W, Grain D, Bobet S, Le Gourrierec J, Thévenin J, et al. (2014) Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in Arabidopsis seed. New phytol 202: 132–144. doi:10.1111/nph.12620.

20. Devic M, Guilleminot J, Debeaujon I, Bechtold N, Bensaude E, et al. (1999) The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J 19: 387–398. doi:10.1046/j.1365-313X.1999.00529.x.

21. Kitamura S, Shikazono N, Tanaka A (2004) TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. Plant J 37: 104–114. doi:10.1046/j.1365-313X.2003.01943.x.

22. Baxter IR, Young JC, Armstrong G, Foster N, Bogenschutz N, et al. (2005) A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. Proc Natl Acad Sci USA 102: 2649–2654. doi:10.1073/pnas.0406377102.

23. Pourcel L, Routaboul J-M, Kerhoas L, Caboche M, Loic Lepiniec, et al. (2005) TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. Plant Cell 17: 2966–2980. doi:10.1105/tpc.105.035154.1.

24. Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, et al. (2007) The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19: 2023–2038. doi:10.1105/tpc.106.046029.

25. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, et al. (2004) TT2, TT8, and TTG1 synergistically specify the expression of

BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J 39: 366–380. doi:10.1111/j.1365-313X.2004.02138.x.

26. Zhao L, Gao L, Wang H, Chen X, Wang Y, et al. (2013) The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. Funct Integr Genomics 13: 75–98. doi:10.1007/s10142-012-0301-4.

27. Johnson CS, Kolevski B, Smyth DR (2002) TRANSPARENT TESTA GLABRA2 , a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell 14: 1359–1375. doi:10.1105/tpc.001404.

28. Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, et al. (2002) The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14: 2463–2479. doi:10.1105/tpc.004127.

29. Sagasser M, Lu G, Hahlbrock K, Weisshaar B (2002) A. thaliana TRANSPARENT TESTA 1 is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes Dev 16: 138–149. doi:10.1101/gad.212702.

30. Madhani HD, Fink GR (1998) The control of filamentous differentiation and virulence in fungi. Trends Cell Biol 8: 348–353. doi:10.1016/S0962-8924(98)01298-7.

31. Gagiano M, Bauer FF, Pretorius IS (2002) The sensing of nutritional status and the relationship to filamentous growth in *Saccharomyces cerevisiae*. FEMS Yeast Res 2: 433–470. doi:10.1111/j.1567-1364.2002.tb00114.x.

32. Bracha AL, Ramanathan A, Huang S, Ingber DE, Schreiber SL (2010) Carbon metabolism-mediated myogenic differentiation. Nat Chem Biol 6: 202–204. doi:10.1038/nchembio.301.

33. McGraw TE, Mittal V (2010) Stem cells: Metabolism regulates differentiation. Nat Chem Biol 6: 176–177. doi:10.1038/nchembio.324.

34. Folmes CDL, Park S, Terzic A (2013) Lipid metabolism greases the stem cell engine. Cell Metab 17: 153–155. doi:10.1016/j.cmet.2013.01.010.

35. Shiraki N, Shiraki Y, Tsuyama T, Obata F, Miura M, et al. (2014) Methionine metabolism regulates maintenance and differentiation of human pluripotent stem cells. Cell Metab 19: 780–794. doi:10.1016/j.cmet.2014.03.017.

36. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621–628. doi:10.1038/nmeth.1226.

37. Baggerly KA, Deng L, Morris JS, Aldaz CM (2003) Differential expression in SAGE: accounting for normal between-library variation. Bioinformatics 19: 1477–1483. doi:10.1093/bioinformatics/btg173.

38. Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38: 64–70. doi:10.1093/nar/gkq310.

39. Doughty J, Aljabri M, Scott R (2014) Flavonoids and the regulation of seed size in Arabidopsis. Biochem Soc Trans 42: 364–369. doi:10.1042/BST20140040.

40. Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development 130: 4859–4869. doi:10.1242/dev.00681.

41. Gonzalez A, Mendenhall J, Huo Y, Lloyd A (2009) TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Dev Biol 325: 412–421. doi:10.1016/j.ydbio.2008.10.005.

42. Li SF, Milliken ON, Pham H, Seyit R, Napoli R, et al. (2009) The Arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. Plant Cell 21: 72–89. doi:10.1105/tpc.108.063503.

43. Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, et al. (2010) VERDANDI is a direct target of the MADS domain ovule

identity complex and affects embryo sac differentiation in Arabidopsis. Plant Cell 22: 1702–1715. doi:10.1105/tpc.109.068627.

44. Riechmann JL, Meyerowitz EM (1997) Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA-binding specificity. Mol Biol Cell 8: 1243–1259.

45. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, et al. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40: 897–903. doi:10.1038/ng.154.

46. Ausín I, Alonso-Blanco C, Jarillo J a, Ruiz-García L, Martínez-Zapater JM (2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nat Genet 36: 162–166. doi:10.1038/ng1295.

47. Benhamed M, Bertrand C, Servet C, Zhou D-X (2006) Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. Plant Cell 18: 2893–2903. doi:10.1105/tpc.106.043489.

48. Ng DW, Chandrasekharan MB, Hall TC (2006) Ordered histone modifications are associated with transcriptional poising and activation of the phaseolin promoter. Plant Cell 18: 119–132. doi:10.1105/tpc.105.037010.

49. Zhou J, Wang X, He K, Charron J-BF, Elling A a, et al. (2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression. Plant Mol Biol 72: 585–595. doi:10.1007/s11103-009-9594-7.

50. De Folter S, Shchennikova A V, Franken J, Busscher M, Baskar R, et al. (2006) A Bsister MADS-box gene involved in ovule and seed development in petunia and Arabidopsis. Plant J 47: 934–946. doi:10.1111/j.1365-313X.2006.02846.x.

51. Li S, Zhou X, Chen L, Huang W, Yu D (2010) Functional characterization of *Arabidopsis thaliana* WRKY39 in heat stress. Mol Cell 29: 475–483. doi:10.1007/s10059-010-0059-2.

52. Chi Y, Yang Y, Zhou Y, Zhou J, Fan B, et al. (2013) Protein-protein interactions in the regulation of WRKY transcription factors. Mol Plant 6: 287–300. Available: http://www.ncbi.nlm.nih.gov/pubmed/23455420. Accessed 24 July 2014.

53. Gao J, Zhu Y, Zhou W, Molinier J, Dong A, et al. (2012) NAP1 family histone chaperones are required for somatic homologous recombination in Arabidopsis. Plant Cell 24: 1437–1447. doi:10.1105/tpc.112.096792.

54. Kim H, Park H, Kim K, Jung M, Han H, et al. (2012) A NAC transcription factor and SNI1 cooperatively suppress basal pathogen resistance in *Arabidopsis thaliana*. Nucleic Acids Res 40: 9182–9192. doi:10.1093/nar/gks683.

55. Shikata M, Koyama T, Mitsuda N, Ohme-Takagi M (2009) Arabidopsis SBP-box genes SPL10, SPL11 and SPL2 control morphological change in association with shoot maturation in the reproductive phase. Plant Cell Physiol 50: 2133–2145. doi:10.1093/pcp/pcp148.

56. Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L (2001) The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13: 2099–2114.

57. Dean G, Cao Y, Xiang D, Provart NJ, Ramsay L, et al. (2011) Analysis of gene expression patterns during seed coat development in Arabidopsis. Mol Plant 4: 1074–1091. doi:10.1093/mp/ssr040.

58. Khan D, Chan A, Millar JL, Girard IJ, Belmonte MF (2014) Plant Science Predicting transcriptional circuitry underlying seed coat development. Plant science 223: 146–152. doi:10.1016/j.plantsci.2014.03.016.

59. Dixon R a, Xie D-Y, Sharma SB (2005) Proanthocyanidins--a final frontier in flavonoid research? New phytol 165: 9–28. doi:10.1111/j.1469-8137.2004.01217.x.

60. Ardi R, Kobiler I, Jacoby B, Keen NT, Prusky D (1998) Involvement of epicatechin biosynthesis in the activation of the mechanism of resistance

of avocado fruits to Colletotrichum gloeosporioides. Physiol Mol Plant Pathol 53: 269–285. doi:10.1006/pmpp.1998.0181.

61. Debeaujon I, Léon-Kloosterziel KM, Koornneef M (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. Plant Physiol 122: 403–414. doi:10.1104/pp.122.2.403.

62. Hanhineva K, Rogachev I, Kokko H, Mintz-Oron S, Venger I, et al. (2008) Non-targeted analysis of spatial metabolite composition in strawberry (Fragaria X ananassa) flowers. Phytochemistry 69: 2463–2481. doi:10.1016/j.phytochem.2008.07.009.

63. Mintz-Oron S, Mandel T, Rogachev I, Feldberg L, Lotan O, et al. (2008) Gene expression and metabolism in tomato fruit surface tissues. Plant Physiol 147: 823–851. doi:10.1104/pp.108.116004.

64. Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, et al. (2009) Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. PLoS Genet 5: e1000777. doi:10.1371/journal.pgen.1000777.

65. Kitamura S, Matsuda F, Tohge T, Yonekura-Sakakibara K, Yamazaki M, et al. (2010) Metabolic profiling and cytological analysis of proanthocyanidins in immature seeds of *Arabidopsis thaliana* flavonoid accumulation mutants. Plant J 62: 549–559. doi:10.1111/j.1365-313X.2010.04174.x.

66. Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, et al. (2011) GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. Plant Cell 23: 4507–4525. doi: 10.1105/tpc.111.088732

67. Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, et al. (2000) Molecular and genetic analyses of the silky1 gene reveal conservation in floral organ specification between eudicots and monocots. Mol Cell 5: 569–579. doi:10.1016/S1097-2765(00)80450-5.

68. Gregis V, Sessa A, Colombo L, Kater MM (2008) AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity

in Arabidopsis. Plant J 56: 891–902. doi:10.1111/j.1365-313X.2008.03648.x.

69. Verwoerd T, Dekker B, Hoekema A (1989) A small-scale procedure for the rapid isolation of plant RNAs. Nucleic Acids Res 17: 2362.

70. Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH (2010) Identification and testing of superior reference genes for a starting pool of transcript normalization in Arabidopsis. Plant Cell Physiol 51: 1694–1706. doi:10.1093/pcp/pcq128.

Networks controlling seed size in *Arabidopsis*

Authors: Gregorio Orozco-Arroyo, Dario Paolo, Ignacio Ezquer, Lucia Colombo*

Authors affiliation:

Dipartimento di BioScienze, Università degli Studi di Milano, Via Giovanni Celoria 26 - 20133 Milano, Italy

* To whom correspondence should be addressed (lucia.colombo@unimi.it)

INDEX

1. Abstract
2. Introduction
3. Endosperm development
    3.1. Genomic and parent-of-origin effects
4. The role of seed coat in seed size determination
    4.1. Factors controlling integuments cell proliferation
    4.2. Factors controlling integuments cell elongation
5. Endosperm-integument crosstalk
6. The role of hormone synthesis and perception in determining seed size
    6.1. Key role of brassinosteroids in seed size regulation
    6.2. The role of auxins in communication
    6.3. Cytokinins
    6.4 Abscisic acid
7. Future Perspectives
8. References

1. Abstract

Human and livestock nutrition is largely based on calories derived from seeds, in particular cereals and legumes. Unveiling the control of seed size is therefore of remarkable importance in the frame of developing new strategies for the crop improvement. The networks controlling the development of the seed coat, the endosperm and the embryo, as well as their interplay, have been described in *Arabidopsis thaliana*. In this review we provide an organic and comprehensive description of the current knowledge about the molecular mechanisms controlling seed size in *Arabidopsis*.

2. Introduction

Increasing seed production is a key goal to meet world demand and consumption of agricultural crops, for food and feed in emerging economies. In this context, the study of the molecular mechanisms controlling seed formation becomes essential for plant scientists as seed size is a major component of seed yield (Adamski et al 2009). Thus, advances in the basic knowledge about seed development in the model species *Arabidopsis thaliana* are of key relevance for the rational design of genetically engineered traits in relevant agronomic crop species that could complement and improve upon traditional breeding systems (Varshney et al 2009; Langridge and Fleury 2011; Feuillet et al 2011; Becker et al 2014).

Arabidopsis seed development (see Fig. 1) starts after a double fertilization event (for a complete seed development review see Nowack et al 2010; Becker et al 2014). During the first fertilization event, the zygotic embryo is generated by the fusion of the egg cell and one sperm cell. The second fertilization event, which triggers the development of the triploid endosperm, starts with the fusion of the central cell of the embryo sac with the second pollen sperm cell (endosperm development is reviewed by Lafon-Placette and Köhler, 2014). The two biparentally derived fertilization products (the embryo and the endosperm) are encased by the maternal sporophytic tissue (the seed coat), which is derived from the ovule integuments (seed coat development has been reviewed recently by Khan et al 2014; Figueiredo and Köhler 2014). The seed coat represents a protective layer that prevents damage from external factors such as UV radiation, toxic chemicals and pathogens, as well as impeding germination until conditions are favorable (Haughn and Chaudhury 2005). Furthermore, the seed

coat plays a major role in controlling communication between the two generations (reviewed by Bencivenga et al 2011).

In spite of the influence of several abiotic factors on plant growth and development, such as temperature, light, day length, etc., the final size of plant organs is reasonably constant within a given species (Tsukaya 2006), indicating that it is mainly the genetic seed developmental plan which determines the rate of growth until the seed reaches a predetermined mass and final size (Conlon and Raff 1999; Day and Lawrence 2000). *Arabidopsis* seed size is mainly attained either during the rapid proliferation and growth of the endosperm (Boisnard-Lorig et al 2001) and proliferation of the seed coat cells. These events span from fertilization to six days after pollination (DAP) of seed development (Fig. 1). From seven to 13 DAP, there is a residual increase in seed volume occurring when the embryo expands at the expense of the endosperm. At this point, seed growth is limited by the seed coat that acts as a constraining physical barrier (Fang et al 2012). Thus, to understand the whole mechanism governing seed size, it is essential to unveil both the mechanisms of endosperm and integument growth and development, as well as the interplay existing between the developmental programs of these structures.

In the last decades, many key regulators of seed size have been identified (reviewed by Kesavan et al 2013 – summarized in Table 1). However, there are still major gaps in knowledge regarding seed size and the available data are still fragmentary and need to be assembled into a global and coherent picture (see Fig. 2). This review provides a summary and an update of the different pathways controlling seed size in *Arabidopsis*. We analyzed seed size regulation in *Arabidopsis*, focusing on different functional categories in order to better describe them singularly. This includes mechanisms underlying the developmental processes of (A) the endosperm, including genomic imprinting

and parent-of-origin effects, and (B) the seed coat / integuments. Moreover, we discuss (C) the crosstalk between endosperm and seed coat and the role of (D) hormone synthesis and perception in determining seed size.

3. Endosperm development

Successful seed development requires the synchronized growth of the endosperm, the embryo and the seed coat (Fig. 1). Coordinated growth and development between these structures is reached through exchange of signals whose nature is still unknown. The profound morphological changes that characterize seed coat development could start only if the endosperm undergoes its developmental program, as embryo development by itself is not sufficient to stimulate seed coat growth and differentiation (Nowack et al 2007; Hehenberger et al 2012). However, it was demonstrated that central cell nuclei could start to proliferate even in the absence of karyogamy between central cell and sperm nucleus (Guitton et al 2004). Notably, despite the absence of the second fertilization event in *FIS-PRC2* (*FERTILIZATION INDEPENDENT SEED - POLYCOMB REPRESSIVE COMPLEX2*) mutants, which develop endosperm autonomously, embryogenesis is completed and viable plants are produced (Nowack et al 2007). Endosperm development has four phases (Fig. 1): syncytial, cellularization, differentiation and death. The syncytial phase is characterized by a series of divisions of the triploid nuclei without cytokinesis (Boisnard-Lorig et al 2001) and parallels the maximal phase of seed growth (Garcia et al 2005). After eight rounds of syncytial mitoses, the cellularization process starts, initially from regions surrounding the embryo and proceeding towards the chalazal region. Cellularization is followed by a differentiation of functional tissues, and eventually most endosperm cells die during seed

maturation. The timing of endosperm cellularization correlates with the end of the main stage of seed growing, therefore the size attained by the endosperm syncytium appear to be a major determinant of seed size (Boisnard-Lorig et al 2001). Consequently, precocious endosperm cellularization results in small seeds, while delayed endosperm cellularization causes the formation of enlarged seeds (Scott et al 1998; Garcia et al 2003; Berger et al 2006). A number of mutations have been described that impair proper endosperm development and the timing of endosperm cellularization. The existence of three redundant pathways that control endosperm cellularization has been recently proposed (Khan et al 2014). The first pathway regulates endosperm cellularization through the action of APETALA 2 (AP2) and MADS-box transcription factors such as AGAMOUS LIKE 80 (AGL80) (Portereiko et al 2006), AGL61 (Steffen et al 2008) and AGL62 (Kang et al 2008). The second endosperm cellularization pathway includes members of the Polycomb group (PcG) proteins and its imprinted genes (discussed below). The third pathway is the *IKU* pathway.

The *AP2* pathway controlling endosperm cellularization has only two members previously linked to seed size determination: *AP2* itself and *AGL62*. *AP2* encodes the founding member of the plant-specific family of transcription factors that contain an AP2/EREBP (Ethylene Responsive Element binding Protein) DNA binding domain (Jofuku et al 1994; Okamuro et al 1997; Riechmann and Meyerowitz 1998). AP2 is involved in a great variety of developmental processes, including endosperm cellularization. *ap2* mutant seeds undergo an extended endosperm proliferation stage, associated with a delay in cellularization (Ohto et al 2009). Additionally, the abnormal endosperm development in *ap2* mutants resulted in other seed defects, such as enlarged embryos that show increased cell number and cell size (Jofuku et al

2005; Ohto et al 2005). AP2 has also been associated with seed coat development and integument-endosperm crosstalk, as we discuss below. The other member belonging to this pathway is *AGL62*, a type I MADS-box transcription factor. The expression level *AGL62* correlates with endosperm cellularization in a dosage-dependent way, suggesting that it represents a key regulator of endosperm cellularization, and consequently of seed size determination. Accordingly, the *agl62* mutants have precocious endosperm cellularization and a small seed phenotype (Kang et al 2008; Kradolfer et al 2013), while increased *AGL62* expression correlates with a delay or a complete absence of cellularization (Erilova et al 2009; Tiwari et al 2010). Interestingly, *AGL62* expression is under negative control of the FIS-PRC2 complex, an indication that the timing of endosperm cellularization is epigenetically controlled (Hehenberger et al 2012).

The second pathway controlling cellularization of the endosperm involves the PcG protein complex and its imprinted genes. Imprinting and its relation with seed size control will be discussed in a separate section of this review.

The IKU pathway is probably the best-described pathway for endosperm cellularization. The genes *HAIKU1* (*IKU1*) and *IKU2* have been shown to be key regulators of seed size in *Arabidopsis* via control of the transition from syncytial phase to the cellularization phase of the endosperm (Garcia et al 2003). *IKU1* encodes a protein containing a VQ motif (Wang et al 2010), while *IKU2* encodes a leucine-rich repeat kinase (Luo et al 2005). *iku1* or *iku2* mutant plants show reduced proliferation of the endosperm, as well as a precocious cellularization process, leading to reduced seed size (Garcia et al 2003). Another member of the IKU pathway is *MINISEED3* (*MINI3*), a WRKY class transcription factor that regulates the endosperm cellularization process (Luo et al 2005). *mini3* mutant plants phenocopy *iku1* and *iku2* small seed phenotypes,

due to precocious cellularization of the endosperm. In addition, the small seed phenotype of *mini3* mutant is ascribable to reduced cell expansion in the seed coat, and reduced cell proliferation that results in a smaller embryo compared to wild type (Garcia et al 2003; Luo et al 2005). Genetic and mutant analyses indicate that *IKU1, IKU2* and *MINI3* are likely to participate in a single pathway, with *IKU1* regulating both *MINI3* and *IKU2*, and *MINI3* regulating *IKU2* (Luo et al 2005). Apparently, *MINI3* could positively regulate *IKU2* by binding to the putative W-box identified in the *IKU2* promoter. Seed size of the double mutants *iku2-1 mini3-1* is similar to the seed size of homozygous mutant alleles of each single locus (Luo et al 2005).

Recently, it has been reported that SHORT HYPOCOTYL BLUE1 (SHB1) binds to the promoters of *IKU2* and *MINI3* (Zhou et al 2009; Kang et al 2013). *SHB1* encodes a nuclear SYG1-homologous protein (Kang and Ni 2006) that is recruited by MINI3 to activate the *IKU2* and *MINI3* expression, and probably other genes required for endosperm development, stimulating the process of endosperm cellularization (Kang et al 2013). *SHB1* was first described to be involved in hypocotyl development (Kang and Ni 2006), and later as a regulator of endosperm proliferation and the timing of cellularization. The gain-of-function overexpression mutant *shb1-D* displayed an enlarged seed size phenotype associated with a delay in endosperm cellularization (Zhou et al 2009).

Thus, independent networks act as key regulators of endosperm growth, by controlling endosperm proliferation and cellularization with a major impact in final seed size (Fig. 2). Further investigation is required to identify all the molecular players in these pathways and to determine if they share downstream targets.

3.1. Genomic imprinting and parent-of-origin effects

In plants, genomic imprinting has been observed in the endosperm. Imprinting of a specific allele depends on the presence of an epigenetic mark on the corresponding locus (reviewed by Ferguson-Smith, 2011). It has been proposed that imprinted genes regulate the transfer of nutrients from the sporophyte to the developing progeny. In particular, maternally expressed genes (MEGs) function to equally allocate nutrients to all seeds, while on the other hand paternally expressed genes (PEGs) function as growth factors that allow their own offspring to extract the maximum amount of nutrients from the mother. Therefore, increased PEGs activation determines formation of larger seeds (Haig and Westoby 1989).

Epigenetic marks have been proposed as a major determinant of parents-of-origin imprinting. Of particular importance is the balance of methylation between maternal and paternal alleles in the central cell. DNA demethylation relies on the enzymatic activity of DEMETER (DME) (Kinoshita et al 2004; Gehring et al 2006) and DNA methylation depends on the enzyme MET1 (Hsieh et al 2011; Jullien et al 2012). *DME* is expressed in the central cell of the female gametophyte (Choi et al 2002) but not in sperm cells (Schoft et al 2011). This leads to specific DNA hypomethylation of the maternally inherited genome. Previous studies showed that altering DNA methylation in a parental-specific manner via MET1 resulted in variation in seed size (Xiao et al 2006). When crossing *MET1::RNAi* pistils with wild type pollen, the result is production of enlarged F1 seeds. Meanwhile, reciprocal crosses generated smaller F1 seeds, as expected from the presence of hypomethylated maternal genome (Adams et al 2000; Luo et al 2000). Thus, the DNA methylation status of the maternal genome in the endosperm directly influences seed size.

The second major mechanism involved in imprinted expression of a subset of genes relies on PcG proteins. PcG proteins are pivotal regulators of cell identity that act as transcriptional repressors in multimeric complexes (Schuettengruber and Cavalli 2009). Among these, the PRC2-complex catalyzes the trimethylation of histone H3 on lysine 27 (H3K27me3), and has been implicated in controlling endosperm development. Specifically, the FIS-PRC2 complex, which comprises the different subunits encoded by *MEDEA* (*MEA*), *FERTILIZATION INDEPENDENT SEED2* (*FIS2*), *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE*), and *MULTICOPY SUPPRESSOR OF IRA1* (*MSI1*) acts in the central cell of the female gametophyte and in the endosperm, targeting DNA hypomethylation sites. For this reason, the FIS-PRC2 complex mainly represses the expression of maternally inherited (and hypomethylated) alleles. Seeds with mutations in *mea*, *fis2* or *fie2* show endosperm proliferation even in the absence of fertilization, but also prolonged endospermal proliferation and late cellularization if fertilization occurs (Grossniklaus et al 1998; Kiyosue et al 1999; Makarevich et al 2008). The phenotypes of these mutants imply that PCR2 complexes promote fast endosperm differentiation after fertilization, thus directly acting on a pathway that greatly influences seed size (Fig. 2).

Finally, it is necessary to mention that perturbation of the relative dosages of the maternal and paternal genomes, typical in the case of interploidy crosses, directly affects endosperm development and seed size (Garcia et al 2003; Luo et al 2005; Kang et al 2008; Zhou et al 2009; Wang et al 2010). The defects and low endosperm viability often observed in seeds of interploidy crosses (as in the case of wheat) can be explained in terms of maternal or paternal genome excess, *i.e.* an imbalance between MEGs and PEGs, and its effect on endosperm growth (Haig and Westoby 1991). However, the negative effects on seed

development of interploidy crosses are reduced in *Arabidopsis*, in which both paternalized (PEGs excess) and maternalized (MEGs excess) seeds show the expected alteration from wild type size, but show normal endosperm viability. This mitigated effect is probably due to the high rate of self-pollination that is characteristic of this model species (Scott et al 1998).

4. The role of the seed coat in seed size determination

The *Arabidopsis* seed coat derives from the ovule integuments, formed by a set of five cell layers in mature ovules (Fig. 1). Two cell layers derive from the outer integument (oi), and three from the inner one (ii). The outer integument consists of two cell layers (oi1 and oi2), and the inner integument consists of three cell layers (ii1, ii1', and ii2) (Beeckman et al 2000; Kunieda et al 2008). The innermost layer of the inner integument, ii1, named the endothelium (Beeckman et al 2000), is in direct contact with the endosperm cells.

The seed coat deeply influences seed size, highlighting a fundamental role for seed maternal tissues in the control of this aspect of seed yield. The seed cavity (the space enclosed by the seed coat) increases in volume after fertilization, partly due to the independent developmental plan of the seed coat and partly as the result of the interplay between the seed coat and the endosperm (Ingouff et al 2006; Roszak and Köhler 2011). Fertilization is in fact followed by intense mitotic activity in the different layers of the seed coat, prior of a sharp decline of the same. Before fertilization, the female gametophyte (embryo sac) seems to have only a moderate importance in generating the signals to stimulate the integuments' proliferation (Ingouff et al 2006); this was proven by demonstrating that mutants defective in embryo sac formation, such as *sporocyteless* (*spl*), are still able to develop integument to some extent (Yang et

al 1999). Numerous studies have identified genes involved in *Arabidopsis* ovule integuments and seed coat development, and some of them have provided a functional characterization of seed size contribution. In particular, seed size mutant phenotypes showing a clear maternal inheritance are mainly due to an alteration of cell proliferation or elongation in the seed coat. The control of these two pathways will be discussed separately.

4.1. Factors controlling integuments cell proliferation

A key player in the control of cell cycle and expansion in *Arabidopsis* is *AUXIN RESPONSE FACTOR 2* (*ARF2*), which encodes a B3-type transcription factor of the *ARF* family (Li et al 2004). *ARF* genes take part in auxin-related responses and recognize specific AuxRE (Auxin Response Elements) consensus elements on target genes (Ulmasov et al 1999). Among the different ARF proteins, ARF2 is thought to act as a transcriptional repressor, exercising a negative control over cell proliferation and expansion (Li et al 2004; Okushima et al 2005; Schruff et al 2006). In particular different *arf2* loss-of-function mutants exhibit abnormal flower morphology and enlarged seeds in comparison to the wild type (Okushima et al 2005), a phenotype characterized in detail in the case of *arf2-9*, which presented more cells in the seed coat compared to respect to wild type seeds. The result of the increased volume of the seed cavity in *arf2-9* is that seeds are 46% heavier than the wild type seeds, showing in some cases additional cell layers in the seed coat (Schruff et al 2006). A further confirmation that *ARF2* is important for the maternal control of seed size comes from the maternal inheritance of *arf2-9* phenotype observed in the reciprocal crosses with wild type plants (Schruff et al 2006). Although the *arf2-9* mutant shows a significant reduction in fertility due to improper flower development

(Schruff et al 2006), the hypothesis that the large-seed phenotype might be an indirect effect (as a result seed size/seed number trade-off) was later refuted in another study (Hughes et al 2008). In fact the defects in the floral morphology of the *arf2-9* mutant were overcome by expressing *ARF2* under the promoter of *APETALA1* (*AP1*). The *pAP1::ARF2 arf2-9* plant improved the fertility, retaining the enlarged seed size phenotype of the original *arf2-9* mutant, thus showing the pivotal role of ARF2 in seed development.

Another negative regulator of cell division is the transcription factor AP2, whose role in endosperm development has been described above. Interestingly, the increased cell proliferation observed in *ap2* is under maternal control, and affects both the seed coat and the endosperm (Jofuku et al 2005; Ohto et al 2005). Notably, *AP2* expression is negatively regulated by miR172 during flower development (Chen 2004), while *ARF2* is negatively regulated by trans-acting small-interfering RNA (tasiRNA) (Williams et al 2005). Similarly, it was reported that mutation in the gene *miR159* results in seeds smaller than wild type (Allen et al 2007). The two known targets of *miR159* that are expressed in developing seeds, *MYB33* and *MYB65*, have no described function in the seed. However, they are responsible for the *mir159ab* seed phenotype, as the quadruple mutant *mir159ab myb33 myb65* showed a reversion of the seed traits (Allen et al 2007). Taken together, these results provide evidence of a fundamental role for post-transcriptional regulation via small RNAs in the control of seed size.

Cytochrome P450 KLUH, encoded in *Arabidopsis* by *CYP78A5/KLU,* is a regulator of organ size (both leaves and floral organs) as well as of plastochron length (Anastasiou et al 2007; Wang et al 2008). It has also been shown that *KLU*, expressed prior to fertilization in the inner integuments of the ovule, acts as a maternal positive regulator of seed size. *klu-2* seeds have a reduced number

of cells in the outer layers of the seed coat in comparison to wild type, with the result that *klu-2* seeds are 13% lighter than seeds of wild type plants. The opposite phenotype was observed in *KLU*-overexpressing plants, whose seeds are 11% heavier (Adamski et al 2009). *KLU* seems to act independently of previously described integument cell proliferation factors as *AP2* and *ARF2*, because seeds of the double mutants *klu arf2* and *klu ap2* were an intermediate seed size between those of the respective single mutants (Adamski et al 2009).

In *Arabidopsis* the importance of ubiquitin pathway in the determination of seed size has been widely investigated over the last decade. Several members involved in this pathway have been identified (reviewed by Li and Li 2014) for their role in maternal control of seed size. Among them, *DA1* and *DA1-RELATED* (*DAR*) encodes for plant-specific ubiquitin receptor protein. While single mutants *da1-ko* and *dar1-1* do not exhibit variation in seed size in comparison to wild type, the double mutant *da1-ko dar1-1* produces larger seeds. Another mutation in the *DA1* sequence (a single arginine-to-lysine aminoacidic change at position 358, the d*a1-1* mutant) results in plants producing seeds with increased cell proliferation in the seed coat, a phenotype also observed in 35S::DA1^{R358K}. This suggests that the mutated DA1 protein might act antagonistically with native DA1 or DAR (Li et al 2008). *DA2* and *ENHANCER OF DA1* (*EOD1*) encode proteins with E3 ubiquitin ligase activity, and are also negative regulators of seed size, as shown by the enlarged seeds of single mutants *da2-1* and *eod1*. They may act synergistically with *DA1*, as observed by the enhanced seed size of *da1-1 da2-1* and *da1-1 eod1* double mutants in comparison to *da1* mutant (Xia et al 2013). *EOD3* encodes cytochrome P450 CYP78A6. The gain-of-function mutant *eod3-1D* proved to be a dominant enhancer of the *da1-1* seed size phenotype, while on the contrary *eod3-ko* produced smaller seeds than wild type (Fang et al 2012). *CYP78A9*

encodes for another cytochrome P450 and is the most closely related gene to *EOD3,* with whom it might act synergistically in promoting the size of the seed coat. This is implied by the additive small seed phenotype observed in *eod3-ko cyp78a9-ko* double mutants in comparison to the single mutants (Fang et al 2012). *UBIQUITIN-SPECIFIC PROTEASE 15* (*UBP15*)/ *SUPPRESSOR OF DA2* (*SOD2*) encodes for a de-ubiquitinating enzyme acting downstream of *DA1* (Li et al 2008; Du et al 2014). The *ubp15* mutant produces small seeds, while the overexpression line of *UBP15* results in larger seeds. This is likely due to a positive effect on cell proliferation in maternal integuments of ovules and developing seeds.

It has been suggested that *da1-1* acts independently of *ARF2* and *AP2*, as the seed phenotype of the double mutants *da1-1 ap2* and *da1-1 arf2* is additive in comparison to the one of the single mutants (Li et al 2008).

4.2. Factors controlling integuments cell elongation

A reduction in cell elongation is observed in the loss-of-function mutant *TRANSPARENT TESTA GLABRA2* (*TTG2*). In the *ttg2* mutant, cell elongation in the integuments is affected, possibly because of the increased physical constraint of the cell walls, or possibly because of disruption of the developmental pathways for elongation. Endosperm development is also affected, probably as a consequence of the defects in integument cells (Garcia et al 2003; Garcia et al 2005). Developing seeds produced by the double mutant *ttg2 iku2* display extremely reduced size in comparison with the single mutants *ttg2* and *iku2* seeds (Garcia et al 2005). The combination of *ttg2* and *iku2* mutations prevents integument cell elongation and growth of the endosperm more severely than in each single mutant. The double homozygous mutant

displays a cumulative phenotype combining the maternal effects of *ttg2* with the zygotic effect of *iku2* (Garcia et al 2003; Garcia et al 2005). The additive reduction of integument cell division and elongation, endosperm growth and seed size when *iku2* and *ttg2* mutations are combined, indicates that each mutation acts in distinct genetic pathways, but has common effectors. In parallel, reduction of the endosperm volume is more evident in the double mutant relative to the single mutants. To achieve the size of the integument, dictated by the size of the syncytial endosperm, integument cells regulate elongation, not cell proliferation. Integument cell elongation plays a key role in the coordination of size between the endosperm and the integument. Accordingly, *TTG2* would modulate the competence of the integument cells to elongate via a maternal integument elongation-dependent pathway (Garcia et al 2005).

Another positive regulator of seed size in *Arabidopsis* is the R2R3 MYB transcription factor, MYB56, which maternally affects seed development by regulating seed size and shape (Zhang et al 2013). The loss-of-function mutant lines of *MYB56* generate smaller seeds, while overexpression of *MYB56* generates larger seeds compared with wild type. *myb56* endothelial cells are smaller and more rounded. Apparently, the role of MYB56 is locally dependent since its altered expression on the endothelial layer affects cell size but not cell number; however, in the two layers of the outer integument *MYB56* controls only cell number but not the cell size (Zhang et al 2013). *MYB56* affects seed size in a regulatory pathway probably independent of other seed coat development regulators such as *TTG2*, *KLU*, *GORDITA* (*GOA*) and *DA1*, because these genes show no expression changes in a *myb56* mutant background (Zhang et al 2013).

SEEDSTICK (*STK*) and *ARABIDOPSIS B SISTER* (*ABS*) are two MADS-box genes that act together to control the formation of one layer of the seed coat, the endothelium, during seed development (Mizzotti et al 2011). *STK* controls ovule identity redundantly with *SHATTERPROOF1* (*SHP1*) and *SHP2*. In addition, *stk* single mutant produces smaller seeds (Pinyopich et al 2003) with respect to wild type, whereas *abs* mutant has no size difference (Nesi et al 2002). The double mutant *stk abs* completely lacks endothelium development and manifests a high level of sterility, due to both ovule and seed abortions (Mizzotti et al 2011). Another MADS-box transcription factor involved in seed coat development is *GOA*. A loss-of-function mutation in *GOA* causes an increase in the seed size when compared with wild type, due to an impact on cell expansion processes, during fruit and seed development (Prasad et al 2010; Erdmann et al 2010).

Very recently, a new actor in the integument development was described, the plasma membrane receptor kinase FERONIA (FER) (Yu et al 2014). FER has been demonstrated previously to be involved in inhibiting pollen tube elongation (Escobar-Restrepo et al 2007) and promoting cell elongation in leaves and root hairs (Guo et al 2009; Duan et al 2010). *FER* is highly expressed on the integuments of developing seeds but it was not detected in embryo or endosperm (Yu et al 2014). *FER*-null mutants develop seed that are 40% - 60% larger than the wild type. At 2DAP the outer integument of *fer-4* contained larger cells and no differences in cell number from the wild type. The authors concluded that *FER* inhibits the elongation of seed coat cells (Yu et al 2014). This conclusion is supported by the fact that FER controls cell elongation in root hairs in response to auxin through recruitment of RHO GTPases (ROP/RAC) to promote or inhibit cell elongation. ROP/RAC signaling pathway regulates several cell responses, such as polarized growth

and differentiation (Duan et al 2010; Yu et al 2014). In the female gametophyte, FER is a receptor of Rapid Alkalinization Factor (RALF), a small peptide whose overexpression or external application promotes cell wall alkalinization and growth inhibition. The FER-RALF interaction causes the phosphorylation of the H^+-ATPase AHA2. AHA2 phosphorylation may have an effect on the cell wall levels of reactive oxygen species (ROS), changing the balance between the ROS promoting/inhibiting cell wall relaxation state (reviewed in Wolf and Höfte, 2014). In this way *FER* could, at least partially, control the cell wall's capacity to elongate. However, further research has to be done to fully understand the role of *FER* in seed development.

5. Endosperm-integument crosstalk

Endosperm and integument growth and development are tightly coupled. As mentioned above, seed coat development influences endosperm proliferation and the timing of cellularization (Fig. 1). At the same time, the endosperm performs a key nourishing function and provides signals to coordinate seed maturation (Berger et al 2006).

Two models have been proposed to explain the crosstalk between endosperm and the seed coat and its role in controlling seed size. The 'Integument Size-Restriction Model' suggests that the expansion of the integument cells represents a physical constraint to the size of the seed cavity, restricting the size of the embryo. As a result, this volume reduction increases the concentration of the factors triggering the cellularization process (Garcia et al 2005; Doughty et al 2014).

In the second model, identified as the 'Cellularization Signaling Model' (Fig. 3), the interplay between seed coat and endosperm is mediated by a signal that

moves between integuments and endosperm. Flavonoids (proanthocyanidins [PAs]) represent excellent candidates for the signal that triggers the endosperm cellularization process since they are synthesized in the endothelium. The accumulation of flavonoids is initiated after fertilization in the endothelium (Debeaujon et al 2003). The relevance of flavonoids in seed size control emerged from the fact that many flavonoid biosynthetic pathway mutants show alterations in the timing of the endosperm cellularization process (Scott et al 2013; Doughty et al 2014). Furthermore, it has been reported that flavonols could interact with the phosphoglycoprotein (PGP) auxin transporters PGP1, PGP4 and PGP19 (Peer and Murphy 2007). Flavonoids inhibit PGP-mediated polar auxin transport (Terasaka et al 2005), which in fact may cause a rapid change in auxin concentration that results in delay/triggering of the endosperm cellularization process (Doughty et al 2014), thus affecting seed development and seed size.

Another type of candidate molecules that could mediate the seed coat-endosperm crosstalk are the polysaccharides. Nutrients from the phloem have to be unloaded from the seed coat into the endosperm and the embryo. The processing of sucrose follows distinct biochemical pathways, such as biosynthesis of cell wall polysaccharides and storage reserves. Thus, maternal tissues are major sites of sugar translocation and partitioning, and are hence considered key determinants of sink strength and seed biomass yield. Since sugar metabolism and transport can be highly compartmentalized in seeds (Morley-Smith et al 2008), even small differences in hexose/sucrose ratio can have dramatic effects on seed development and storage metabolism. For instance, *AP2* seems to modulate the nutritional supply from maternal tissues by changing the ratio of hexose to sucrose during seed development, opening the

possibility that *AP2* may also control seed mass through its effects on sugar metabolism (Ohto et al 2009).

6. The role of hormone synthesis and perception in determining seed size

As stated before, the complex structure forming the developing seed requires the coordination in growth of multiple tissues and cells with different patterns of proliferation and differentiation. This coordinated growth demands a precise spatio-temporal organization that can be achieved thanks to the synthesis and perception of signals in different seed tissues. This sophisticated communicative system between seed compartments is crucial not only to regulate their balance in growth, but also to control the progression of the whole developmental process within each tissue. The function of hormones in this communicative role to coordinate seed development has been well characterized by studies performed on hormone deficient and insensitive mutants of *Arabidopsis*. Several hormonal pathways such as brassinosteroids, cytokinins, auxins and abscisic acid have been already proposed to play a crucial role in seed development (Sun et al 2010). In this last part of the review, we provide a global panorama of the regulation of seed development by phytohormonal stimuli, emphasizing their impact on seed size (for a review of hormones controlling seed development see Locascio et al 2014).

6.1. Key role of brassinosteroids in seed size regulation

The function of brassinosteroids (BR) in seed development has been well characterized by studies of BR deficient and insensitive mutants in several species such as *Arabidopsis*, *Oryza sativa*, *Pisum sativum*, and *Vicia faba* (for a

review, see Jiang and Lin 2013). At the cellular level, low endogenous concentrations of BR have been shown to exert a positive effect on cell elongation, meanwhile saturating levels of BR lead to the opposite effects with reduced cell elongation (Fujioka et al 1997; Turk et al 2003). Brassinosteroids are required for proper plant growth and deficiencies in their synthesis and signal transduction pathway leads to severe dwarfed phenotypes (Fujioka et al 1997). An *Arabidopsis* dwarf mutant overexpressing the P450 monooxygenase gene *CYP72C1* (*shk1-D*) showed a reduction of endogenous BR levels and produced smaller seeds than the wild type, probably due to an effect on cell elongation (Takahashi et al 2005). A similar small seed phenotype was reported in the *DWARF5* (*DWF5*) loss-of-function mutant. *DWF5* encodes a sterol reductase gene involved in the BR biosynthesis pathway (Choe et al 2000). The weak BR deficient mutant *de-etiolated2* (*det-2*), in which seed size was rescued by exogenous BR application, and the BR insensitive mutant (*BRASSINOSTEROID INSENSITIVE1*) *bri1-5* produced smaller seeds than wild type seeds.

The mechanism of BR regulation of seed size is twofold: 1) expanding the seed cavity and endosperm volume, promoting embryo development and 2) controlling integument cell length (Jiang et al 2013). BR regulates embryo and endosperm development through the BRASSINAZOLE-RESISTANT1 (BZR1) transcription factor which controls the IKU-pathway by binding to the promoter regions of *SHB1* or *IKU1,* or alternatively through binding to the promoter of *IKU2* (Jiang et al 2013).

On the other hand, evidence supporting BR control of seed size by regulating integument development comes from the significant decrease of integument cell length in *det2* (Jiang et al 2013) and from the mutant *arf2*, which develops larger seeds due to extra integument cell divisions (Schruff et al 2006). *ARF2* is

a direct target of BZR1, and its transcription is negatively regulated by BR (Jiang et al 2013). Thus it seems that BR might regulate seed size through BZR1 binding and repressing *ARF2* promoter to positively regulate the integument development (Jiang et al 2013). As a result, *ARF2* has been proposed to mediate the crosstalk between auxins and BR. BIN2, a kinase regulated by BR, phosphorylates ARF2 *in vitro*. Apparently, this phosphorylation would allow the detachment of ARF2 from DNA, inhibiting its transcriptional repression activity (Vert et al 2008). The proposed scenario establishes that BR affects BIN2 target specificity promoting a change from BRZ1/BES1 to ARF2. The presence of auxin and/or BR will determine an increment or persistence of the target genes expression (Krizek 2009). Interestingly, the *fer* mutants are hypersensitive to BR (24-epibrassinolide), suggesting that *FER* can act as a critical modulator of the brassinosteroid-signaling pathway during hypocotyl development (Deslauriers and Larsen 2010). The clarification of the relation between *FER*, BR and seed size promises to be very interesting to be clarified. Last but not least, BR can act as global regulator, acting at the same time over both integuments, endosperm, and embryo development through BZR1 binding to the *AP2* promoter (Jiang et al 2013).

6.2. The role of auxins in communication

At the cellular level, auxin is involved in many processes, including pattern formation, cell division and cell expansion (Vandenbussche and Van Der Straeten 2004; Leyser 2005). In addition, auxins exert a key role during the first steps of seed development (Hamann et al 2002; Friml et al 2003; Jenik and Barton 2005; Cheng et al 2007; Wabnik et al 2013). Schruff and colleagues

proposed that *ARF2* is a general repressor of cell division in many aerial organs of the plant by controlling expression of *CYCD3;1*, a D-type cyclin involved in cell cycle entry, and *AINTEGUMENTA* (*ANT*), a transcription factor involved in organ growth and cell division control (Klucher et al 1996; Schruff et al 2006).

6.3. Cytokinins

Several studies have highlighted the importance of cytokinins (CK), together with auxin, in promoting growth by cell division, development and differentiation (Bishopp et al 2011; Vanstraelen and Benková 2012). High levels of CK are present during early seed development in many species (Yang et al 2002). In *Arabidopsis*, limited information comes from a few reports (Werner et al 2003; Garcia et al 2005; Day et al 2008) and CK function has not yet been exhaustively characterized. Studies performed on the genetics of CK production have shown that during early stages of seed development transcriptional changes are mostly associated with effects of the hormone on the development of endosperm and seed coat. These data reinforce the idea that the control of seed size would involve a crosstalk occurring between maternal and zygotic tissues (Garcia et al 2005). Transcriptome analysis of the endosperm at 4 DAP revealed an over-representation of CK biosynthetic and response genes, supporting the hypothesis that the predominant role of CK is in cell proliferation of the early endosperm (Lur and Setter 1993; Day et al 2008). Overexpression of two cytokinin oxidase dehydrogenases (*CKX1* & *CKX3*) produced larger seeds with larger embryos. The enlargement found in these transgenic seeds is attributable to increases in cell number and size (Werner et al 2003). Larger seeds were also produced by the triple mutant of the CK

receptor genes *ARABIDOPSIS HISTIDINE KINASE 2* (AHK2), AHK3 and CYTOKININ RESPONSE1/AHK4 (CRE1/AHK4). In this case, an increase of almost two times the seed size was reported, when compared with wild type seeds, due to an enlargement of the embryo size, with approximately 15% greater cell number and 30% greater cell size. Reciprocal crosses with wild type plants suggested that the increase found in seed size was likely to be regulated by maternal and/or endospermal genotypes (Riefler et al 2006).
Recently, it was concluded that the control of endosperm size by the IKU pathway is regulated by the cytokinin catabolic pathway through the activation of *CKX2* (*CYTOKININ OXIDASE 2*) by MINI3 (Li et al 2013). CKX2 is also co-regulated by maternal genome dosage and methylation, and both phenomena suppress *CXK2* transcription. These data establish a link between hormonal and epigenetic factors in the regulation of seed size in *Arabidopsis* (Li et al 2013).

5.4. Abscisic acid

The predominant role of abscisic acid (ABA) regulation involves key processes occurring during the maturation stages of seed development. Key aspects of this development are accumulation of storage compounds in the embryo, seed dormancy, and the inhibition of precocious germination (McCarty 1995; Finkelstein et al 2002; Kanno et al 2010). ABA biosynthesis exhibits two peaks during seed development: initially biosynthesis is induced in the embryo, then levels accumulate to a second peak during the late maturation stage, where it is thought that ABA mainly originates from the maternal tissues (Finkelstein et al 2002; Finkelstein 2004). ABA has been proposed to act mainly as an endosperm development regulator since the mutants *abscisic acid deficient2* (*aba2*) and *abscisic acid-insensitive5* (*abi5*) develop larger seeds than the wild

type (Cheng et al 2014). *ABA2* encodes a dehydrogenase/reductase involved in ABA biosynthesis (González-Guzmán et al 2002) and *ABI5* encodes a transcription factor involved in ABA signaling (Brocard et al 2002). Interestingly, *aba2* mutants have delayed endosperm cellularization. The model of action suggests that endogenous ABA levels in the seed are raised by ABA2 action, resulting in an enhancement of *ABI5* transcription. ABI5 negatively regulates *SHB1* expression by directly binding to its promoter region. Therefore ABA regulates proper endosperm development and cellularization processes in a SHB1-dependent way (Cheng et al 2014). ABA slowly induced *DA1* expression but other growth regulators such as jasmonic acid, auxin, CK, BR, gibberellins or glucose, failed to induce its expression. It therefore seems that the mechanism that restricts proliferative growth under the control of *DA1* control could include ABA signaling (Li et al 2008).

7. Future perspectives

Unraveling seed development and its genetic control is important due to the critical role of seeds as a food source for mankind and livestock, as well as the growing interest in seeds as a renewable source of energy. Recently, genomic-based research and other modern technologies have made it possible to identify most of the genes involved in seed development, providing a vast amount of information that could be used in the engineering and design of transgenic crops. However, there are many gaps in the field regarding the functional characterization and determination of the biological relevance of these genes in model species. Unveiling a complete and accurate map of the process remains a major challenge for plant biologist. Achieving these goal will require not only the integration of multiple disciplines including proteomics, metabolomics and

functional genomics, but also the development and improvement of automatized computational tools to analyze complex datasets. A comprehensive analysis of large-scale datasets will provide the required tools to enhance the nutritional quality of seeds, and also to increase resistance to adverse environmental conditions and/or biological attacks. A second major challenge for plant genomics will be finding an integrative and rational way to apply that information to crop species to improve their agronomic performance. This could be achieved either by using the basic knowledge arising from studies of *Arabidopsis*, or using the tools and techniques refined with *Arabidopsis* (or other model species), to generate and analyze extensive datasets for important crop species.

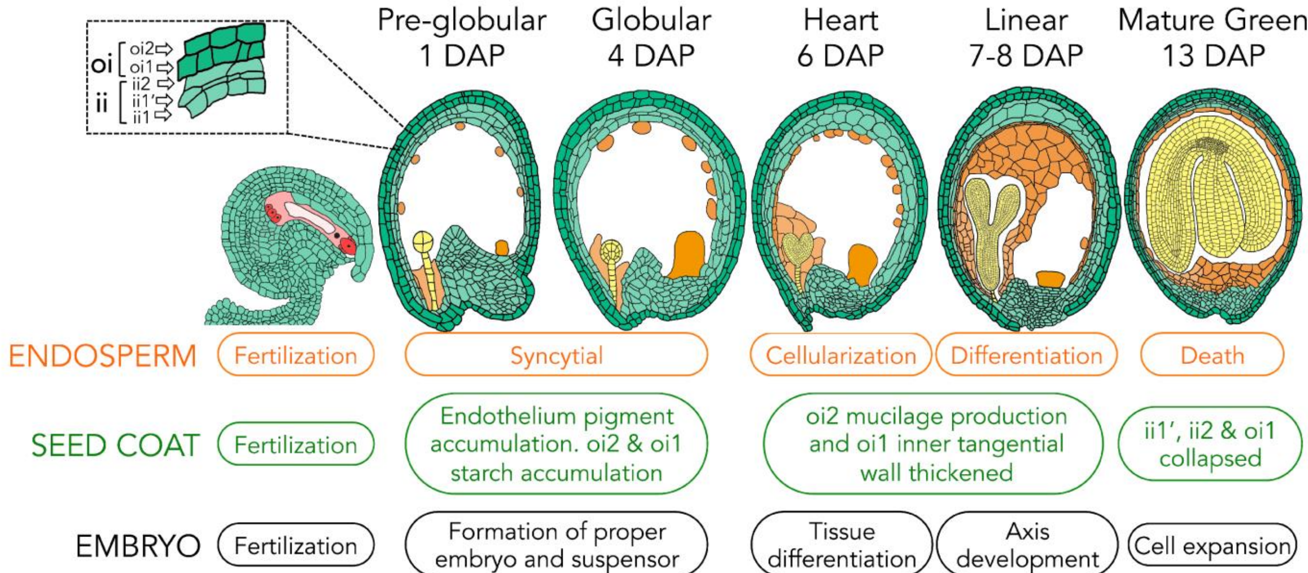


Fig. 1 Schematic representation of seed development in *Arabidopsis*. Diagrams of an unfertilized ovule and five stages of seed development from the pre-globular (1 Day After Pollination - DAP) to mature green (13 DAP). Embryo developmental stages were adapted from Le et al 2010. Detail of the five seed coat layers according to Appelhagen et al 2014. Seed coat developmental events are specified by Beeckman et al 2000. Endosperm development is presented according to Lafon-Placette and Köhler, 2014. Drawings are not to scale. Abbreviations and color-code: *oi* outer integument, *ii* inner integument, *DAP* days after pollination. *Green* seed coat, *Orange* endosperm, *Yellow* embryo.

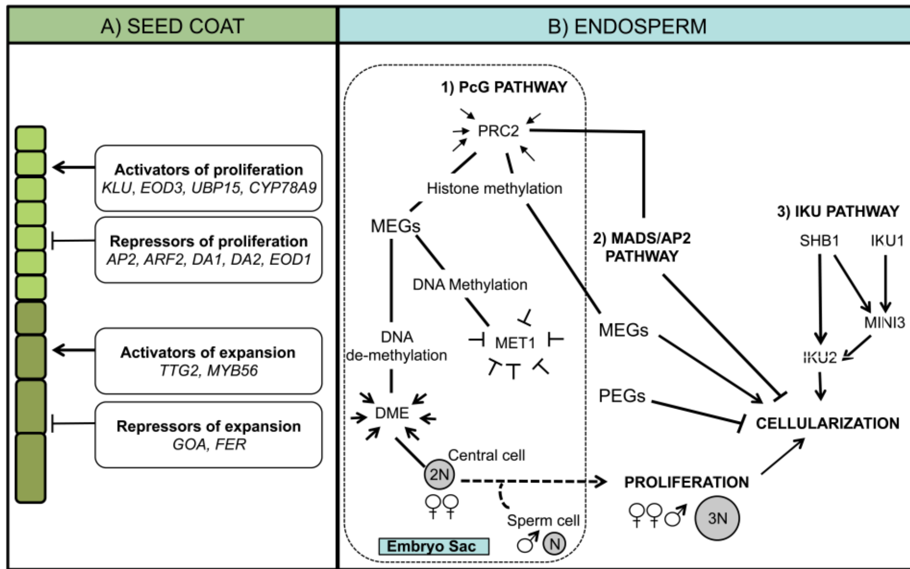


Fig. 2 Model indicating the pathways determining seed size in *Arabidopsis*. The presented model illustrates the main networks and/or key regulators characterized in the literature, based on their role in development of the A) seed coat or of B) the endosperm which are discussed in this review. Part C) shows the proposed crosstalk mechanism involving the seed coat and the endosperm, illustrated in detail in Fig. 3. A) Seed coat. Genetic pathways involving the activation/repression of cell proliferation and cell expansion affecting seed coat development, thus controlling seed size in a maternal way. The distinct four functional categories (pink boxes) involved are refer to characterization studies provided by the respective publications. B) Endosperm. Schematic representation of factors influencing endosperm cellularization and thus, seed size. One of the mechanisms involved in parents-of-origin effects include activation of DME in the central cell, and simultaneous repression of MET1, that results in a misbalanced ratio in the endosperm that favors the expression of MEGs over the PEGs. White boxes indicate the three pathways that regulate the timing of endosperm cellularization, discussed in detail in this work.

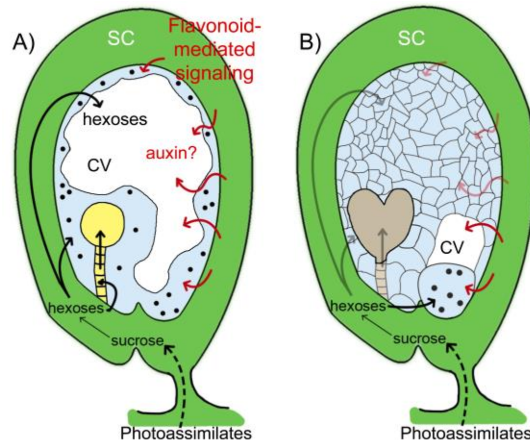


Fig. 3. Schematic representation of endosperm-seed coat crosstalk in *Arabidopsis* according to the 'cellularization signaling model'. Seed coat layers are not shown for clarity. Black circles represent the endosperm nuclei at syncytial stage. A) Seeds on globular stage embryo (4 DAP), the endosperm progresses from the syncytial to cellularized stage. In this suggested model, transport of a cellularization signal between the integuments and the endosperm would be controlled by flavonoid biosynthesis. Adding support to this thesis, several mutants defective in the flavonoid biosynthesis pathway with reduced seed size were found to display a precocious endosperm cellularization (Scott et al 2013). Hexose concentrations may also play an important regulatory role driving growth of the endosperm, since a higher hexose/sucrose ratio may stimulate mitotic activity and promote cellular proliferation leading to a greater seed size (Ohto et al 2005). During the early stages of seed development sucrose is actively transported into plant “sink” tissues like seeds and enters the seed coat via the vascular bundle of the funiculus (black arrows). Sucrose is cleaved in the seed coat and the resultant hexoses are used by developing embryo and endosperm. Signaling mechanisms, originated in the seed coat, may enter to the syncytium from the seed coat and later reach the embryo (red

arrows). This could be done directly from the syncytial endosperm, or indirectly via the suspensor. B) The accumulation of these signals triggers the endosperm cellularization process at later stages of seed development (heart stage embryo). Abbreviations: *SC* seed coat.

Gene	Locus	Mutants characterized	Effect on seed size	Expression Patterns		Functional Category	Reference(s)
				Ovule (prefertilization)	Seed		
ARF2	AT5G62000	arf2-6; arf2-7; arf2-8	+	Y	Embryo	Transcription Factor	Okushima et al. 2005; Schruff et al. 2006
IKU1	AT2G35230	iku1	-	Y	Endosperm	VQ motif protein	Garcia et al. 2003; Wang et al. 2010
KLU	AT1G13710	klu-2; klu-4	-	Y		Cytochrome P450 oxygenase	Adamski et al. 2009
TTG2	AT2G37260	ttg2-1	-	Y	Seed coat,endosperm	Transcription Factor	Johnson et al. 2002; Dilkes et al. 2008
STK	AT4G09960	stk-1; stk-2	-	Y		Transcription Factor	Pinyopich et al. 2003; Mizzotti et al. 2012
FER	AT3G51550	fer-4	+		Seed coat	Receptor kinase	Yu et al. 2014
AHK2	AT5G35750			Y	Developing siliques, embryo	CK receptor	Higuchi et al. 2004; Riefler et al. 2006
AHK3	AT1G27320	ahk2-5 ahk3-7 cre1-2*	+		Developing siliques, embryo	CK receptor	Nishimura et al. 2004
CRE1/AHK4	AT2G01830			Y	Developing siliques, embryo	CK receptor	Riefler et al. 2006; Nishimura et al. 2004
FIS2	AT2G35670	fis2	POE	Y	Endosperm	Chromatin remodeller	Luo et al. 2000
DA1	AT1G19270	da1-1	+	Y	Embryo	Ubiquitin receptor	Li et al. 2008
IKU2	AT3G19700	iku2-3	-	Y	Endosperm	Receptor kinase	Garcia et al. 2003; Luo et al. 2005
MINI3	AT1G55600	mini3-1	-		Endosperm, embryo	Transcription Factor	Luo et al. 2005; Li et al. 2013
CKX1	AT2G41510	35S:AtCKX1****	+	Y	ND	CK biosynthesis	Werner et al. 2003
CKX5	AT1G75450	ckx3 ckx5*	+	Y	ND	CK biosynthesis	Bartrina et al. 2011
CKX3	AT5G56970	35S:AtCKX3****	+	ND	ND	CK biosynthesis	Werner et al. 2005
MYB56	AT5G17800	myb56-1	-		Developing seeds	Transcription Factor	Zhang et al. 2013a
SHB1	AT4G25350	shb1-D**	+		Endosperm, embryo	SYG1 homologous protein	Zhou et al. 2009
AP2	AT4G36920	ap2-11	+		Seed coat, endosperm and embryo	Transcription Factor	Ohto et al. 2005; Kinoshita et al. 2004
EOD3	AT2G46660	eod3-ko1	-	ND	ND	Ubiquitin ligase	Fang et al. 2012
MET1	AT5G49160	met1-6	POE		Endosperm, embryo	DNA methylation	Xiao et al. 2006
CYP78A9	AT3G61880	cyp78a9-ko	-		Developing seeds	Cytochrome P450 oxygenase	Fang et al. 2012
EOD1	AT3G63530	eod1-2	+	Y	Embryo	Ubiquitin ligase	Li et al. 2008; Xia et al. 2013
UBP15/SOD2	AT1G17110	ubp15-1	-	Y	NS	Ubiquitin protease	Du et al. 2014
GOA	AT1G31140	goa-1	+	Y	Developing siliques, seed coat, embryo	Transcription Factor	Prasad et al. 2010; Erdmann et al. 2010
CYP72C1	AT1G17060	shk1-D***	-		Developing siliques	Cytochrome P450 oxygenase	Takahashi et al. 2005
DET2	AT2G38050	det2	-	Y		BR biosynthesis	Fujioka et al. 1997; Jiang et al. 2013
BRI1	AT4G39400	bri1-5	-		Developing seeds	BR signaling	Jiang et al. 2013; Hategan et al. 2014
ABA2	AT1G52340	aba2-1	+		Developing siliques	ABA biosynthesis	Cheng et al. 2014
ABI5	AT2G36270	abi5-2	+		Developing siliques, dry seeds	Transcription Factor	Finkelstein and Lynch 2000; Cheng et al. 2014
MEA	AT1G02580	f644/mea	POE	Y	Developing siliques, embryo	Chromatin remodeller	Grossniklaus et al. 1998; Kiyosue et al. 1999
DA2	AT1G78420	da2-1	+	Y	NS	Ubiquitin ligase	Xia et al. 2013
AGL62	AT5G60440	agl62-2	-		Developing seeds	Transcription Factor	Hehenberger at al. 2012

Y - Yes
NS - Not Specified
ND - Not Detected
* only double or triple mutant presented seed size phenotype
** gain-of-function
***Activation line of *CYP72C1*
****overexpression line
POE – Parent-of-origin effects

Table 1 List of *Arabidopsis* genes regulating seed size. Gene product (name), locus (AtG), mutant allele, functional group (TF, hormones, etc.), expression pattern (endosperm, integuments/seed coat, embryo) and reference.

8. References

Adams S, Vinkenoog R, Spielman M, et al (2000) Parent-of-origin effects on seed development in Arabidopsis thaliana require DNA methylation. Development 127:2493–502.

Adamski N, Anastasiou E, Eriksson S, et al (2009) Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. Proc Natl Acad Sci U S A 106:20115–20. doi: 10.1073/pnas.0907024106

Allen RS, Li J, Stahle MI, et al (2007) Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. Proc Natl Acad Sci U S A 104:16371–6. doi: 10.1073/pnas.0707653104

Anastasiou E, Kenz S, Gerstung M, et al (2007) Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. Dev Cell 13:843–56. doi: 10.1016/j.devcel.2007.10.001

Appelhagen I, Thiedig K, Nordholt N, et al (2014) Update on transparent testa mutants from Arabidopsis thaliana: characterisation of new alleles from an isogenic collection. Planta 240:955–70. doi: 10.1007/s00425-014-2088-0

Bartrina I, Otto E, Strnad M, et al (2011) Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in Arabidopsis thaliana. Plant Cell 23:69–80. doi: 10.1105/tpc.110.079079

Becker MG, Hsu S-W, Harada JJ, Belmonte MF (2014) Genomic dissection of the seed. Front Plant Sci 5:464. doi: 10.3389/fpls.2014.00464

Beeckman T, De Rycke R, Viane R, Inzé D (2000) Histological Study of Seed Coat Development in Arabidopsis thaliana. J Plant Res 113:139–148. doi: 10.1007/PL00013924

Bencivenga S, Colombo L, Masiero S (2011) Cross talk between the sporophyte and the megagametophyte during ovule development. Sex Plant Reprod 24:113–21. doi: 10.1007/s00497-011-0162-3

Berger F, Grini PE, Schnittger A (2006) Endosperm: an integrator of seed growth and development. Curr Opin Plant Biol 9:664–70. doi: 10.1016/j.pbi.2006.09.015

Bishopp A, Benková E, Helariutta Y (2011) Sending mixed messages: auxin-cytokinin crosstalk in roots. Curr Opin Plant Biol 14:10–6. doi: 10.1016/j.pbi.2010.08.014

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

Brocard IM, Lynch TJ, Finkelstein RR (2002) Regulation and role of the Arabidopsis abscisic acid-insensitive 5 gene in abscisic acid, sugar, and stress response. Plant Physiol 129:1533–43. doi: 10.1104/pp.005793

Chen X (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303:2022–5. doi: 10.1126/science.1088060

Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis. Plant Cell 19:2430–9. doi: 10.1105/tpc.107.053009

Cheng ZJ, Zhao XY, Shao XX, et al (2014) Abscisic acid regulates early seed development in Arabidopsis by ABI5-mediated transcription of SHORT HYPOCOTYL UNDER BLUE1. Plant Cell 26:1053–68. doi: 10.1105/tpc.113.121566

Choe S, Tanaka A, Noguchi T, et al (2000) Lesions in the sterol delta reductase gene of Arabidopsis cause dwarfism due to a block in brassinosteroid biosynthesis. Plant J 21:431–43.

Choi Y, Gehring M, Johnson L, et al (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. Cell 110:33–42.

Conlon I, Raff M (1999) Size control in animal development. Cell 96:235–44.

Day RC, Herridge RP, Ambrose BA, Macknight RC (2008) Transcriptome analysis of proliferating Arabidopsis endosperm reveals biological implications for the control of syncytial division, cytokinin signaling, and gene expression regulation. Plant Physiol 148:1964–84. doi: 10.1104/pp.108.128108

Day SJ, Lawrence PA (2000) Measuring dimensions: the regulation of size and shape. Development 127:2977–87.

Debeaujon I, Nesi N, Perez P, et al (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. Plant Cell 15:2514–31. doi: 10.1105/tpc.014043

Deslauriers SD, Larsen PB (2010) FERONIA is a key modulator of brassinosteroid and ethylene responsiveness in Arabidopsis hypocotyls. Mol Plant 3:626–40. doi: 10.1093/mp/ssq015

Dilkes BP, Spielman M, Weizbauer R, et al (2008) The maternally expressed WRKY transcription factor TTG2 controls lethality in interploidy crosses of Arabidopsis. PLoS Biol 6:2707–20. doi: 10.1371/journal.pbio.0060308

Doughty J, Aljabri M, Scott RJ (2014) Flavonoids and the regulation of seed size in Arabidopsis. Biochem Soc Trans 42:364–9. doi: 10.1042/BST20140040

Du L, Li N, Chen L, et al (2014) The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in Arabidopsis. Plant Cell 26:665–77. doi: 10.1105/tpc.114.122663

Duan Q, Kita D, Li C, et al (2010) FERONIA receptor-like kinase regulates RHO GTPase signaling of root hair development. Proc Natl Acad Sci U S A 107:17821–6. doi: 10.1073/pnas.1005366107

Erdmann R, Gramzow L, Melzer R, et al (2010) GORDITA (AGL63) is a young paralog of the Arabidopsis thaliana B(sister) MADS box gene ABS (TT16) that has undergone neofunctionalization. Plant J 63:914–24. doi: 10.1111/j.1365-313X.2010.04290.x

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

Escobar-Restrepo J-M, Huck N, Kessler S, et al (2007) The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. Science 317:656–60. doi: 10.1126/science.1143562

Fang W, Wang Z, Cui R, et al (2012) Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. Plant J 70:929–39. doi: 10.1111/j.1365-313X.2012.04907.x

Ferguson-Smith AC (2011) Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet 12:565–75. doi: 10.1038/nrg3032

Feuillet C, Leach JE, Rogers J, et al (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88. doi: 10.1016/j.tplants.2010.10.005

Figueiredo DD, Köhler C (2014) Signalling events regulating seed coat development. Biochem Soc Trans 42:358–63. doi: 10.1042/BST20130221

Finkelstein R (2004) Plant Hormones: Biosynthesis, Signal Transduction, Action!, Kluwer Aca. 549–573.

Finkelstein RR, Gampala SSL, Rock CD (2002) Abscisic acid signaling in seeds and seedlings. Plant Cell 14 Suppl:S15–45.

Friml J, Vieten A, Sauer M, et al (2003) Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. Nature 426:147–53. doi: 10.1038/nature02085

Fujioka S, Li J, Choi YH, et al (1997) The Arabidopsis deetiolated2 mutant is blocked early in brassinosteroid biosynthesis. Plant Cell 9:1951–62. doi: 10.1105/tpc.9.11.1951

Garcia D, Fitz Gerald JN, Berger F (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. Plant Cell 17:52–60. doi: 10.1105/tpc.104.027136

Garcia D, Saingery V, Chambrier P, et al (2003) Arabidopsis haiku mutants reveal new controls of seed size by endosperm. Plant Physiol 131:1661–70. doi: 10.1104/pp.102.018762

Gehring M, Huh JH, Hsieh T-F, et al (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124:495–506. doi: 10.1016/j.cell.2005.12.034

González-Guzmán M, Apostolova N, Bellés JM, et al (2002) The short-chain alcohol dehydrogenase ABA2 catalyzes the conversion of xanthoxin to abscisic aldehyde. Plant Cell 14:1833–46.

Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. Science 280:446–50.

Guitton A-E, Page DR, Chambrier P, et al (2004) Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. Development 131:2971–81. doi: 10.1242/dev.01168

Guo H, Ye H, Li L, Yin Y (2009) A family of receptor-like kinases are regulated by BES1 and involved in plant growth in Arabidopsis thaliana. Plant Signal Behav 4:784–6. doi: 10.1073/pnas.0812346106

Haig D, Westoby R (1989) Parent-Specific Gene-Expression and the Triploid Endosperm. Am Nat 134:147–155.

Haig D, Westoby R (1991) Genomic imprinting in endosperm: its effects on seed development in crosses between species, and between different ploidies of the same species, and its implications for the evolution of apomixis. Philos Trans Biol Sci 333:1–13.

Hamann T, Benkova E, Bäurle I, et al (2002) The Arabidopsis BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. Genes Dev 16:1610–5. doi: 10.1101/gad.229402

Hategan L, Godza B, Kozma-Bognar L, et al (2014) Differential expression of the brassinosteroid receptor-encoding BRI1 gene in Arabidopsis. Planta 239:989–1001. doi: 10.1007/s00425-014-2031-4

Haughn G, Chaudhury A (2005) Genetic analysis of seed coat development in Arabidopsis. Trends Plant Sci 10:472–7. doi: 10.1016/j.tplants.2005.08.005

Hehenberger E, Kradolfer D, Köhler C (2012) Endosperm cellularization defines an important developmental transition for embryo development. Development 139:2031–9. doi: 10.1242/dev.077057

Higuchi M, Pischke MS, Mähönen AP, et al (2004) In planta functions of the Arabidopsis cytokinin receptor family. Proc Natl Acad Sci U S A 101:8821–6. doi: 10.1073/pnas.0402887101

Hsieh T-F, Shin J, Uzawa R, et al (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A 108:1755–62. doi: 10.1073/pnas.1019273108

Hughes R, Spielman M, Schruff MC, et al (2008) Yield assessment of integument-led seed growth following targeted repair of auxin response factor 2. Plant Biotechnol J 6:758–69. doi: 10.1111/j.1467-7652.2008.00359.x

Ingouff M, Jullien PE, Berger F (2006) The female gametophyte and the endosperm control cell proliferation and differentiation of the seed coat in Arabidopsis. Plant Cell 18:3491–501. doi: 10.1105/tpc.106.047266

Jenik PD, Barton MK (2005) Surge and destroy: the role of auxin in plant embryogenesis. Development 132:3577–85. doi: 10.1242/dev.01952

Jiang W-B, Huang H-Y, Hu Y-W, et al (2013) Brassinosteroid Regulates Seed Size and Shape in Arabidopsis. PLANT Physiol pp.113.217703–. doi: 10.1104/pp.113.217703

Jiang W-B, Lin W-H (2013) Brassinosteroid functions in Arabidopsis seed development. Plant Signal Behav 8:1–5. doi: 10.4161/psb.25928

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. Plant Cell 6:1211–25.

Jofuku KD, Omidyar PK, Gee Z, Okamuro JK (2005) Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc Natl Acad Sci U S A 102:3117–22. doi: 10.1073/pnas.0409893102

Jullien PE, Susaki D, Yelagandula R, et al (2012) DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol 22:1825–30. doi: 10.1016/j.cub.2012.07.061

Kang I-H, Steffen JG, Portereiko MF, et al (2008) The AGL62 MADS domain protein regulates cellularization during endosperm development in Arabidopsis. Plant Cell 20:635–47. doi: 10.1105/tpc.107.055137

Kang X, Li W, Zhou Y, Ni M (2013) A WRKY transcription factor recruits the SYG1-like protein SHB1 to activate gene expression and seed cavity enlargement. PLoS Genet 9:e1003347. doi: 10.1371/journal.pgen.1003347

Kang X, Ni M (2006) Arabidopsis SHORT HYPOCOTYL UNDER BLUE1 contains SPX and EXS domains and acts in cryptochrome signaling. Plant Cell 18:921–34. doi: 10.1105/tpc.105.037879

Kanno Y, Jikumaru Y, Hanada A, et al (2010) Comprehensive hormone profiling in developing Arabidopsis seeds: examination of the site of ABA

biosynthesis, ABA transport and hormone interactions. Plant Cell Physiol 51:1988–2001. doi: 10.1093/pcp/pcq158

Kesavan M, Song JT, Seo HS (2013) Seed size: a priority trait in cereal crops. Physiol Plant 147:113–20. doi: 10.1111/j.1399-3054.2012.01664.x

Khan D, Millar JL, Girard IJ, Belmonte MF (2014) Transcriptional circuitry underlying seed coat development in Arabidopsis. Plant Sci 219-220:51–60. doi: 10.1016/j.plantsci.2014.01.004

Kinoshita T, Miura A, Choi Y, et al (2004) One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science 303:521–3. doi: 10.1126/science.1089835

Kiyosue T, Ohad N, Yadegari R, et al (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. Proc Natl Acad Sci U S A 96:4186–91.

Klucher KM, Chow H, Reiser L, Fischer RL (1996) The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. Plant Cell 8:137–53. doi: 10.1105/tpc.8.2.137

Kradolfer D, Hennig L, Köhler C (2013) Increased maternal genome dosage bypasses the requirement of the FIS polycomb repressive complex 2 in Arabidopsis seed development. PLoS Genet 9:e1003163. doi: 10.1371/journal.pgen.1003163

Krizek BA (2009) Making bigger plants: key regulators of final organ size. Curr Opin Plant Biol 12:17–22. doi: 10.1016/j.pbi.2008.09.006

Kunieda T, Mitsuda N, Ohme-Takagi M, et al (2008) NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in Arabidopsis. Plant Cell 20:2631–42. doi: 10.1105/tpc.108.060160

Lafon-Placette C, Köhler C (2014) Embryo and endosperm, partners in seed development. Curr Opin Plant Biol 17:64–9. doi: 10.1016/j.pbi.2013.11.008

Langridge P, Fleury D (2011) Making the most of “omics” for crop breeding. Trends Biotechnol 29:33–40. doi: 10.1016/j.tibtech.2010.09.006

Le BH, Cheng C, Bui AQ, et al (2010) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proc Natl Acad Sci U S A 107:8063–70. doi: 10.1073/pnas.1003530107

Leyser O (2005) Auxin distribution and plant pattern formation: how many angels can dance on the point of PIN? Cell 121:819–22. doi: 10.1016/j.cell.2005.06.005

Li H, Johnson P, Stepanova A, et al (2004) Convergence of signaling pathways in the control of differential cell growth in Arabidopsis. Dev Cell 7:193–204. doi: 10.1016/j.devcel.2004.07.002

Li J, Nie X, Tan JLH, Berger F (2013) Integration of epigenetic and genetic controls of seed size by cytokinin in Arabidopsis. Proc Natl Acad Sci U S A 110:15479–84. doi: 10.1073/pnas.1305175110

Li N, Li Y (2014) Ubiquitin-mediated control of seed size in plants. Front Plant Sci 5:332. doi: 10.3389/fpls.2014.00332

Li Y, Zheng L, Corke F, et al (2008) Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. Genes Dev 22:1331–6. doi: 10.1101/gad.463608

Locascio A, Roig-Villanova I, Bernardi J, Varotto S (2014) Current perspectives on the hormonal control of seed development in Arabidopsis and maize: a focus on auxin. Front Plant Sci 5:412. doi: 10.3389/fpls.2014.00412

Luo M, Bilodeau P, Dennis ES, et al (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. Proc Natl Acad Sci U S A 97:10637–42. doi: 10.1073/pnas.170292997

Luo M, Dennis ES, Berger F, et al (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. Proc Natl Acad Sci U S A 102:17531–6. doi: 10.1073/pnas.0508418102

Lur HS, Setter TL (1993) Role of Auxin in Maize Endosperm Development (Timing of Nuclear DNA Endoreduplication, Zein Expression, and Cytokinin). Plant Physiol 103:273–280.

Makarevich G, Villar CBR, Erilova A, Köhler C (2008) Mechanism of PHERES1 imprinting in Arabidopsis. J Cell Sci 121:906–12. doi: 10.1242/jcs.023077

McCarty DR (1995) Genetic Control and Integration of Maturation and Germination Pathways in Seed Development. Annu Rev Plant Physiol Plant Mol Biol 46:71–93. doi: 10.1146/annurev.pp.46.060195.000443

Mizzotti C, Mendes MA, Caporali E, et al (2011) The MADS box genes SEEDSTICK and ARABIDOPSIS B(sister) play a maternal role in fertilization and seed development. Plant J 1–12. doi: 10.1111/j.1365-313X.2011.04878.x

Morley-Smith ER, Pike MJ, Findlay K, et al (2008) The transport of sugars to developing embryos is not via the bulk endosperm in oilseed rape seeds. Plant Physiol 147:2121–30. doi: 10.1104/pp.108.124644

Nesi N, Debeaujon I, Jond C, et al (2002) The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14:2463–79.

Nowack MK, Shirzadi R, Dissmeyer N, et al (2007) Bypassing genomic imprinting allows seed development. Nature 447:312–5. doi: 10.1038/nature05770

Nowack MK, Ungru A, Bjerkan KN, et al (2010) Reproductive cross-talk: seed development in flowering plants. Biochem Soc Trans 38:604–12. doi: 10.1042/BST0380604

Ohto M, Floyd SK, Fischer RL, et al (2009) Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. Sex Plant Reprod 22:277–89. doi: 10.1007/s00497-009-0116-1

Ohto M-A, Fischer RL, Goldberg RB, et al (2005) Control of seed mass by APETALA2. Proc Natl Acad Sci U S A 102:3123–8. doi: 10.1073/pnas.0409858102

Okamuro JK, Szeto W, Lotys-Prass C, Jofuku KD (1997) Photo and hormonal control of meristem identity in the Arabidopsis flower mutants apetala2 and apetala1. Plant Cell 9:37–47. doi: 10.1105/tpc.9.1.37

Okushima Y, Mitina I, Quach HL, Theologis A (2005) AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. Plant J 43:29–46. doi: 10.1111/j.1365-313X.2005.02426.x

Peer WA, Murphy AS (2007) Flavonoids and auxin transport: modulators or regulators? Trends Plant Sci 12:556–63.

Pinyopich A, Ditta GS, Savidge B, et al (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature 424:85–8. doi: 10.1038/nature01741

Portereiko MF, Lloyd A, Steffen JG, et al (2006) AGL80 is required for central cell and endosperm development in Arabidopsis. Plant Cell 18:1862–72. doi: 10.1105/tpc.106.040824

Prasad K, Zhang X, Tobón E, Ambrose B a (2010) The Arabidopsis B-sister MADS-box protein, GORDITA, represses fruit growth and contributes to integument development. Plant J 62:203–14. doi: 10.1111/j.1365-313X.2010.04139.x

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem 379:633–46.

Riefler M, Novak O, Strnad M, Schmülling T (2006) Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. Plant Cell 18:40–54. doi: 10.1105/tpc.105.037796

Roszak P, Köhler C (2011) Polycomb group proteins are required to couple seed coat initiation to fertilization. Proc Natl Acad Sci U S A 108:20826–31. doi: 10.1073/pnas.1117111108

Schoft VK, Chumak N, Choi Y, et al (2011) Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte. Proc Natl Acad Sci U S A 108:8042–7. doi: 10.1073/pnas.1105117108

Schruff MC, Spielman M, Tiwari S, et al (2006) The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. Development 133:251–61. doi: 10.1242/dev.02194

Schuettengruber B, Cavalli G (2009) Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development 136:3531–42. doi: 10.1242/dev.033902

Scott R, Tratt J, Bolbol A (2013) Polyploid and Hybrid Genomics, John Wiley. Oxford, UK

Scott RJ, Spielman M, Bailey J, Dickinson HG (1998) Parent-of-origin effects on seed development in Arabidopsis thaliana. Development 125:3329–41.

Steffen JG, Kang I-H, Portereiko MF, et al (2008) AGL61 interacts with AGL80 and is required for central cell development in Arabidopsis. Plant Physiol 148:259–68. doi: 10.1104/pp.108.119404

Sun X, Shantharaj D, Kang X, Ni M (2010) Transcriptional and hormonal signaling control of Arabidopsis seed development. Curr Opin Plant Biol 13:611–20. doi: 10.1016/j.pbi.2010.08.009

Takahashi N, Nakazawa M, Shibata K, et al (2005) shk1-D, a dwarf Arabidopsis mutant caused by activation of the CYP72C1 gene, has altered brassinosteroid levels. Plant J 42:13–22. doi: 10.1111/j.1365-313X.2005.02357.x

Terasaka K, Blakeslee JJ, Titapiwatanakun B, et al (2005) PGP4, an ATP binding cassette P-glycoprotein, catalyzes auxin transport in Arabidopsis thaliana roots. Plant Cell 17:2922–39. doi: 10.1105/tpc.105.035816

Tiwari S, Spielman M, Schulz R, et al (2010) Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. BMC Plant Biol 10:72. doi: 10.1186/1471-2229-10-72

Tsukaya H (2006) Mechanism of leaf-shape determination. Annu Rev Plant Biol 57:477–96. doi: 10.1146/annurev.arplant.57.032905.105320

Turk EM, Fujioka S, Seto H, et al (2003) CYP72B1 inactivates brassinosteroid hormones: an intersection between photomorphogenesis and plant steroid signal transduction. Plant Physiol 133:1643–53. doi: 10.1104/pp.103.030882

Ulmasov T, Hagen G, Guilfoyle TJ (1999) Activation and repression of transcription by auxin-response factors. Proc Natl Acad Sci U S A 96:5844–9.

Vandenbussche F, Van Der Straeten D (2004) Shaping the shoot: a circuitry that integrates multiple signals. Trends Plant Sci 9:499–506. doi: 10.1016/j.tplants.2004.08.002

Vanstraelen M, Benková E (2012) Hormonal interactions in the regulation of plant development. Annu Rev Cell Dev Biol 28:463–87. doi: 10.1146/annurev-cellbio-101011-155741

Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol 27:522–30. doi: 10.1016/j.tibtech.2009.05.006

Vert G, Walcher CL, Chory J, Nemhauser JL (2008) Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. Proc Natl Acad Sci U S A 105:9829–34. doi: 10.1073/pnas.0803996105

Wabnik K, Robert HS, Smith RS, Friml J (2013) Modeling framework for the establishment of the apical-basal embryonic axis in plants. Curr Biol 23:2513–8. doi: 10.1016/j.cub.2013.10.038

Wang A, Garcia D, Zhang H, et al (2010) The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. Plant J 63:670–9. doi: 10.1111/j.1365-313X.2010.04271.x

Wang J-W, Schwab R, Czech B, et al (2008) Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. Plant Cell 20:1231–43. doi: 10.1105/tpc.108.058180

Werner T, Motyka V, Laucou V, et al (2003) Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. Plant Cell 15:2532–50. doi: 10.1105/tpc.014928

Williams L, Carles CC, Osmont KS, Fletcher JC (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. Proc Natl Acad Sci U S A 102:9703–8. doi: 10.1073/pnas.0504029102

Wolf S, Höfte H (2014) Growth Control: A Saga of Cell Walls, ROS, and Peptide Receptors. Plant Cell 26:1848–1856. doi: 10.1105/tpc.114.125518

Xia T, Li N, Dumenil J, et al (2013) The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in Arabidopsis. Plant Cell 25:3347–59. doi: 10.1105/tpc.113.115063

Xiao W, Brown RC, Lemmon BE, et al (2006) Regulation of seed size by hypomethylation of maternal and paternal genomes. Plant Physiol 142:1160–8. doi: 10.1104/pp.106.088849

Yang J, Zhang J, Huang Z, et al (2002) Correlation of cytokinin levels in the endosperms and roots with cell number and cell division activity during endosperm development in rice. Ann Bot 90:369–77.

Yang WC, Ye D, Xu J, Sundaresan V (1999) The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev 13:2108–17.

Yu F, Li J, Huang Y, et al (2014) FERONIA receptor kinase controls seed size in Arabidopsis thaliana. Mol Plant 7:920–2. doi: 10.1093/mp/ssu010

Zhang Y, Liang W, Shi J, et al (2013) MYB56 encoding a R2R3 MYB transcription factor regulates seed size in Arabidopsis thaliana. J Integr Plant Biol 55:1166–78. doi: 10.1111/jipb.12094

Zhou Y, Zhang X, Kang X, et al (2009) SHORT HYPOCOTYL UNDER BLUE1 associates with MINISEED3 and HAIKU2 promoters in vivo to regulate Arabidopsis seed development. Plant Cell 21:106–17. doi: 10.1105/tpc.108.064972

Part III

Contents

"The genetic network involving *SEEDSTICK* and *AUXIN REPONSE FACTOR2* in the control of seed size in *Arabidopsis thaliana*"
Manuscript in preparation

The genetic network involving *SEEDSTICK* and *AUXIN REPONSE FACTOR2* in the control of seed size in *Arabidopsis thaliana*.

Authors: Dario Paolo[1], Gregorio Orozco-Arroyo[1], Simona Masiero[1], Elisabetta Caporali[1], Arp Schnittger[2], Stefan de Folter[3], Lucia Colombo[1*]

Author affiliation:

[1] Dipartimento di BioScienze, Università degli Studi di Milano, Via Giovanni Celoria 26 - 20133 Milano, Italy

[2] Institut de Biologie Moleculaire des Plantes, 12 Rue du Général Zimmer - 67000 Strasbourg, France

[3] Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato, México.

[*] To whom correspondence should be addressed (lucia.colombo@unimi.it)

ABSTRACT

Seed size is the result of complex molecular networks controlling the development of the seed coat (of maternal origin) and of the embryo and the endosperm (the fertilization products). In this study we characterized the role of two transcription factors in the control of seed size of *Arabidopsis thaliana*: SEEDSTICK (STK), a MADS-domain protein, and AUXIN RESPONSE FACTOR 2 (ARF2), belonging to the ARF family. The loss-of-function mutant *stk* produces smaller seeds than wild type, while the seeds of the *arf2-8* mutant are larger than wild type. We show that these phenotypes are due to alterations of seed coat morphology, that we characterized in details. *STK* is mainly a positive regulator of cell expansion, and *ARF2*, whose expression is activated by *STK*, acts as a repressor of cell proliferation and cell expansion, in a pathway also controlled by the brassinosteroids. Interestingly, we proved that STK has also a function in directly promoting cell cycle progression via the transcriptional activation of *E2fa*, via a molecular mechanism that involves chromatin remodeling at the *E2fa* locus. Furthermore, we characterized the seeds of double mutant *stk arf2-8*, and we concluded that while *STK* and *ARF2* have antagonistic activities in controlling the increase in dimension of the seed coat, they also maintain independent roles in controlling other aspects of seed coat metabolism, such as proanthocyanidins biosynthesis and mucilage release. Altogether, our results unveil a new genetic network active in maternal control of seed size in Arabidopsis.

INTRODUCTION

In Spermatophyte plants, seeds are the vehicle by which plants rely for their propagation, and the reproductive success depends on the possibility of producing a healthy seed set. In this context, seed size represents one of the aspect that mostly influences plant fitness, with several studies suggesting that the seedlings of large-seed plants are better able to tolerate many of the stresses encountered during seedling establishment, whereas plants with smaller seeds have higher colonization abilities related to the possibility of producing a large number of seeds, as it happens for example in model plant *Arabidopsis thaliana* (Lord and Westoby, 2012; Moles et al., 2005). Moreover, in agriculture, food grains have been subjected to selection and breeding for size, as well as for other qualities, with the result that nowadays crops have usually larger and heavier seeds in comparison to wild-type relatives (Fan et al., 2006; Kesavan et al., 2013; Kim et al., 2009; Orsi and Tanksley, 2009).

Although the importance of seed size determination and the fact that this trait has been used and modified by breeders over the last centuries, only in the last few years molecular regulators of seed size begun to be identified, mainly through studies on Arabidopsis (Van Daele et al., 2012). Several factors influence seed size, among them the parent-of-origin effect (Haig, 1987; Haig, 2013), the overall plant fertility (Hughes et al., 2008), cell cycle/expansion regulation processes and hormonal signaling, particularly in relation to auxin and brassinosteroids which cascades that interact with all the factors mentioned above (Jiang and Lin, 2013).

Final seed size is achieved through coordinated growth of the three parts that compose the seeds: the seed coat, the endosperm and the embryo (Garcia et al., 2005). All of these three parts are characterized by a different genotype: the

seed coat, deriving from the ovule integuments, is entirely of maternal origin, whereas both embryo and endosperm are the result of the double fertilization. The Arabidopsis seed size is mainly determined at 4 days after fertilization (DAF) by the dimension of the seed cavity, due to the coordinated increase in size of the seed coat, that temporally precedes the proliferation of the endosperm and the growth of the embryo (Sun et al., 2010a).

The development of the seed coat in terms of cell proliferation and cell expansion is highly regulated. For instance, *CYP78A5 (KLUH)* promotes cell proliferation in the seed coat, thus affecting seed size (Adamski et al., 2009), while *AUXIN RESPONSE FACTOR2 (ARF2)* exerts an opposite role on cell proliferation and *arf2* mutant seeds are known to be both larger and heavier than wild-type due to enlarged integuments (Okushima et al., 2005; Schruff et al., 2006). *ARF2* encodes a B3-type transcription factor of the ARF family (Li et al., 2004), and the activity of its encoded protein is negatively regulated by brassinosteroids (Jiang et al., 2013; Vert et al., 2008). The brassinosteroids (BRs) are a class of steroid hormones that control several aspects of plant development, including stress responses (Krishna, 2003) and regulation of yield is mainly related to a positive effect on cell expansion in the vegetative tissues and in the reproductive organs (Divi and Krishna, 2009; Sasse, 2003). BRs regulate grain filling and dimension in rice (Tanabe et al., 2005; Wu et al., 2008) and the characterization of the BR-deficient mutant d*e-etiolated 2* (*det2*) has further confirmed their positive correlation with seed size also in Arabidopsis (Jiang et al., 2013). Although the BRs cascade is yet to be completely unveiled, we know that the BRs signal is perceived by the cell surface receptor BR-INSENSITIVE1 (BRI1), which in turn activates the two main players of the cascade, the transcriptional factors BRASSINAZOLE-RESISTANT 1 (BZR1) and the BRI1-EMS-SUPPRESSOR 1 (BES1), which

control a large number of BR-related genes; among the identified BR-regulated genes, several are associated to pathways determining cell wall structural properties, cytoskeleton formation and biosynthesis of hormones (Fang et al., 2012; Goda et al., 2002; He et al., 2005; Kim et al., 2009; Müssig et al., 2002; Sun et al., 2010b; Vert et al., 2005; Yin et al., 2002). Beside the brassinosteroids, also auxin has been reported to be involved in the determination of seed size. *ARF2* has been proposed as a possible molecular link between the BRs and the auxin cascades in organ size determination (Vert et al., 2008).

There are also many example of mutants affected in seed size due to alteration in cell expansion. Among them *TRANSPARENT TESTA GLABRA2 (TTG2)* increases seed size by promoting cell expansion in the integuments (Garcia et al., 2005), on the contrary, *APETALA2 (AP2)* represses cell expansion in the seed coat (Jofuku et al., 2005; Ohto et al., 2005; Ohto et al., 2009). Finally, the ubiquitination pathway influences the dimensions of the seed coat, further highlighting the maternal control and regulation in determining the seed size (reviewed by Li and Li, 2014).

Given the importance of the maternal tissues in determination of final seed size, we decided to focus on the Arabidopsis network affecting the formation of the seed coat directly influencing the dimension of the seeds. Loss-of-function mutants for the MADS-domain transcription factor *SEEDSTICK* (*STK*), produce smaller seeds than wild type, albeit with no obvious defects in overall plant fertility (Pinyopich et al., 2003). *STK* has been largely reported in relation to its redundant function in ovule identity determination and for its role in the specification of one layer of the seed coat, the endothelium (Pinyopich et al., 2003; Favaro et al., 2003; Mizzotti et al., 2011). *STK* is also known for its fundamental role in flavonoids (proanthocyanidins/ PAs) biosynthesis in the

seed coat (Mizzotti et al., 2014, *in press*, see Part II of this thesis). PAs are involved in the seed coat-integument cross talk, and act as as positive regulators of endosperm cellularization and seed size, possibly upstream of auxin cascade (Doughty et al., 2014).

In this study we characterize in detail the phenotype of *stk* and *arf2* seeds, revealing a genetic network involving both these transcription factors, for which we present a novel role in the control of seed coat development.

Our data show evidences that *ARF2* and *STK* respectively control cell proliferation and expansion regulation in maternal tissues, and that they are both pivotal regulator of seed size. The hormonal regulation of this genetic network via the brassinosteroids cascade is presented as well, and the data obtained are used to propose and discuss a model that explain the interaction these factors studied in the determination of seed size.

RESULTS

ARF2 and *STK* control seed coat development

To unveil one of the genetic network determining seed size in Arabidopsis, we analyzed the role of two different genes during seed development: *STK* that encodes a MADS-domain transcription factor (Colombo et al., 2008; Favaro et al., 2003; Pinyopich et al., 2003) and *ARF2*, of the *ARF* family (Li et al, 2004). It was previously reported that both *STK* and *ARF2* influence seed size in opposite ways (Pinyopich et al., 2003; Schruff et al., 2006).

To define the possible functional connection between these two genes, we further characterized the phenotype of the two single mutants as well of the double mutant *stk arf2*. In this study we used loss-of-function insertional mutants *stk* and *arf2-8* (Okushima et al., 2005; Pinyopich et al., 2003) (Suppl. Fig. 1).

In order to quantify the differences in seed size, the different mutants were grown in controlled conditions together with wild-type *Col-0* control plants and three weeks after manual pollination (DAP) mature dry seeds were harvested from individual plants to evaluate their size and shape. Manual pollination was necessary as it was already described that self-pollination of the *arf2-8* mutant is impaired due to elonged pistils in the flowers (Schruff et al., 2006), a trait that is maintained in double mutant *stk arf2-8* pistils (data not shown). Mutant and wild-type seeds differ in average seed mass (Fig. 1), with *stk* seeds being smaller than *Col-0* (approximately -15%) and *arf2-8* seeds being larger (circa +14%). Interestingly, we found that the double mutation in both *STK* and *ARF2* only partially recovered wild-type seed mass, as the *stk arf2-8* plants still develop seeds significantly larger than wild-type (+9%). Moreover, *stk* seeds are characterized by rounder-like shape, different from the more oval-like shape

typical of the wild-type seeds, a characteristic that is much reduced in the double mutant *stk arf2-8* (Fig. 1 A).

It has been reported that *STK* controls seed coat development by the coordination of the biosynthesis and release of the mucilage and flavonoid synthesis and accumulation of PAs (Ezquer and Colombo, personal communication; Mizzotti et al., *in press*) therefore we also explored the role of *ARF2.* in these processes. The mucilage is a polysaccharide synthetized and released by the cells of the outer layer of the seed coat. Seeds hydration leads to mucilage inflation and consequent bursting of the primary cell wall, so that mucilage is released upon germination (Western et al., 2000; Windsor et al., 2000). Ruthenium Red, a pectin-staining dye has been used to follow mucilage release of *Col-0, arf2-8, stk* and *stk arf2-8* seeds (Hanke and Northcote, 1975). *Col-0* and *arf2-8* seeds similarly released mucilage, while no release was observed in the *stk* and in *stk arf2-8* mutant seeds (Fig.2). We performed scanning electron microscope observations of the mutant seeds to further analyze the seeds morphology. From the comparison of the seeds surface layers, *stk* epidermal cells are smaller than wild-type ones. Conversely *arf2*-8 and *stk arf2-8* seeds both presented seeds with altered surface, where the distinction between the different cells was less clear than wild-type (Fig. 3).

PAs accumulation can be visualized using a vanillin staining (Debeaujon et al., 2003); *stk* seeds progressively accumulate PAs also ectopically outside the endothelium in the seed coat (Mizzotti et al., 2014*, in press*) while *ARF2* instead does not appear involved in this pathway, since *arf2-8* staining is identical to the one of *Col-0* seeds and *stk arf2-8* mutant seeds phenocopy *stk* seeds (Fig. 4)

STK and *ARF2* maternally control seed size.

It is well established that in Arabidopsis final seed size largely depends on the formation of the seed cavity, the empty space that is enclosed by the maternal seed coat, and that will be occupied by the proliferating endosperm and later on by the growing embryo (Sun et al., 2010; Sundaresan, 2005). We previously show that that *STK* is expressed in developing ovules and seeds (Brambilla et al., 2007; Mizzotti et al., 2011). Temporal and spatial confocal analyses of the transgenic *pSTK::STK-GFP* seeds reveal that STK protein is present in the seed coat until 4 days after fertilization (Mizzotti et al., 2014*, in press*). The activity of a putative *ARF2* promoter during seed development was already evaluated with reporter construct *pARF2::3XGFP-SV40* and it was similarly detected that *ARF2* is expressed throughout the seed coat after fertilization (Schruff et al., 2006). We therefore wanted to access the genetics of the phenotypes observed in the *stk* and *arf2-8* mutants by performing reciprocal crosses with wild-type *Col-0* plants. By using these genetic assays we could demonstrate that the seed size phenotypes of both *stk* and *arf2-8* were independent of the pollination partner, indicating that these transcription factors influence seed size by acting in the seed coat (Tab. 1), thus maternally controlling this trait.

Female genotype		**Male genotype**	**Average seed mass (µg)**
Col-0	x	*Col-0*	14,38 ± 0,16
stk	x	*stk*	12,10 ± 0,21
Col-0	x	*stk*	14,13 ± 0,29
stk	x	*Col-0*	12,33 ± 0,24
arf2-8	x	*arf2-8*	16,20 ± 0,32
Col-0	x	*arf2-8*	14,51 ± 0,23
arf2-8	x	*Col-0*	16,11 ± 0,28

Tab. 1 Reciprocal crosses with wild-type plants.

Consequently, in order to better analyze the morphology of the seed coat, seeds from the mutants taken in consideration were Fuelgen-stained (Braselton et al., 1996) and observed through confocal microscope. This cell-wall specific staining was performed to quantify the main differences between the mutants in terms of cell proliferation (average cell number counted in the seed coat) and cell expansion (average cell length). Seeds were observed at 4 DAP (Fig 5 A), a time point were the increase in dimension of the seeds reaches a peak, and the final volume of the inner seed cavity is established (Sun et al., 2010). We performed measurements for both the outermost layer of the seed coat (oi2) and for the innermost one (ii1) (Fig 5 B). Interestingly, the seed phenotype of the *stk* mutant is likely to be ascribable to defects in cell expansion in the seed coat, as *stk* cells of the oi2 resulted shorter than in the wild type, while at the same time we have not been able to detect a significant reduction of cell proliferation in *stk* seeds (Fig 5 B). On the other hand, coherently with what was previously reported about the repressive action of *ARF2* on proliferation (Schruff et al., 2006), the seed coat of the *arf2-8* mutant presented more cells in the layers analyzed, as well as increased average cell length in ii1 (Fig. 5 B). The double mutant *stk arf-8* only partially recovers the wild-type phenotype because while average cell number is intermediate between *arf2-8* and *stk* single mutants, the *arf2-8 stk* seed coat still clearly shows defective cell length, comparable to the *stk* single mutant (Fig. 5 B).

Since the phenotype of *stk* is partially rescued in the *stk arf2-8* mutant, we wanted to test if *STK* and *ARF2* regulate each other. We performed qRT-PCR on cDNA obtained from siliques between 0-4 DAP of the mutants and of the wild type. While *STK* negatively controls *ARF2* expression level, as we concluded from the overexpression of *ARF2* in the *stk* mutant, we didn't detect reciprocal control of *STK* expression via *ARF2* (Fig. 6).

To better understand the phenotype of the mutants we also evaluated, in the different backgrounds, the expression levels of cell-cycle related genes in developing seeds. This was done in order to understand if genes previously reported for their control on cell cycle could be downstream of *STK* or *ARF2*.
We observed that *ARF2* negatively controls *CYCD3;1*, a positive marker of mitotic activity in the Arabidopsis seed (Collins et al., 2012). Also the expression level of *AINTEGUMENTA* (*ANT*) is increased in the *arf2-8* mutant; *ANT* encodes for an AP2-like transcription factors that promotes cell proliferation in different organs, including the flower, and it was shown to be crucial for the correct development of ovule integuments (Baker et al., 1997; Krizek and Anderson, 2013). Differently from what happens the *arf2-8* mutant background, the expression levels of *CYCD3;1* and *ANT* do not change in the *stk* mutant background. On the other hand master regulators of cell cycle progression *CDKB1;1* (Boudolf et al., 2004) and *E2Fa* (De Veylder et al., 2002; Vlieghe et al., 2005) are positively regulated by *STK* but not by *ARF2* (Fig. 6).

STK controls cell cycle progression via direct activation of *E2Fa*

The over-expression of *E2Fa* in *stk* is of particular interest because, despite the fact that cell proliferation in the seed coats of the *stk* mutant is not significantly altered (Fig. 5 B), we were able to identify differences between *stk* and wild-type seeds when comparing them in terms of cell cycle progression via flow cytometry. With this technique, thanks to the fluorence-activated cell sorting (FACS) it is in fact possible to discriminate, in the heterogeneous cell population of developing seeds, relative quantitative representation of each cell type. For this reason, 6 DAP seeds of *Col-0* and *stk* plants were collected, DAPI-stained and subsequently analyzed, comparing their DNA-content

profiles. We concluded that *stk* seeds did not differ from wild-type ones in terms of presence of triploid nuclei corresponding to endosperm (3c and 6c peaks, Fig 7 A, B); the fact that cell cycle progression is not affected in *stk* confirms the genetic data shown before, that indicated that the *stk* phenotype is due to altered seed coat development. The only difference observed is in terms of 4c and 2c peaks, as it was previously suggested (Mizzotti et al., 2011) (Fig. 7 A, B). While, in *stk* seeds, the 4c peak is representative of close to 50% of the total sample, in wild-type seeds represents less than 35% of the total sample (Fig. 7 B). Interestingly, 4c peak at this stage of seed development is mainly influenced by seed coat cell in G_2 phase of the cell cycle and a high 4c/2c ratio measured in an organ usually is interpreted as an alteration of the balance between mitosis / endoreduplication, a checkpoint known to be positively regulated by *E2Fa* (Boudolf et al., 2004; He et al., 2004; Sozzani et al., 2006).
To understand the nature of *E2Fa* regulation via *STK*, we analyzed the *E2Fa* locus and identified two putative CaRG boxes, the $CC[A/T]_6GG$ consensus sequences known to be recognized by MADS-domain transcription factors, including STK (Matias-Hernandez et al., 2010; Nurrish and Treisman, 1995) (Fig. 8 A). We tested whether *E2Fa* could be a direct target of STK, via a chromatin immunoprecipitation (ChIP) assay with an anti-STK antibody. Unspecific binding of the anti-STK antibody was achieved by using the double mutant *shp1 shp2* as positive control, as suggested previously (Matias-Hernandez et al., 2010), while negative control was represented by the *stk* mutant. Immunoprecipitated chromatin obtained from 0-4 DAP fruits was tested via qPCR analysis, showing for one of the STK putative binding of *E2Fa* a significant enrichment over the negative control (Fig. 8 B). These data show indeed *E2Fa* expression is directly regulated by STK in developing seeds.

To understand the molecular mechanism involved in *E2Fa* regulation via STK, we investigated on the possibility that the binding of STK on its target might correlate with a local change in the chromatin landscape, similarly to *BANYULS (BAN)* regulation (Mizzotti et al., 2014, *in press*). For this reason we tested, by means of ChIP experiment using a specific antibody the level of H3-lysine9-acetylation (H3K9ac), an epigenetic mark usually associated with activation of expression (Lauria and Rossi, 2011). The level of the H3K9ac mark was assessed at the STK binding site on *E2Fa*, on both wild-type and *stk* 0-4 DAP seeds. This region presented a lower level of the H3K9ac in *stk* mutant background (Fig. 8 C) and therefore we conclude that the down-regulation of *E2fa* in *stk* seeds correlates with a reduced level of this euchromatic at the STK binding site

Interplay between brassinosteroids, STK and ARF2 in seed size determination.

It is well established that auxins and brassinosteroids (BRs) control several aspects of plant growth, in the frame of regulation of cell proliferation and expansion. Strong evidences support the idea of interdependence and synergism between these two pathways (Halliday, 2004; Nemhauser et al., 2004) and a recent study investigated in detail on the role of BRs in the regulation of seed size and shape in Arabidopsis (Jiang et al., 2013). Interestingly, the post-translational regulation of ARF2 protein has been proposed as one of the possible molecular links between the two hormonal cascades of auxin and BRs in the control of organ size (Vert et al., 2008). For this reason we wanted to evaluate whether also the regulation of seed size via-*STK* is downstream of the BRs signaling. To test this, exogenous application of brassinolide (BL), a compound that represents the first isolated of the brassinosteroid, was

performed repeatedly every 24h on young buds and resulting siliques until 4 days after fertilization, on both wild-type *Col-0* and *stk* plants. Mature seeds were then harvested 3 weeks later to be measured (Fig. 9 A). The results indicate that BL-application increased seed size in comparison to mock-control both in the wild type and in the *stk* mutants; however *stk* mutant seeds only partially recovered wild-type size (Fig. 9 B), further suggesting that the *STK* function in indispensable for the achievement of final seed size. Finally, we also reported that *ARF2* expression is not affected in BL-treated siliques from 0-4 DAP and mock control, nor in wild-type nor in *stk* background, and that similarly the expression of *STK* does not change upon BL-treatment in wild-type and in *arf2-8* background (Fig 9 C).

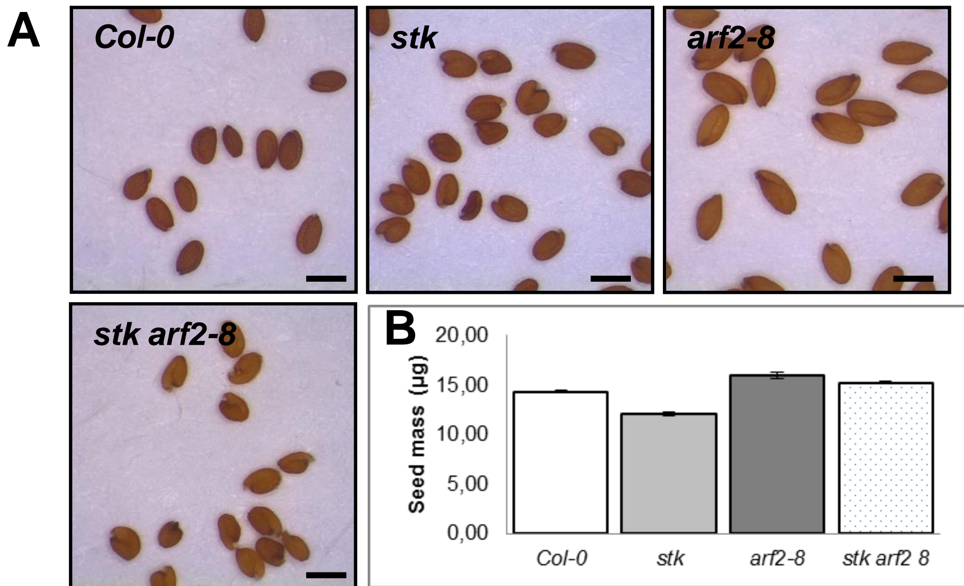


Fig.1 Size and weight comparison of mature seeds.

(A) Comparison of mature seeds collected from siliques 3 weeks after pollination of wild-type, single and double homozygous mutant plants. Scale bar = 500 µm. **(B)** Seed mass measured 3 weeks after pollination. Error bars represent st. error of the sample. All the samples are significantly different from each other (p<0,01).

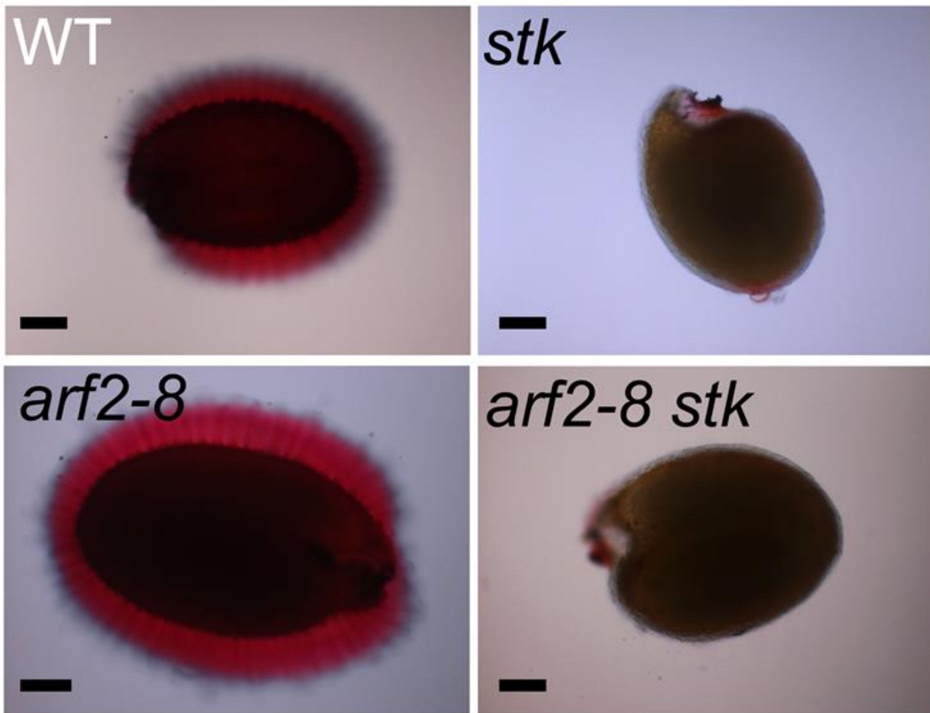


Fig.2 Mucilage extrusion assay.

Mature seeds show mucilage-specific pink staining upon hydration and Ruthenium-Red treatment. *stk* and *stk arf2-8* seeds show lack of mucilage release in comparison to wild-type and *arf2-8* seeds, indicating alteration in the structure of cell wall in the seed coat. Scale bar = 100 μm

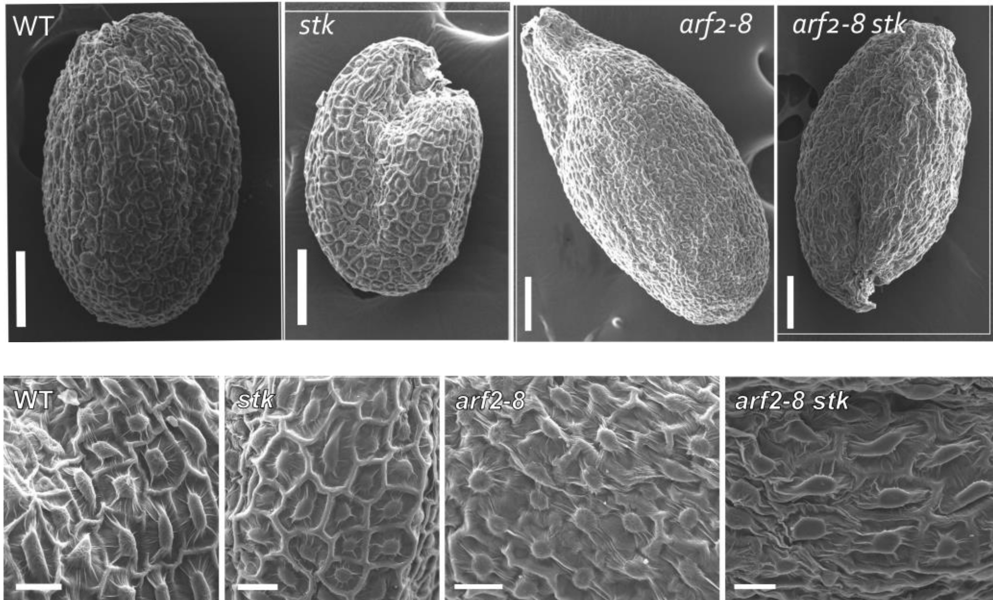


Fig.3 Scanning electron microscopy observations

SEM imaging of mature seeds. Upper part: *stk* seeds are both smaller and more rounded than wild-type seeds, while *arf2-8* and *stk arf2-8* are larger than wild-type (scale bar = 100 µm). Lower part: *arf2-8* and *stk arf2-8* epidermis structure appears altered compared to wild type, and cell-cell distinction is less evident. (Scale bar = 20 µm)

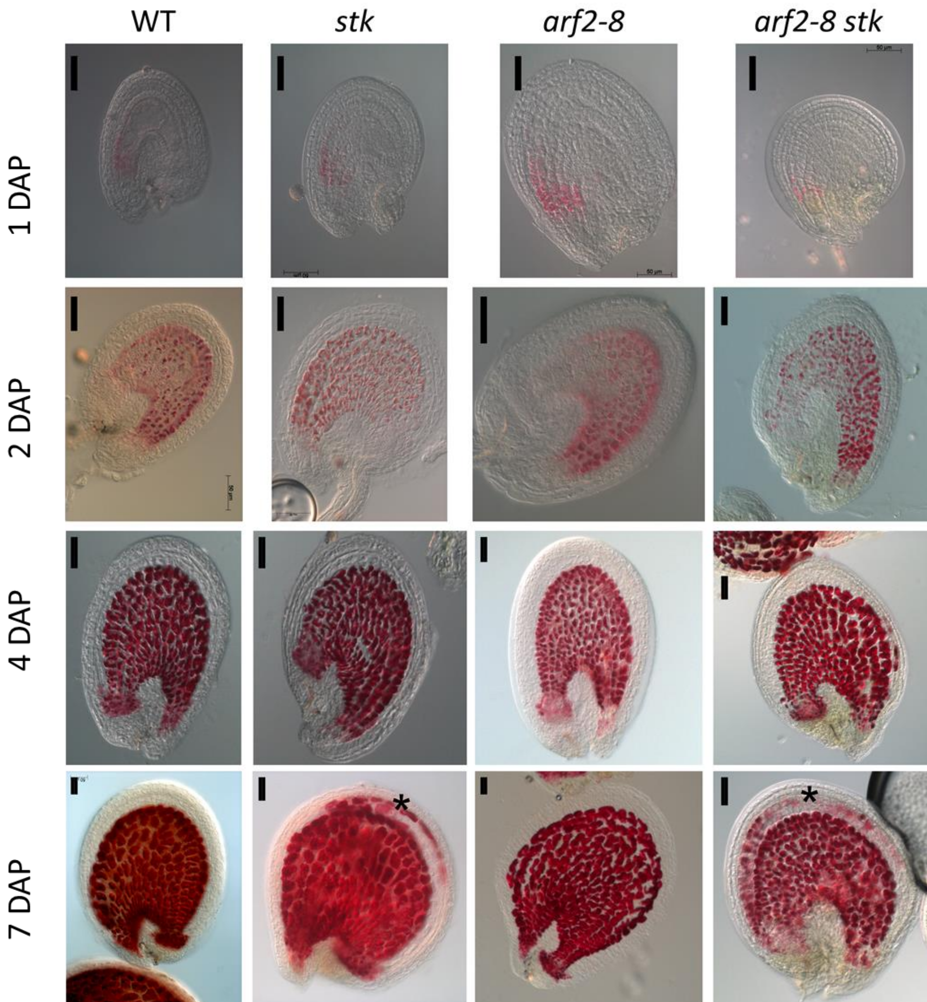


Fig.4 Vanillin assay on wild-type and mutant seeds reveal PAs accumulation.

Pink staining indicate PAs accumulation in the endothelium of developing seeds. PAs biosynthesis starts in the endothelium after fertilization and gradually procedes from the micropylar region of the seed towards the chalazal pole. In *stk* and *stk arf2-8* ectopic accumulation of PAs is observed outside of the endothelium in the seed coat (asterisks). DAP = days after pollination. Scale bar = 50 µm

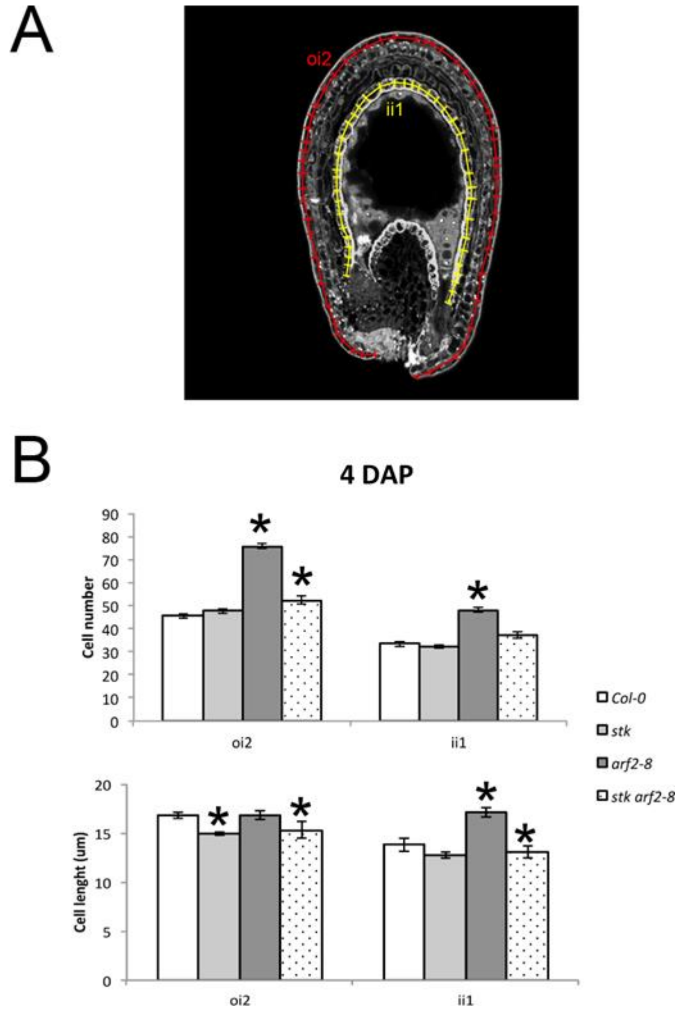


Fig.5 Seed coat morphology analysis (A) Fuelgen-stained seed at 4 DAP. oi2 = outer integument 2; ii1 = inner integument 1 **(B)** Comparison of average cell number and average cell length in layer oi2 and ii1. Error bars represent st. error of the sample. Asterisks indicate that the sample is significantly different from wild-type *Col-0* control (p<0,01).

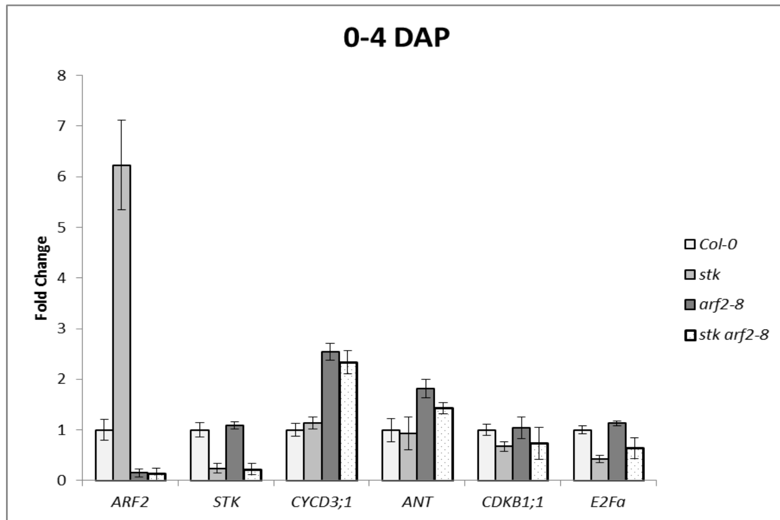


Fig.6 Expression analysis of cell cycle related genes in developing seeds (0-4 DAP).

Relative expression level (normalized for *ACTIN* and *UBIQUITIN*) of different genes related to cell cycle control, tested on 4 DAP siliques.

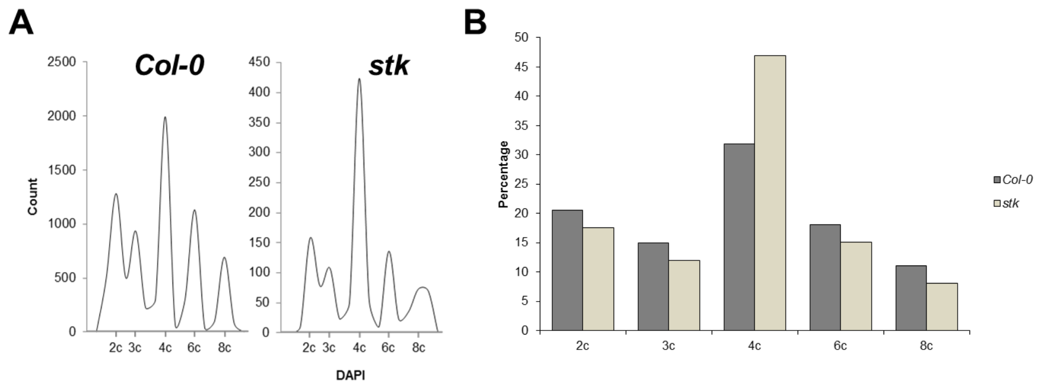


Fig.7 DNA-content profile obtained from FACS analysis of seeds.

(A) DNA-content profile of wild-type *Col-0* and of *stk* mutant seeds showing the different peaks related to diploid nuclei (2c, 4c), highly endo-reduplicated nuclei (8c) and triploid/endospermal nuclei (3c, 6c). **(B)** quantitative representation of each peak relative to total measured samples.

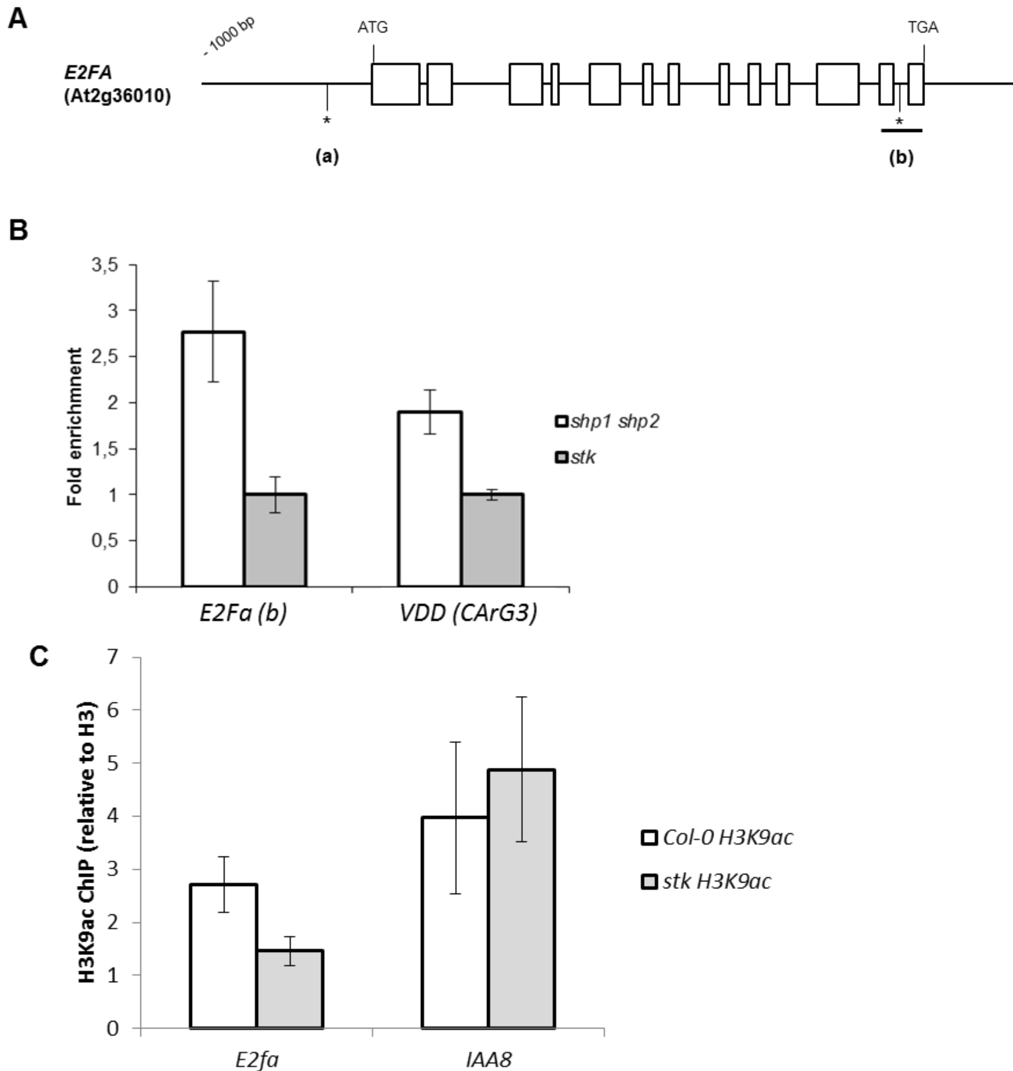


Fig.8 *E2Fa* regulation via STK protein.

(A) Schematic representation of *E2Fa* gene with intron/exon structure, with positions of CArG boxes (region a, b) tested by ChIP. **(B)** ChIP enrichment tests by qPCR show that region “b” of *E2Fa* is a direct target of STK. **(C)** H3K9 acetylation level in *E2Fa* region “b” of wild-type and stk expressed as % input, normalized on total H3 level (See Material and Methods).

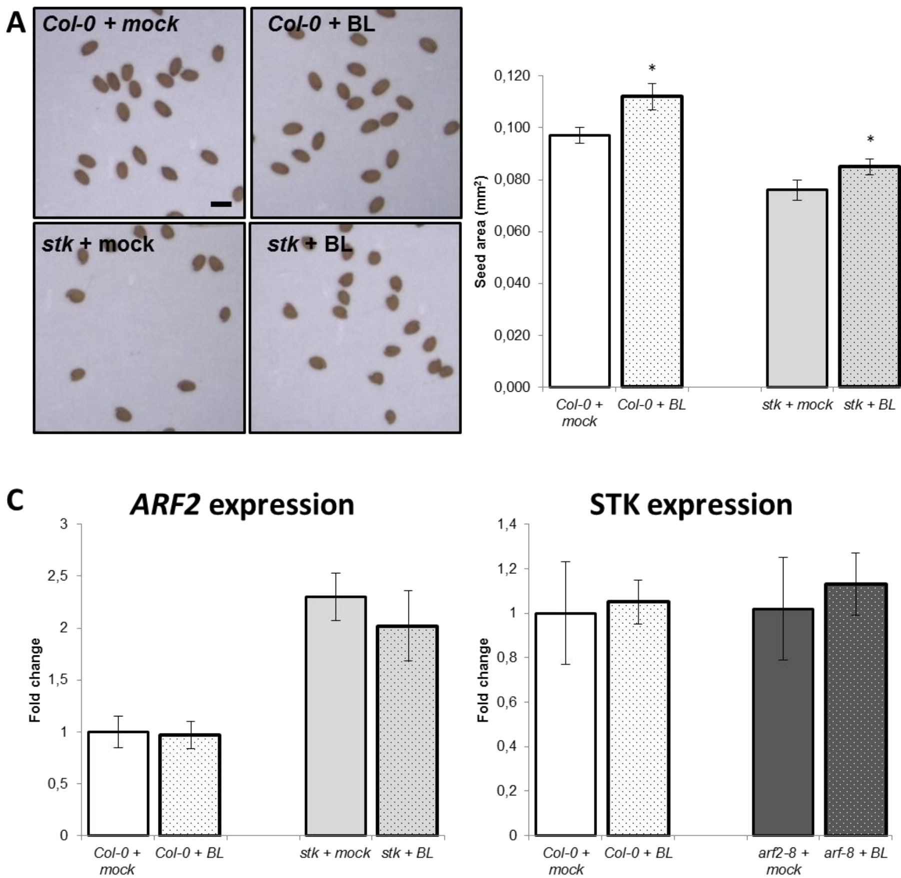


Fig.9 Effects of brassinolide treatment on *Col-0* and *stk* mutant seeds.

(A) Comparison between mature seeds collected from siliques treated with 1 μm brassinolide (BL) and relative mock-control **(B)** Seed size comparison of plants treated with BL and mock-control. Asterisks indicate that the sample is significantly different from relative mock control (p<0,01) **(C)** Relative expression level (normalized for *ACTIN* and *UBIQUITIN*) of ARF2 and STK upon BL-treatment.

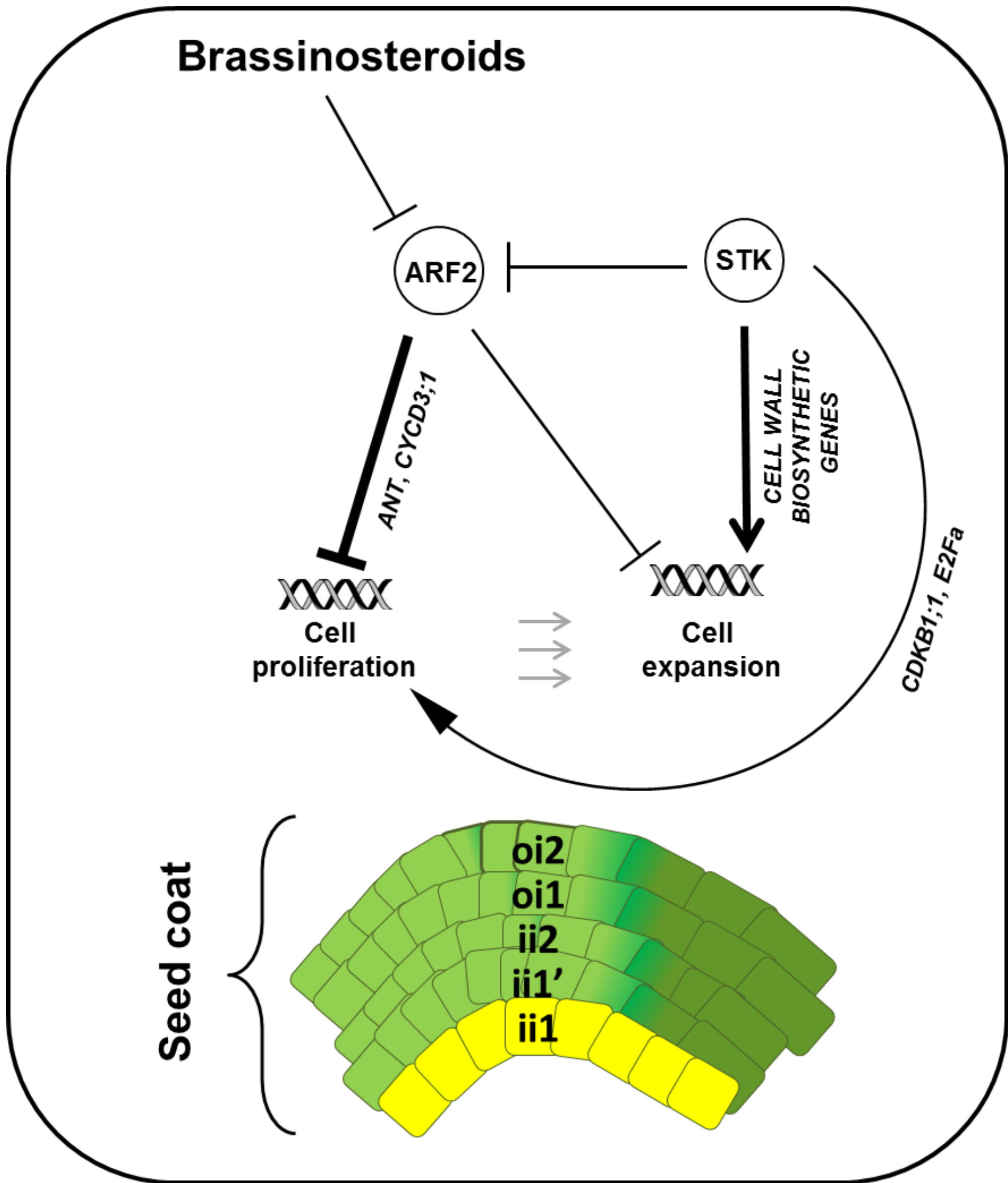
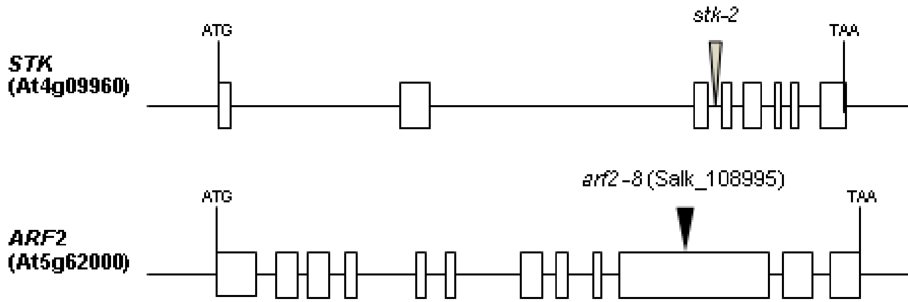


Fig.10 Proposed model of regulation of seed coat cell proliferation and expansion.



Suppl. Fig.1 Mutant alleles analyzed.

Schematic representation of *STK* and *ARF2* genes with intron/exon structure. Triangles represent the position of the insertion in the mutant alleles *stk-2* and *arf2-8*.

Primers for genotyping	
stk-2	5’-GCTTGTTCTGATAGCACCAACACTAGCA-3’
	5’-CCACTAACCATTTGATGATGGTGTTGT-3’
arf2-8	5’-TGAGTCCTGTTCCAATGCC-3’
	5’-GAAAGATCCACTGAACGGC-3’
Primers for expression analysis	
STK	5’-ACGCGCAGAAAAGGGAGATTGAGC-3’
	5’- TGTCGGGATCAGAGTAAGAACCTCC-3’
ARF2	5’- CCTCATCCGAAGGATGCTCAAACG-3’
	5’- GGAGCCATCAACTCTCCATTGAACTC-3’
CYCD3;1	5’- GGAGAAGGAACATACGGGAAAGTC-3’
	5’- ACGAAGAGCAGTTGGTGGAATAC-3’
ANT	5’- CAAGCACGGATTGGTAGAGTCG-3’
	5’- CATTAGGGTTTGATGTCCAAGG-3’
E2fa	5’- CTTTGGAAGGGAGTTGATGCGTGTC-3’
	5’- TTGTGCCATGAGGAGCTTTGACG-3’
CDKB1;1	5’- GGAGAAGGAACATACGGGAAAGTC-3’
	5’- ACGAAGAGCAGTTGGTGGAATAC-3’
ACT	5’- CTCAGGTATTGCAGACCGTATGAG-3’
	5’- CTGGACCTGCTTCATCATACTCTG-3’
UBQ	5’- CTGTTCACGGAACCCAATTC-3’
	5’- GGAAAAAGGTCTGACCGACA-3’
Primers for ChIP experiment	
E2fa	5’- ACTGGCTTCTCTCAAATGCTGAG-3’
	5’- GGTGCTATTTCGCCCATTCCT-3’
VDD	5’- GGAAATATGACGTTGTCTTTTTTAG-3’
	5’- CAGAAACAGCAATATGGTCGTG-3’
IAA8	5’- GAACGTAACTACTTGGGTCT-3’
	5’- GTCTTTAGAAGGTAGCAAC-3’
ACT	5’- CCAATCGTGAGAAAATGACTCAG-3’
	5’- CCAAACGCAGAATAGCATGTGG-3’

Suppl. Tab.1 List of primers used in this study.

DISCUSSION

The embryo and the endosperm develop protected by the maternal seed coat; the coordinated growth of these three parts requires signaling and interplay between them, although their development is completed in different moments. In Arabidopsis, the formation of the seed cavity precedes the complete development of the endosperm an the embryo (Sun et al., 2010a), thus final seed size in largely dependent on seed coat formation.

We provide evidences that both the MADS-domain transcription factor STK and Auxin Response Factor ARF2 are master regulator of seed size, albeit with opposite roles. *ARF2* and *STK* affect seed coat development through the regulation of cell proliferation and of cell expansion in the seed coat, besides STK regulates *ARF2* expression in the developing seeds. We shed light on the putative pathways downstream of these two transcription factors in the control of cell cycle and on the connection between their activity and the BRs cascade, already known to regulate ARF2 and seed size (Jiang et al., 2013; Vert et al., 2008).

It has been previously reported that STK is active in the seed coat and controls different aspects of its development; it controls endothelium formation (redundantly with another MADS-domain gene *ARABIDOPSIS B-SISTER*) (Mizzotti et al., 2011), as well as secondary metabolism of flavonoids (PAs) (Mizzotti et al., 2014, *in press*), compounds thought to be involved in the seed coat-endosperm cross-talk. This interplay between the seed coat and the endosperm is a central mechanism in the control of the timing of endosperm cellularization, and consequently in the determination of seed size (Doughty et al., 2014).

Our data indeed confirm that STK controls the biosynthesis and accumulation of PAs, as *stk* ectopically accumulates PAs also outside of the endothelium, in different layers of the seed coat. On the contrary *ARF2* does not seem to be involved in the PAs metabolic pathway as *arf2-8* single mutant does not present ectopic accumulation of PAs, nor do *stk arf2-8* seeds (Fig. 4).

The defects in expansion observed in *stk*, whose seed coat cells are clearly shorter than in wild type (Fig. 5), can be explained by the pivotal role of STK in determining the biochemical and structural properties of cell walls in the seed coat. As matter of fact the *stk* seed coat at 2 DAP is significantly stiffer and more rigid than in wild-type seeds, as it was experimentally proven through the use of an atomic force microscopy approach to investigate the elastic modulus of the cell wall at the epidermis (the outermost layer) of the seed coat (Ezquer and Colombo, personal communication). Furthermore structural defects of the cell walls in *stk* seeds would explain lack of mucilage extrusion from the *stk* mutant, as improper mucilage release could be linked to increased mechanical resistance of the outer primary cell wall (Western et al., 2001) (Fig. 2).

In addition, we proved that STK is also involved in the progression of the cell cycle in the seed coat. In this frame, it is however interesting to highlight that positive markers of the mitosis (such as *E2Fa* and *CDKB1;1*) are down-regulated in the *stk* mutant during seed development (Fig. 6), although we did not observe significant reduction in cell proliferation in the *stk* seed coat. This observation brings to assume redundancy in the activation of cell proliferation via different pathways also in absence of STK, a common compensation effect of disruption between cell number/cell length ratio in developing organs. Strikingly, *E2Fa* and *CDKB1;1* control G2/M transition in cell cycle, and their down-regulation results in high 4c/2c cell ratio organs (Boudolf et al., 2004). We indeed observed via FACS that a high 4c/2c ratio is a characteristic of

developing *stk* seeds, and it indicates alteration in cell cycle progression in the seed coat (Fig. 7). The altered 4c/2c ratio in *stk* could be therefore explained by the down-regulation of *E2Fa* in *stk* seeds. We additionally revealed via ChIP that *E2Fa* is a direct target of STK during seed development, and consequently we propose that STK directly activates *E2Fa* by binding to its genomic region (Fig. 8). Previous studies indicated that protein complexes involved in chromatin remodeling can be recruited on target genes by transcription factors of different families, such as BnSCL1, a SCARECROW-like protein of *Brassica napus* (Gao et al., 2004), AtERF7 (Song et al., 2005), LEUNIG (LEU) (Gonzalez et al., 2007) and, among MADS-domain factors, AGAMOUS LIKE 15 (AGL15) (Hill et al., 2008). It has been recently shown that the regulation of STK on *BAN*, a key gene in PAs biosynthesis, likely involves local chromatin remodeling (Mizzotti et al., 2014, *in press*). Our data add new information on the possibility that epigenetic regulation of expression might be common among STK target genes. Mizzotti and colleagues showed in fact that *BAN* is over-expressed in *stk* seeds, and that coherently with that the *BAN* locus presents a higher level of H3K9ac (an euchromatic marker). The ChIP data here presented, on the contrary, indicate that the level H3K9ac seeds at the STK binding site on *E2Fa* is lower in *stk* seeds than in wild type ones, in accordance with *E2Fa* down-regulation in *stk* and the alteration in the cell cycle described above.

The activation of *E2Fa* expression has been pointed out as a branching step in the arising of lateral roots in Arabidopsis (De Veylder et al., 2002), and notably, the triple mutant *stk shp1 shp2* completely lacks lateral roots (Moreno-Risueno et al., 2010): therefore, the mechanism of positive control of *STK* on *E2Fa* expression, that we show during seed development, could be conserved also in the development of secondary radical apparatus, and the study of parallelism

between these two pathways would be of great interest in deciphering how MADS-domain transcription factor control the development of lateral organs, after their initial identity specification.

ARF2 has the opposite effect on seed size in comparison to *STK*, since prove that *ARF2* represses cell proliferation in the seed coat, acting via a distinct pathway compared to *STK*. *ARF2* acts in fact as a repressor of two positive markers of mitotic activity such as *CYCD3;1* and *ANT* also in young developing seed (Fig. 6), a regulatory mechanism already proposed for mature organs (Schruff et al., 2006). Overexpression of *ANT*, a major regulator of floral organ size (Krizek, 2009; Krizek and Anderson, 2013) has been shown to induce cell proliferation, and indeed we proved that cell proliferation is increased in the seed coats of *arf2-8*. The expression *ANT* is known to be induced also by brassinosteroids (Jiang et al., 2013), hormones that negatively regulate the ARF2 protein activity via BIN2-mediated phosphorilation. This post-translational modification loosens ARF2 binding to Aux-RE elements on target genes (Ulmasov et al., 1999) and it determines an increase in organ size (Vert at al., 2008). For these reasons the fact that *ANT* is downstream of ARF2 in developing seeds is of remarkable interest as it suggests that ARF2 regulation of cell proliferation in the seed coat might be controlled by the BRs. We indeed proved that exogenous application of brassinolide on developing fruits results in increased seed size, possibly because of the removal of ARF2-block on cell proliferation and expansion. Coherently with a mainly post-translational regulation of ARF2 via BRs, we observed that *ARF2* expression does not change after BL-treatment (Fig. 9).

Moreover, the fact that the *arf2-8* mutant has also larger cells in one of the seed coat layer analyzed (Fig. 5) indicates that *ARF2* might also be involved, besides its negative regulation on cell proliferation, also in the repression of cell

expansion; the existence of such a mechanism contributes to explain the reduced expansion observed in *stk*: the absence of STK leads not only to structural defects in the cell walls described above, that inhibit cell expansion, but it also determines overexpression of *ARF2,* that further represses elongation (Fig. 10).

From the detailed characterization of the seed coat of the *stk arf2-8* mutant, we concluded that the partial recover of the wild-type seed phenotype is due to a reduction in cell proliferation in comparison with the single mutant *arf2-8*, while the defects in cell expansion (due to the loss of STK activity) are not overcome by the simultaneous loss of ARF2 function. These characteristics of the seed coat in the *stk arf2-8* mutant indicate that *STK* is partially epistatic to *ARF2* in controlling cell expansion, while ARF2 activity is dominant in controlling cell proliferation.

Interestingly, while a significant increase in seed size (over *stk* mock control) is also observed when treating *stk* mutant plants with BL, the effect of the brassinosteroids is not sufficient to completely recover wild-type seed dimensions. This data, together with the fact that STK expression does not change upon BL-treatment, point out that the regulation of seed size via *STK* is BR-independent.

In conclusion we have characterized in detail the role of these two transcription factors, showing how their antagonistic activity in seed size determination is exerted through different pathways. We show that *STK* mainly induces cell expansion in the seed coat, even if it seems also to have a redundant function in promoting cell cycle progression. *ARF2* mainly represses cell proliferation in the same layers of the seed coat where *STK* acts. These two reciprocally influenced pathways are both pivotal in the achievement of seed size and they are differently regulated by hormonal signals, as the brassinosteroids cascade.

Our study identifies two master regulators of seed yield, confirming the importance of maternal control in seed size determination.

MATERIAL AND METHODS

Plant material and growth condition

Arabidopsis thaliana wild-type (ecotype Columbia, *Col-0*) and transgenic lines were grown on soil at 22°C under short-day (8h light / 16h dark) and long-day condition (16h light / 8h dark). *stk-2* contains a 74 nt insertion near the splice site of the 3rd intron (Pinyopich et al., 2003); *arf2-8* (SALK_108995) was previously identified as allele of *ARF2* (Schruff et al., 2006).

Genotyping

PCR-based genotyping for mutant alleles was performed with the following primer pairs: *stk-2* Atp_204 / Atp_561; *arf2-8* Atp_3666 / Atp_3667 (Supplemental Table 1).

Seed size measurement

Average seed mass was determined by weighing mature dry seeds in batches of 500 and the weights of at least three replicates were measured for each seed lot. The measurements of seed area was performed using SmartGrain software upon observation via stereo microscope (Tanabata et al., 2012).

Tissue staining and microscopy

Ruthenium-Red (0.01% w/v for 90 min, Sigma-Aldrich) staining and evaluation of mucilage extrusion on mature seeds was performed upon water imbibition as previously described (Western et al., 2001).

Scanning electron microscopy of mature dry seeds was performed by gold coating them using a sputter coater (SEMPREP2; Nanotech) and observed with a LEO 1430 scanning electron microscope (LEO Electron Microscopy).

Detailed seed coat analysis was performed upon Fuelgen-staining of 4 DAP seeds to visualize the cell walls (Braselton et al., 1996), and then observed through confocal microscopy.
Vanillin assay for PA detection was performed as described previously (Debeaujon et al., 2000). Vanillin (vanilaldehyde) condenses to PAs and flavan-3-ol precursors to give a bright-red product in acidic conditions. Microscopic observations were performed using a Zeiss Axiophot D1 microscope (http://zeiss.com/) equipped with differential interface contrast (DIC) optics. Images were recorded with an Axiocam MRc5 camera (Zeiss) using the Axiovision program (version 4.1).

Expression analyses

Total RNA was extracted from siliques from 0 until 4 days after pollination (DAP), using the LiCl method (Verwoerd et al., 1989). Total RNA was treated using the Ambion TURBO DNA-free^{TM} DNase (Ambion, www.lifetechnologies.com - Cat. AM1907) and then reverse transcribed using the Bio-Rad iScript^{TM} kit (Bio-Rad, www.bio-rad.com - Cat. 170-8891). The cDNAs were standardized relative to *UBIQUITIN10* (*UBI10*) and *ACTIN2-8* (*ACT2-8*) transcripts and gene expression analysis was performed using the iQ5 Multicolor real-time PCR detection system (Bio-Rad, www.bio-rad.com) with SYBR Green PCR Master Mix (Bio-Rad). Base line and threshold levels were set according to the manufacturer's instructions. Primers used in the experiments are listed in Supplemental Table 1.

Ploidy analysis

For flow cytometry analysis, flowers were manually emasculated 24h before manual pollination with pollen of the same genotype. Seeds were collected

from the siliques at 6 DAF and prepared for the analysis with CyStain® UV Precise kit (Partec, www.partec.com – Cat. 05-5002). The staining with DAPI (4',6-diamidino-2-phenylindole) was performed as previously described (Nowack et al., 2010). The ploidy level was calibrated against the 2C nuclear DNA content peak derived from a preparation of young rosette leaves (Mizzotti et al., 2011).

Chromatin immunoprecipitation assay

ChIP experiments to evaluate enrichment of *E2Fa* were performed with an antibody specific for STK protein, using fertilized flowers and siliques until 4 DAF. Plant material was collected from the *shp1 shp2* double mutant and from *stk* plants; qPCR data were obtained using the iQ5 Multicolor real-time PCR detection system (Bio-Rad, www.bio-rad.com) with SYBR Green PCR Master Mix (Bio-Rad) and, for the region of interest, fold enrichment over negative control (*ACT7*) was evaluated with ΔΔCt method (Matias-Hernandez et al., 2010). Similarly, relative differences in acetylation levels between the wild-type *Col-0* and the *stk* mutant were tested by ChIP using the same developmental stages mentioned above. We used an antibody against an unmodified isoform of H3 ("total H3", Upstate, www.millipore.com – Cat. 06-753) and one specific for H3K9ac (Upstate, Cat. 07-532). For each sample, percentage of enrichment in the acetylation versus input was calculated as previously shown (Lin et al., 2012), normalizing for total H3 level. IAA8 was chosen as positive control (Mizzotti et al. 2014, *in press*).

Hormonal treatments

Brassinolide (Sigma-Aldrich Cat. E1641) was diluted to 20 mM with absolute ethanol and a working solution diluted to 1 μM with water was used for

exogenous treatment on plants. The solution was applied directly on inflorences and the treatment was repeated every 24h until 4 DAP, with sterile water used as mock control.

REFERENCES

Adamski, N., Anastasiou, E., Eriksson, S., O'Neill, C. M. and Lenhard, M. (2009). Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20115–20.

Baker, S. C., Robinson-Beers, K., Villanueva, J. M., Gaiser, J. C. and Gasser, C. S. (1997). Interactions among genes regulating ovule development in Arabidopsis thaliana. *Genetics* **145**, 1109–24.

Boudolf, V., Vlieghe, K., Beemster, G. T. S., Magyar, Z., Torres Acosta, J. A., Maes, S., Van Der Schueren, E., Inzé, D. and De Veylder, L. (2004). The plant-specific cyclin-dependent kinase CDKB1;1 and transcription factor E2Fa-DPa control the balance of mitotically dividing and endoreduplicating cells in Arabidopsis. *Plant Cell* **16**, 2683–92.

Brambilla, V., Battaglia, R., Colombo, M., Masiero, S., Bencivenga, S., Kater, M. M. and Colombo, L. (2007). Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in Arabidopsis. *Plant Cell* **19**, 2544–56.

Braselton, J. P., Wilkinson, M. J. and Clulow, S. A. (1996). Feulgen staining of intact plant tissues for confocal microscopy. *Biotech. Histochem.* **71**, 84–7.

Collins, C., Dewitte, W. and Murray, J. A. H. (2012). D-type cyclins control cell division and developmental rate during Arabidopsis seed development. *J. Exp. Bot.* **63**, 3571–3586.

Colombo, L., Battaglia, R. and Kater, M. M. (2008). Arabidopsis ovule development and its evolutionary conservation. *Trends Plant Sci.* **13**, 444–50.

De Veylder, L., Beeckman, T., Beemster, G. T. S., de Almeida Engler, J., Ormenese, S., Maes, S., Naudts, M., Van Der Schueren, E., Jacqmard, A., Engler, G., et al. (2002). Control of proliferation, endoreduplication and differentiation by the Arabidopsis E2Fa-DPa transcription factor. *EMBO J.* **21**, 1360–8.

Debeaujon, I., Léon-Kloosterziel, K. M. and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiol.* **122**, 403–14.

Debeaujon, I., Nesi, N., Perez, P., Devic, M., Grandjean, O., Caboche, M. and Lepiniec, L. (2003). Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. *Plant Cell* **15**, 2514–31.

Divi, U. K. and Krishna, P. (2009). Brassinosteroid: a biotechnological target for enhancing crop yield and stress tolerance. *N. Biotechnol.* **26**, 131–6.

Doughty, J., Aljabri, M. and Scott, R. J. (2014). Flavonoids and the regulation of seed size in Arabidopsis. *Biochem. Soc. Trans.* **42**, 364–9.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X. and Zhang, Q. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–71.

Fang, W., Wang, Z., Cui, R., Li, J. and Li, Y. (2012). Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. *Plant J.* **70**, 929–39.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M. and Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* **15**, 2603–11.

Gao, M.-J., Parkin, I., Lydiate, D. and Hannoufa, A. (2004). An auxin-responsive SCARECROW-like transcriptional activator interacts with histone deacetylase. *Plant Mol. Biol.* **55**, 417–31.

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

Goda, H., Shimada, Y., Asami, T., Fujioka, S. and Yoshida, S. (2002). Microarray analysis of brassinosteroid-regulated genes in Arabidopsis. *Plant Physiol.* **130**, 1319–34.

Gonzalez, D., Bowen, A. J., Carroll, T. S. and Conlan, R. S. (2007). The transcription corepressor LEUNIG interacts with the histone deacetylase HDA19 and mediator components MED14 (SWP) and CDK8 (HEN3) to repress transcription. *Mol. Cell. Biol.* **27**, 5306–15.

Haig, D. (1987). Kin conflict in seed plants. *Trends Ecol. Evol.* **2**, 337–40.

Haig, D. (2013). Kin Conflict in Seed Development: An Interdependent but Fractious Collective. *Annu. Rev. Cell Dev. Biol.*

Halliday, K. J. (2004). Plant hormones: the interplay of brassinosteroids and auxin. *Curr. Biol.* **14**, R1008–10.

Hanke, D. E. and Northcote, D. H. (1975). Molecular visualization of pectin and DNA by ruthenium red. *Biopolymers* **14**, 1–17.

He, S. S., Liu, J., Xie, Z., O'Neill, D. and Dotson, S. (2004). Arabidopsis E2Fa plays a bimodal role in regulating cell division and cell growth. *Plant Mol. Biol.* **56**, 171–84.

He, J.-X., Gendron, J. M., Sun, Y., Gampala, S. S. L., Gendron, N., Sun, C. Q. and Wang, Z.-Y. (2005). BZR1 is a transcriptional repressor with dual roles in brassinosteroid homeostasis and growth responses. *Science* **307**, 1634–8.

Hill, K., Wang, H. and Perry, S. E. (2008). A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. *Plant J.* **53**, 172–85.

Hughes, R., Spielman, M., Schruff, M. C., Larson, T. R., Graham, I. A. and Scott, R. J. (2008). Yield assessment of integument-led seed growth following targeted repair of auxin response factor 2. *Plant Biotechnol. J.* **6**, 758–69.

Jiang, W.-B. and Lin, W.-H. (2013). Brassinosteroid functions in Arabidopsis seed development. *Plant Signal. Behav.* **8**, 1–5.

Jiang, W.-B., Huang, H.-Y., Hu, Y.-W., Zhu, S.-W., Wang, Z.-Y. and Lin, W.-H. (2013). Brassinosteroid regulates seed size and shape in Arabidopsis. *Plant Physiol.* **162**, 1965–77.

Jofuku, K. D., Omidyar, P. K., Gee, Z. and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3117–22.

Kesavan, M., Song, J. T. and Seo, H. S. (2013). Seed size: a priority trait in cereal crops. *Physiol. Plant.* **147**, 113–20.

Kim, T.-W., Guan, S., Sun, Y., Deng, Z., Tang, W., Shang, J.-X., Sun, Y., Burlingame, A. L. and Wang, Z.-Y. (2009). Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors. *Nat. Cell Biol.* **11**, 1254–60.

Krishna, P. (2003). Brassinosteroid-Mediated Stress Responses. *J. Plant Growth Regul.* **22**, 289–297.

Krizek, B. A. (2009). Making bigger plants: key regulators of final organ size. *Curr. Opin. Plant Biol.* **12**, 17–22.

Krizek, B. A. and Anderson, J. T. (2013). Control of flower size. *J. Exp. Bot.* **64**, 1427–37.

Lauria, M. and Rossi, V. (2011). Epigenetic control of gene regulation in plants. *Biochim. Biophys. Acta* **1809**, 369–78.

Li, N. and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* **5**, 332.

Li, H., Johnson, P., Stepanova, A., Alonso, J. M. and Ecker, J. R. (2004). Convergence of signaling pathways in the control of differential cell growth in Arabidopsis. *Dev. Cell* **7**, 193–204.

Lin, X., Tirichine, L. and Bowler, C. (2012). Protocol: Chromatin immunoprecipitation (ChIP) methodology to investigate histone modifications in two model diatom species. *Plant Methods* **8**, 48.

Lord, J. M. and Westoby, M. (2012). Accessory costs of seed production and the evolution of angiosperms. *Evolution* **66**, 200–10.

Matias-Hernandez, L., Battaglia, R., Galbiati, F., Rubes, M., Eichenberger, C., Grossniklaus, U., Kater, M. M. and Colombo, L. (2010). VERDANDI is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. *Plant Cell* **22**, 1702–15.

Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R. and Colombo, L. (2011). The MADS box genes SEEDSTICK and ARABIDOPSIS B(sister) play a maternal role in fertilization and seed development. *Plant J.* 1–12.

Moles, A. T., Ackerly, D. D., Webb, C. O., Tweddle, J. C., Dickie, J. B. and Westoby, M. (2005). A brief history of seed size. *Science* **307**, 576–80.

Moreno-Risueno, M. A., Van Norman, J. M., Moreno, A., Zhang, J., Ahnert, S. E. and Benfey, P. N. (2010). Oscillating gene expression determines competence for periodic Arabidopsis root branching. *Science* **329**, 1306–11.

Müssig, C., Fischer, S. and Altmann, T. (2002). Brassinosteroid-regulated gene expression. *Plant Physiol.* **129**, 1241–51.

Nemhauser, J. L., Mockler, T. C. and Chory, J. (2004). Interdependency of brassinosteroid and auxin signaling in Arabidopsis. *PLoS Biol.* **2**, E258.

Nowack, M. K., Ungru, A., Bjerkan, K. N., Grini, P. E. and Schnittger, A. (2010). Reproductive cross-talk: seed development in flowering plants. *Biochem. Soc. Trans.* **38**, 604–12.

Nurrish, S. and Treisman, R. (1995). DNA binding specificity determinants in MADS-box transcription factors. *Mol. Cell. Biol.* **15**, 4076–4085.

Ohto, M.-A., Fischer, R. L., Goldberg, R. B., Nakamura, K. and Harada, J. J. (2005). Control of seed mass by APETALA2. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3123–8.

Ohto, M., Floyd, S. K., Fischer, R. L., Goldberg, R. B. and Harada, J. J. (2009). Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. *Sex. Plant Reprod.* **22**, 277–89.

Okushima, Y., Mitina, I., Quach, H. L. and Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. *Plant J.* **43**, 29–46.

Orsi, C. H. and Tanksley, S. D. (2009). Natural variation in an ABC transporter gene associated with seed size evolution in tomato species. *PLoS Genet.* **5**, e1000347.

Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E. and Yanofsky, M. F. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85–8.

Sasse, J. M. (2003). Physiological Actions of Brassinosteroids: An Update. *J. Plant Growth Regul.* **22**, 276–288.

Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N. and Scott, R. J. (2006). The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs. *Development* **133**, 251–61.

Song, C.-P., Agarwal, M., Ohta, M., Guo, Y., Halfter, U., Wang, P. and Zhu, J.-K. (2005). Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* **17**, 2384–96.

Sozzani, R., Maggio, C., Varotto, S., Canova, S., Bergounioux, C., Albani, D. and Cella, R. (2006). Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development. *Plant Physiol.* **140**, 1355–66.

Sun, X., Shantharaj, D., Kang, X. and Ni, M. (2010a). Transcriptional and hormonal signaling control of Arabidopsis seed development. *Curr. Opin. Plant Biol.* **13**, 611–20.

Sun, Y., Fan, X.-Y., Cao, D.-M., Tang, W., He, K., Zhu, J.-Y., He, J.-X., Bai, M.-Y., Zhu, S., Oh, E., et al. (2010b). Integration of brassinosteroid

signal transduction with the transcription network for plant growth regulation in Arabidopsis. *Dev. Cell* **19**, 765–77.

Sundaresan, V. (2005). Control of seed size in plants. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17887–8.

Tanabata, T., Shibaya, T., Hori, K., Ebana, K. and Yano, M. (2012). SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. *Plant Physiol.* **160**, 1871–80.

Tanabe, S., Ashikari, M., Fujioka, S., Takatsuto, S., Yoshida, S., Yano, M., Yoshimura, A., Kitano, H., Matsuoka, M., Fujisawa, Y., et al. (2005). A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length. *Plant Cell* **17**, 776–90.

Ulmasov, T., Hagen, G. and Guilfoyle, T. J. (1999). Activation and repression of transcription by auxin-response factors. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5844–9.

Van Daele, I., Gonzalez, N., Vercauteren, I., de Smet, L., Inzé, D., Roldán-Ruiz, I. and Vuylsteke, M. (2012). A comparative study of seed yield parameters in Arabidopsis thaliana mutants and transgenics. *Plant Biotechnol. J.* 488–500.

Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. and Chory, J. (2005). Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21**, 177–201.

Vert, G., Walcher, C. L., Chory, J. and Nemhauser, J. L. (2008a). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9829–34.

Vert, G., Walcher, C. L., Chory, J. and Nemhauser, J. L. (2008b). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9829–34.

Verwoerd, T. C., Dekker, B. M. and Hoekema, A. (1989). A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Res.* **17**, 2362.

Vlieghe, K., Boudolf, V., Beemster, G. T. S., Maes, S., Magyar, Z., Atanassova, A., de Almeida Engler, J., De Groodt, R., Inzé, D. and De Veylder, L. (2005). The DP-E2F-like gene DEL1 controls the endocycle in Arabidopsis thaliana. *Curr. Biol.* **15**, 59–63.

Western, T. L., Skinner, D. J. and Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant Physiol.* **122**, 345–56.

Western, T. L., Burn, J., Tan, W. L., Skinner, D. J., Martin-McCaffrey, L., Moffatt, B. A. and Haughn, G. W. (2001). Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in Arabidopsis. *Plant Physiol.* **127**, 998–1011.

Windsor, J. B., Symonds, V. V, Mendenhall, J. and Lloyd, A. M. (2000). Arabidopsis seed coat development: morphological differentiation of the outer integument. *Plant J.* **22**, 483–93.

Wu, C., Trieu, A., Radhakrishnan, P., Kwok, S. F., Harris, S., Zhang, K., Wang, J., Wan, J., Zhai, H., Takatsuto, S., et al. (2008). Brassinosteroids regulate grain filling in rice. *Plant Cell* **20**, 2130–45.

Yin, Y., Wang, Z. Y., Mora-Garcia, S., Li, J., Yoshida, S., Asami, T. and Chory, J. (2002). BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109**, 181–91.

Ringraziamenti

Questa tesi è il risultato di un percorso di quattro anni, durante i quali la mia vita è cambiata molto, sicuramente più di quanto io potessi immaginare.

E’ stata un’esperienza formativa ricca, che mi ha rivelato molti aspetti del lavoro di scienziato, primo fra tutti che non bisogna mai smettere di essere curiosi, desiderosi di imparare e di trasmettere questo entusiasmo. E’ una lezione che credo valga la pena portarsi dietro, qualunque sia la strada scelta.

Ringrazio Lucia Colombo per avermi offerto l’opportunità di lavorare nel suo gruppo di ricerca, e faccio lo stesso con Simona Masiero, fonte inesauribile di confronto e consigli, non necessariamente lavorativi.

Ringrazio i colleghi con cui ho collaborato a Milano e all’estero, troppi da elencare senza dimenticarmi di nessuno. Faccio un’eccezione per Goyo, con il quale ho condiviso dall’inizio la ricerca sui semi, e Nacho, interista vero e magnifica persona.

Dal dottorato porterò con me anche sincere amicizie. Fra, sono sempre pronto a ascoltare i tuoi numerosi progetti di vita, non importa quanto folli. Mara, dovunque sarò nel mondo farò riferimento a te per informarmi sui pettegolezzi più scottanti. Chiara, non credere di essere quella seria fra noi solo perché fai la mamma e sbucci la frutta, perché fra le altre cose poi non la mangi nemmeno. Michela ed Elia, anche se alla fine non ho capito che piastrelle abbiamo scelto.

Ci sarebbe poi da ringraziare Francesco, ci ho pensato, ma davvero non saprei come farlo. Tuorn’! ‘sta casa aspett a tte!

Dedico infine questo lavoro ai miei genitori, Margherita e Giancarlo, che mi hanno sempre sostenuto e che non hanno mai smesso di essere fieri di me. Lo dedico a Valentina, la mia compagna di vita, alla quale va il merito di avermi aiutato sempre a crederci. “Due buoni compagni di viaggio non dovrebbero lasciarsi mai”.