Characterization of T-DNA insertion lines for *FLA1*, *FLA4* and *FLA8* during *Arabidopsis thaliana* sexual reproduction

Márcio Filipe Moreira do Couto
Biologia Funcional e Biotecnologia de Plantas
Departamento de Biologia
2018

Orientador
Isabel Amorim, Professor auxiliar, FCUP

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto,
_____/_____/____
_____

Acknowledgements

Fica difícil para mim decidir como começar este texto, tendo chegado agora a hora de agradecer a todos aqueles que fizeram parte do meu percurso ao longo destes últimos 5 anos. Devo começar pelos meus amigos? Agradecer a professores pelas oportunidades? Agradecer ao FCP pelas alegrias? Não sei mesmo…

Acho que vou começar pelo início e pelas pessoas que sempre se sacrificaram por mim e me permitiram ter estes 5 anos de estudo. Tudo o que tenho, devo-o aos meus pais e avós. Tudo aquilo que sou aprendi com eles, as coisas boas e as menos boas, e uma grande parte do apoio que tive sempre veio deles. Apoio incondicional que me fez subir até este patamar e que espero ter honrado. O resto da minha família não fica atrás, a todos os meus primos e tios (que não vou enumerar pois tenho a certeza que a lista de nomes iria começar a desformatar o ficheiro que tenho em mãos) tenho de agradecer pelos bons momentos passados e pelos inúmeros convívios em família, desde casamentos recheados de danças “exóticas” e tesourinhos deprimentes que nunca vou esquecer, aos momentos de perda em que todos estivemos lá uns para os outros para ajudar a aguentar e superar. Não há necessidade de referir nomes, pois sempre formamos 1 família. Que venham muitos mais momentos bons e que nos maus continuemos unidos. E claro sem esquecer a minha irmã, que com o seu péssimo gosto para música e tendência para deixar as redes sociais com os seus dados ligados, me mostrou o quão sortudo fui quando os meus pais me apresentaram à nova inquilina da nossa casa. O mais recente inquilino também não podia faltar nesta declaração de amor e agradecimento, pois sem cão um homem nunca teria um melhor amigo. Ao nosso Leo agradeço por me mostrar que o amor incondicional dado por um cão é uma das melhores coisas que podemos ter na nossa vida.

Saindo agora da família passo a falar dos bons amigos que me acompanharam ao longo desta mão cheia de anos. Ao trio Odemira V2.0 da FCUP, Clécio Patrício, Inês Pinto e Maria João (ordem alfabética para não criar ilusões de favoritismo) tenho de agradecer pelos ótimos momentos passados ao longo destes anos. Mesmo nos momentos em que meio mundo de distância nos separava mantivemos sempre ótimas conversas científicas sobre temas adequados à nossa extensa área de estudo, como por exemplo Dona Branca, Maria Leal, Gato Fedorento, India Malhoa, Casa do Toy, entre muitos outros que a minha mente preenchida não tem agora capacidade de debitar.

Mantendo-me no tema de amizades tenho também de agradecer aos meus grandes amigos de Fiães City e arredores. Ao Daniel, Filipa, Andreia, Xana, Pedro (já estavas incluído acima nos primos, mas não custa nada mencionar) tenho de agradecer por todas as nossas saídas e por todos os bons momentos que passamos e certamente passaremos no futuro. E claro, pelo fluxo de memes e boa disposição que continuamos a perpetuar. Ao "Sôtor" Costa tenho de agradecer por me apresentar ao grande guru da minha vida, o Professor Mamadu Djalo e ao seu conhecimento infinito sobre todas as coisas no cosmos, e por se ter mantido sempre por perto, mesmo estando longe.

Ao pessoal da FCUP não podia faltar um destaque especial ao Nuno Ponte e à Sara Pinto que me mostrarem o patamar de excelência de estudantes de mestrado e de "mini-orientadores" que a FCUP tem para oferecer. Não tenho dúvidas que serão bem-sucedidos em todas as atividades em que se envolverem.

Para finalizar tenho ainda de agradecer à Professora Isabel Amorim por me ter convidado a associar-me a este projeto, à Professora Sílvia Coimbra pela oportunidade de conhecer, aprender e receber o apoio de todos os membros do grupo do laboratório e à Professora Paula Melo pelo apoio prestado ao longo deste período.

Uma nota final de apreciação a todos os meus colegas do Mestrado em Biologia Funcional e Biotecnologia de Plantas que vão agora terminar estes dois anos de mestrado e um desejo de boa sorte para todos aqueles que estão agora a iniciar este percurso.

Este trabalho foi financiado pela Fundação para a Ciência e a Tecnologia (FCT) através do projeto FCT/ PTDC/AGR-FOR/3356/2014.

Resumo

Este projeto tem como objetivo a compreensão do papel das proteínas arabinogalactânicas semelhantes a fasciclina (FLAs) nos processos de reprodução em *Arabidopsis thaliana*. As FLAs são uma sub-família de proteínas arabinogalactânicas (AGPs) quiméricas que possuem 1 a 2 domínios semelhantes a fasciclina (FAS). Estes domínios proteicos estão relacionados com propriedades de adesão celular e estão presentes em animais, leveduras, bactérias, algas, líquenes e plantas superiores. As proteínas FLA1, FLA4 e FLA8 foram encontradas, com base em análises bioinformáticas, nos tecidos reprodutivos de *A. thaliana*. Resultados publicados anteriormente indicam que a FLA4 pode ser importante no desenvolvimento da semente. Os papéis de FLA1 e FLA8 permanecem praticamente desconhecidos, especialmente no que diz respeito à reprodução. *FLA1*, *FLA4* e *FLA8* foram estudados usando uma abordagem com múltiplos passos: a região promotora foi analisada para investigar que tipos de fatores de transcrição poderiam potencialmente ligar-se a estes (usando ferramentas de bioinformática); o conjunto de sementes de linhas mutantes individuais para cada gene foram avaliadas fenotipicamente; a taxa de germinação e a morfologia das sementes foram também comparadas com as plantas Col-0, do tipo selvagem, e foram criadas linhas marcadoras para cada gene.

Os resultados obtidos mostraram que a *FLA8* pode ter um papel importante na formação de síliquas e na maturação das sementes. As plantas *fla1* revelaram defeitos na progressão da maturação das sementes, sugerindo que a *FLA1* pode ser importante durante esses processos. O *knock-out* da expressão de *FLA4* não revelou defeitos no desenvolvimento de sementes nas síliquas do mutante *fla4*. Nenhum dos mutantes apresentou perturbações significativas na taxa de germinação e morfologia das sementes maduras.

Como conclusão deste estudo, propomos um possível papel destes genes na rede de controle da reprodução das plantas e sua relação com os fatores de transcrição relacionados com reprodução em plantas.

Palavras-chave: *Arabidopsis*, proteínas arabinogalactânicas semelhantes a fasciclina, reprodução nas plantas, caracterização fenotípica, desenvolvimento de semente, controlo da transcrição

Abstract

This project focuses on understanding the role of fasciclin-like arabinogalactan proteins (FLAs) in the processes required for reproduction in *Arabidopsis thaliana.* FLAs are a sub-family of chimeric arabinogalactan proteins (AGPs) that possess variable numbers of 1 to 2 fasciclin (FAS) domains. These domains have been previously linked to cell adhesion properties and appear in animals, yeast, bacteria, algae, lichens, and higher plants. FLA1, FLA4 and FLA8 were revealed by a bioinformatics analysis to be present in reproductive tissues. Previously published results indicate that FLA4 was shown to be important in seed development. The roles of FLA1 and FLA8 remain mostly unknown, especially in regard to reproduction*. FLA1*, *FLA4* and *FLA8* were studied using a multistep approach: their promoter region was analysed to uncover the types of transcription factors that can potentially bind to them (using bioinformatic tools); the seed set of single mutant lines for each gene was evaluated; the germination rate and seed morphology were also compared to those of wild-type Col-0 plants and marker lines were created for each gene.

The results obtained showed that *FLA8* may have an important role in silique formation and seed maturation. The *fla1* plants revealed defects in the progression of seed maturation, suggesting that *FLA1* may be important during these processes. The knock-out of *FLA4* expression did not reveal defects in the seed set of *fla4* siliques. None of the mutants showed significant disturbances in the germination rate and mature seed morphology.

As a conclusion to this study, we would like to propose a possible role for these genes on the network controlling plant reproduction and their relation to transcription factors related to plant reproduction.

Key words: *Arabidopsis*, fasciclin-like arabinogalactan proteins, plant reproduction, phenotypical characterization, seed development, transcriptional control

Index

List of figures and tables

List of abbreviations

AG – AGAMOUS

AGL13 – AGAMOUS LIKE13

AGP – Arabinogalactan protein

AP – APETALA1

ChIP-seq – Chromatin Immunoprecipitation sequencing

Col-0 – Columbia 0 variety

DNA – Deoxyribonucleic acid

FLA – Fasciclin-like arabinogalactan protein

FAS – Fasciclin- like domain

gDNA – Genomic deoxyribonucleic acid

GFP – Green fluorescent protein

GPI – Glycosylphosphatidylinositol anchor

GUS – β-glucuronidase gene

HRGP – Hydroxyproline (Hyp)-rich O-glycoprotein

LB – Luria-Bertani

LRR-RLK – leucine-rich repeat receptor-like kinase

O.N – Overnight

PCR – Polymerase chain reaction

qRT-PCR – Quantitative reverse-transcriptase polymerase chain reaction

RNA – Ribonucleic acid

RNA-seq – Ribonucleic acid sequencing

SEP – SEPALLATA

SHP – SHATERPROOF

STK – SEEDSTICK

TF – Transcription factor

TTS – Transcription start site

UTR – Untranslated region

WT – Wild type

1. Introduction

1.1. The reproduction of flowering plants

The mechanisms of pollination and fertilization have been observed since the beginning of agricultural practices, thousands of years ago. The first recorded evidence of human activity related to speculating on the processes behind seed production is attributed to a ritual, performed more than 5000 years ago, by Assyrian priests consisting of dusting pollen from the male date palms on the inflorescences of female trees (Boavida *et al*., 2005). Since then, the emergence of Modern Science provided the tools for countless studies regarding the intricacies that underlie phenomena such as the formation of the pollen tube and its passage down the style into the ovule (originally described by Amici, 1824, 1830, 1847) and the occurrence of double fertilization in angiosperms, firstly described by Nawaschin (1898) and Guignard (1899).

The life cycle of flowering plants is divided into two phases, a dominant diploid sporophyte phase and a transient haploid gametophyte phase (Fan *et al*., 2008). The evolutionary success of higher plants heavily relies on this short gametophytic phase, which underlies the sexual reproduction cycle (Boavida *et al*., 2005). The haploid generation of flowering plants develops within the diploid sporophytic tissues of the ovule, a characteristic that is exclusive of seed plants. The sexual reproduction cycle in higher plants is a complex process that can be divided into three key events: pollination, fertilization and embryogenesis (Zhang *et al*., 2014). Angiosperms display gametophyte heteromorphy, meaning that female and male gametophytes have distinct form and functions. The microgametophyte, the male gametophyte, is called pollen grain and, in most of higher plants, it consists of three cells (two of them are gametes produced through the mitotic division of a generative cell into two sperm nuclei, with the third cell being a vegetative cell). The megagametophyte, the female gametophyte, is called embryo sac. *Arabidopsis thaliana* displays a Polygonum-type embryo sac, possessing seven cells: two synergid cells, three antipodal cells, one binucleate central cell and one egg cell.

Since flowering plants do not produce motive sperm cells that reach the female gametophyte by themselves, the fertilization process requires a complex system of pollination, that includes the development of pollen tubes. The pollen grain lands on the stigma (the receptive surface of the carpel), if compatible germinates, and grows by an highly directional tip growth through the style, into the ovary. During this pathway, the

pollen tube rapidly transports the sperm cells while negotiating its way through the female tissues. Once the pollen tube reaches its destination, it discharges the two sperm cells into the embryo sac and double fertilization takes place (Franklin-Tong, 2010). One of the sperm cells fuses with the egg cell giving rise to the diploid embryo, while the other will fuse with the central cell, resulting in the formation of a triploid endosperm. A seed will, then, be formed, surrounded by the seed coat, which is originated from the maternal integuments of the ovule (Ehlers *et al*., 2016).

The processes leading to the formation of a healthy embryo are carefully controlled. The seed can be considered as a functional unit required for the protection and propagation of the offspring represented by the embryo. Vast networks controlled by master regulator genes act to ensure that key events in the life cycle of flowering plants such as fertilization and seed formation are not impaired. In *Arabidopsis*, ovule primordia arise from the placenta at stage 8 of flower development, and differentiation is completed at stage 13 (Smyth *et al*., 1990), when the embryo sac is mature and can be fertilized (Schneitz *et al*., 1995; Mizzotti *et al*., 2012).

1.2. Role of MADS-box genes in plant reproduction

MADS-box genes are a family of genes involved in a myriad of processes related to the development of plants. Their designation is derived from the initials of four loci: MCMI of *Saccharomyces cerevisiae*; AG of *Arabidopsis thaliana*; DEF of *Antirrhinum majus* and SRF of *Homo sapiens*. All these genes encode proteins containing the MADS-box domain, a conserved protein domain with approximately 58 amino acids. They encode transcription factors in all eukaryotic organisms studied thus far. Plant MADS-box proteins contain a DNA-binding domain, an intervening domain, a keratin-like domain and a C-terminal C-domain, thus plant MADS-box proteins are of the MIKC type (Saedler *et al*., 2001). The evolution of land plants is closely related to an increase in number and functional diversity of this type of proteins. This would eventually culminate in the appearance and development of higher flowering plants, where these genes are known to dominate the control of reproductive development from early to later stages (Kaufmann *et al*., 2005). MADS proteins commonly bind DNA sequence elements called 'CArG' boxes (generic consensus: $CC[A,T]_6GG$ or $CC[A,T]_7G$) as tetramers in which 2 transcription factors bind as a homo- or heterodimer that then bind to each other, forming the aforementioned tetramer.

One of the classical examples of the role these genes play in the regulation of the reproductive development of higher plants is the determination of the ABC model of flower development, originally compiled by Coen and associates, 1991. This model indicated that organ identity in each whorl is specified by a unique combination of three classes of homeotic genes that could act alone or in association with each other. Class A genes determine sepal formation. The combination of A and B trigger petal formation. The amalgamation of B and C specifies the formation of stamens and C alone controls the development of carpels. Further studies about this model determined that the ABC genes were required but not sufficient for the specification of floral organ identity. The original model was later improved upon with the addition of the D and E class of genes (Kaufmann *et al*., 2005). The A-class genes are *APETALA 1* (*AP1*) and *AP2*. The B-class genes are *APETALA 3* (*AP3*) and *PISTILLATA* (*PI*). *AGAMOUS* (*AG*) is the only C-class gene. The D-class genes are *SEEDSTICK* (*STK*), *SHATERPROOF 1* (*SHP1*) and *SHP2*. The E-class genes are *SEPALLATA 1* (*SEP1*), *SEP2*, *SEP3* and *SEP4*. *AP2* is the only one of these genes that does not encode a MADS-box transcription factor. The regulation of the genes that give shape to the floral organs, whose identity is controlled by the ABCDE model, occurs through the binding of the tetrameric complexes to two adjacent cis-regulatory CArG-boxes, creating a DNA loop that increases chromatin accessibility at an enhancer or promoter region (Yan *et al*., 2016). The tetramers responsible for the identity of each floral organ are represented in Figure 1.

Besides being a part of the complex necessary for the ovule identity, *SEEDSTICK* (*STK*) is also known to be a master regulator of fertilization and seed development (Mizzotti *et al*., 2012; Mizzotti *et al*., 2014). STK acts together with SHP1 and SHP2 in the determination of ovule identity. In the *stk shp1 shp2* triple mutant, integuments are converted into carpelloid structures and female gametophyte development is arrested just after megasporogenesis (Battaglia *et al*.,2008). It was also shown that STK and SEP3 together regulate the reproductive meristem transcription factor gene *VERDANDI* (*VDD*), which controls cell identity in the female gametophyte (Matias-Hernandez *et al*., 2010). STK was also shown to act on the seed coat by regulating several metabolic processes such as the accumulation of proanthocyanidins (a class of phenylpropanoid metabolites) in the innermost layer of the seed coat (Mizzotti *et al*., 2014). RNA sequencing (RNA-seq) data from *stk* mutants obtained by Mizzotti *et al*. (2014) and Chromatin Immunoprecipitation sequencing (ChIP-seq) data, performed for STK (unpublished data) have demonstrated that STK also affects the expression of

hydroxyproline (Hyp)-rich *O*-glycoproteins, a superfamily of proteins highly present in reproductive tissues.

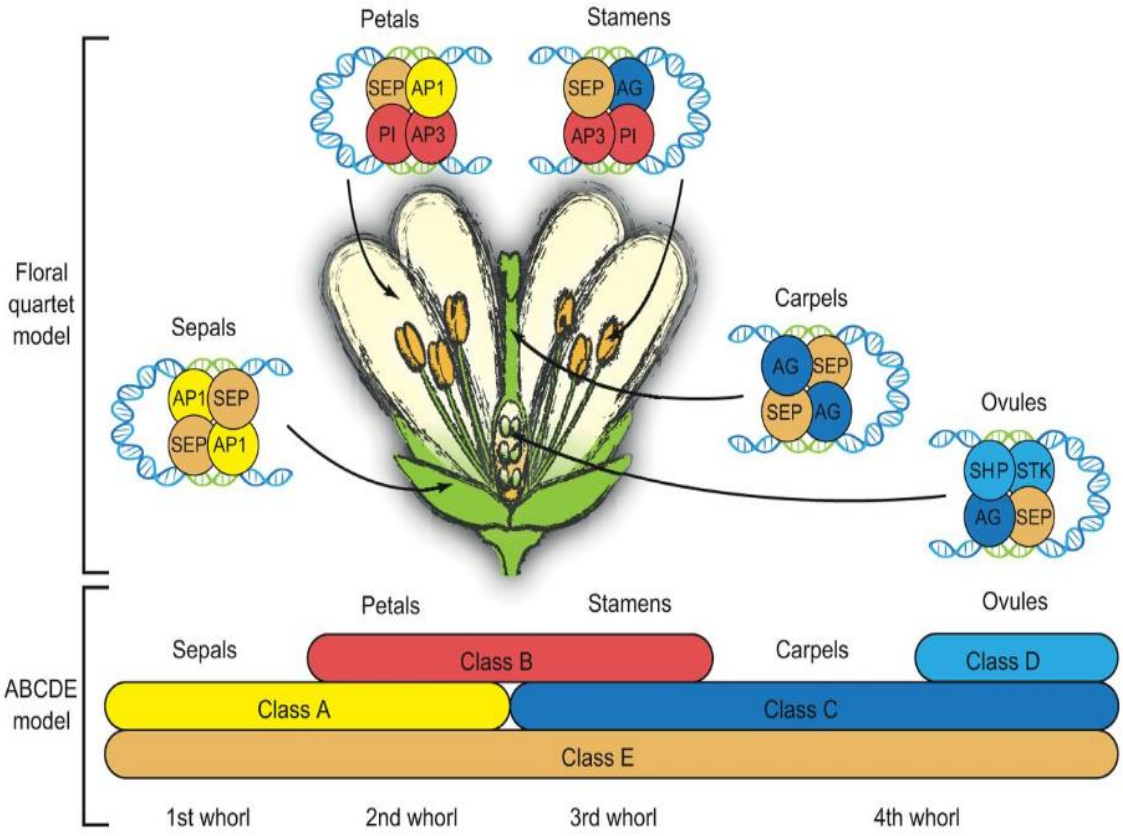


Figure 1. Representation of the ABCDE model of flower development. MADS-box proteins form tetrameric complexes that control the identity of each floral organ (sepals, petals, stamens, carpels and ovules). The formation of these tetramers promotes the binding of this complex to two adjacent *cis*-regulatory DNA binding sites (CArG-boxes, represented in green) creating a DNA loop. The bottom part of the figure illustrates the five sets of homeotic genes with overlapping functions in flowers of *A. thaliana* :A, B, C, D and E. Class A genes are expressed in the organ primordia of the 1st and the 2nd whorl of the flower, class B genes in the 2nd and 3rd whorl, class C genes in whorls 3 and 4, class D genes in parts of the 4th whorl (only in ovule primordia), and class E genes are expressed throughout all four whorls. Class A and E genes specify first whorl sepals. A, B and E specify second whorl petals, class B, C, and E genes specify third whorl stamens, class C and E genes specify fourth whorl carpels, and class C, D and E genes control the development of the ovules within the fourth whorl carpels. *AG – AGAMOUS*; *AP1 – APETALA1*; *AP3 – APETALA3* ; *PI- PISTILLATA*; *SEP – SEPALLATA*; *SHP – SHATERPROOF*; *STK – SEEDSTICK* (From Theißen *et al.*, 2016).

1.3. HRPGs in plant development

One of the most interesting gene super-families associated with plant reproduction is the hydroxyproline (Hyp)-rich *O*-glycoproteins family also known as HRGPs. These proteins are a part of the cell wall and represent a complex aggregate of macromolecules with

diverse structures and functions. Identified several decades ago, these proteins were classified into three categories: moderately glycosylated extensins, hyperglycosylated arabinogalactan proteins (AGPs) and Hyp/Pro-rich proteins that may be non-, weakly- or highly-glycosylated (Hijazi *et al*., 2014). Each HRGP sub-family is characterized by repetitive consensus sequences, which determine the way they are glycosylated according to the so-called Hyp-O-glycosylation code (Kieliszewski, 2001; Hijazi *et al*., 2014). Functionally, these proteins are very diverse: AGPs are implicated in a variety of physiological processes including cell expansion, reproductive development, embryogenesis, signalling pathways, defensive purposes, root development, hormone responses and programmed cell death (Seifert and Roberts, 2007; Costa *et al*., 2015); extensins are described as structural proteins able to form covalent scaffolds (Velasquez *et al*., 2012; Hijazi *et al*., 2014) and Hyp/Pro-rich proteins seem to be involved in plant defence against different kinds of stress, both biotic and abiotic (Kishor *et al*., 2015).

1.4. The multiproteic AGP family

AGPs are some of the most intriguing cell wall proteins, not only due to the extensive list of roles they play in the development of a large variety of plant tissues, but also because they are undoubtedly one of the most complex families of macromolecules found in plants (perhaps matched only by the polyphenolic compounds – lignins, cutins and suberins - and pectins) (Ellis *et al*., 2010; Showalter *et al.,* 2010). This complexity derives mainly from the diversity of glycans decorating the protein backbone, the heterogeneity of their glycosylation, the diversity of protein backbones containing arabinogalactan glycomodules and the existence of diverse protein domains, creating chimeric versions of classical AGPs. These ubiquitous proteins are found on the plasma membrane, in the wall, in the apoplastic space, and in secretions like the stigma surface and wound exudates (Ellis *et al*., 2010). They have been found in detergent-resistant membranes in *A. thaliana*, suggesting their presence in lipid rafts (Borner *et al*., 2005). Key distinguishing features (with notable exceptions) of AGPs appear in their carbohydrate motifs, primarily *O*-linked to the Hyp residues of the protein backbone, that constitute 90% to 98% (w/w) of the molecule, usually as branched type II arabino-3,6-galactans. Their aminoacidic composition, that typically constitutes 1% to 10% of the total weight of the protein, is rich in Hyp/Pro, Ala, Ser, and Thr, with the dipeptide motifs Ala-Hyp, Ser-Hyp, Thr-Hyp, Val-Pro, Gly-Pro and Thr-Pro as distinguishing features, although Ser-(Hyp) (extensin) motifs can also be present. Another important distinguishable aspect is

the presence of a glycosylphosphatidylinositol (GPI) membrane anchor, predicted on most, but not all AGP protein backbones, based on the presence of a hydrophobic C-terminal domain in the protein. This anchor is responsible for the attachment of these proteins to the outer leaflet of the cell membrane. Most, but not all AGPs bind a class of synthetic chemical dyes, the Yariv reagents, in particular to the *β*-glucosyl Yariv reagent (Yariv *et al*., 1967), which has proven extremely useful in their detection, quantification, and precipitation from solution, a useful early step in their purification. Although the precise mechanism of Yariv binding is mostly unknown, it requires the presence of both carbohydrate and protein moieties and is highly variable in strength (Pettolino *et al*., 2006).

The vast diversity of features displayed by the AGP family led to the subdivision of this large family of HRGPs into the classical AGPs, the arabinogalactan peptides (AG peptides) and the chimeric AGPs. In the plant Kingdom, there are assorted chimeric AGPs from the most frequent chimeric subfamilies: Fasciclin-like (FLA); Phytocyanin-like(PAG); Xylogen-like (XYLP); Protein kinase-like (PK); Formin homology 2-like (FH); Glycosyl hydrolase-like (GH) and Pollen allergen (“Ole e I” type). Other least frequent chimeric AGPs are also found, such as Pectin methyl esterase inhibitor-like, Pectate lyase-like domains (PCL) and hydrolase-type esterase-like (SGNH) (Ma et al.2017).

Classical AGPs are characterized by the presence of an N-terminal signal peptide (removed from the core polypeptide during the maturation process) that targets it for secretion, a proline/hydroxyproline-rich core domain and a C-terminal GPI anchor addition signal sequence, which is cleaved to produce the mature protein. The other AGP sub-families are structurally similar to the classical AGPs (Seifert and Roberts ,2007; Showalter *et al*., 2010). Figure 2 represents the wattle blossom model, one of the models currently accepted for the structure of AGPs.

AGPs can be detected during the formation of both female and male sporocytes and subsequent gametophytes and are also present during the stages leading up to double fertilization. This information was obtained through techniques that rely on the specific recognition of certain carbohydrate epitopes by monoclonal antibodies such as JIM8, JIM13, JIM14, LM2, MAC207 and LM6 (Costa *et al.,* 2015, Lopes *et al*., 2016), and the binding of the previously mentioned Yariv reagent to AGPs. Moreover, the use of promoter analysis assays with the GFP and GUS reporter genes, in situ hybridization, microarray data analysis, RNAi transgenic lines and real-time-PCR were also essential techniques used for these discoveries. These studies revealed important roles of AGPs,

per example: AGP22 and AGP24 are important in megaspore mother cell and functional megaspore formation and development (Tucker *et al*., 2012; Tucker and Koltunow ,2014); AGP19, a Lys-rich AGP characterized by Yang *et al*. (2007), is expressed only in the female tissue (style, ovary walls, and siliques); AGP4/Jagger is essential for persistent synergid degeneration and polytubey block (Pereira *et al*., 2015). A p*AGP6:GFP* line demonstrated that *AGP6* was expressed as soon as the locules of the anthers began their development and the GFP fluorescence was restricted to pollen and pollen tubes (Coimbra *et al*., 2008). AGP11 was also shown to be restrictively expressed in pollen, and an *agp6 agp11*double mutant demonstrated arrested pollen grain development, reduction in pollen germination, reduced pollen tube growth and premature pollen germination (Coimbra *et al.,* 2009; Costa *et al.*, 2013). These interesting expression patterns and the various roles they seem to play in plant development, especially in regard to plant reproduction, made AGPs a very fascinating study subject (reviewed in Su S and Higashiyama T, 2018).

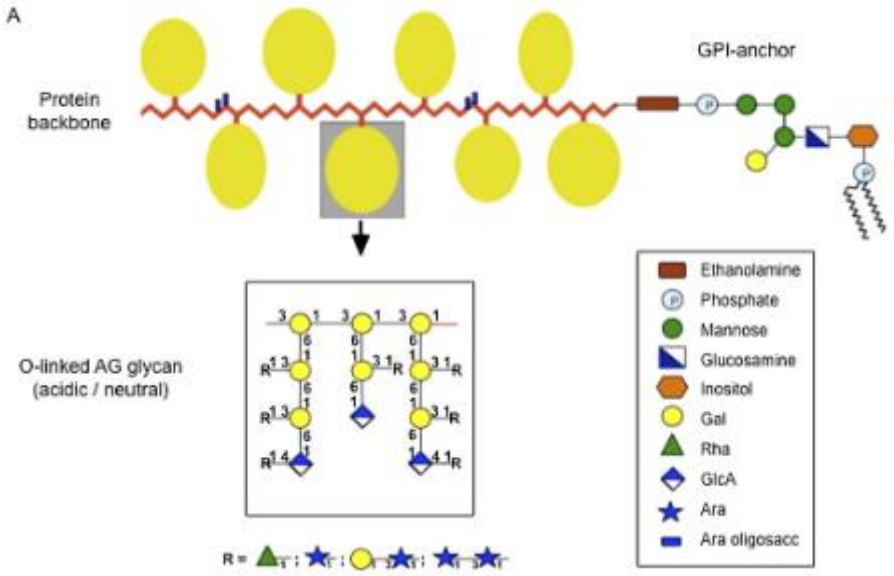


Figure 2. The wattle blossom model of the structure of AGPs with a GPI membrane anchor attached. In this model, there are approximately 25 Hyp residues. Most Hyp residues are non-contiguous and are predicted to bear an AG chain. Each AG chain may contain 15 or more repeats of a β-(1-3)-linked Gal oligosaccharide. (From Ellis *et al.*, 2010)

1.5. Fasciclin-like arabinogalactan proteins in plants

Proteins with variable numbers of fasciclin I (FASI) domains, generally one to four, were first identified in axon fascicles in the embryonic central nervous system of insects (Zinn *et al*., 1988).

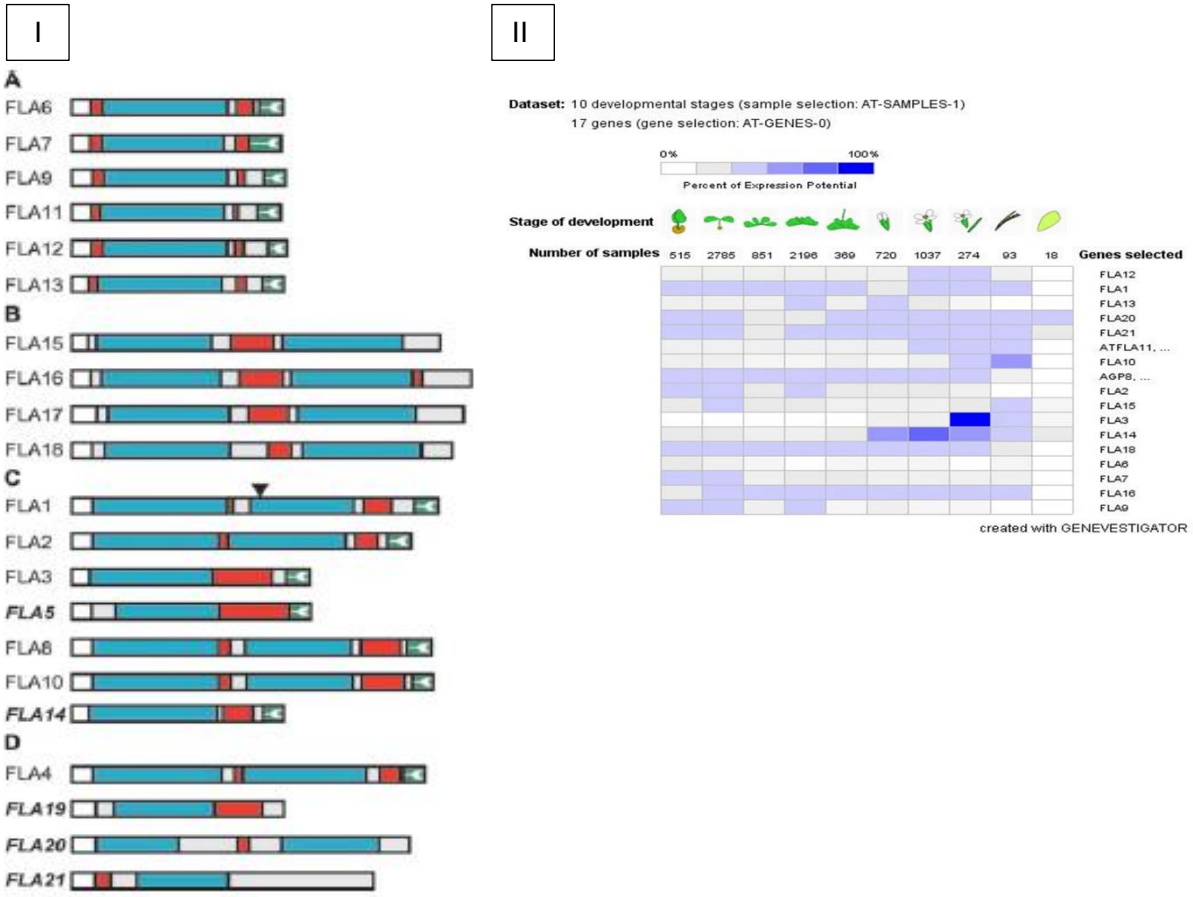


Figure 3. Schematic representation of the *A. thaliana* FLAs. I) The FLAs are grouped into four groups (A–D) based on phylogenetic analysis and pair-wise sequence comparison. The protein backbone of FLAs contains either one or two fasciclin-like domains (blue) and one or two AGP regions (red). Only the gene for FLA1 is predicted to have an intron (triangle), and where there are no ESTs, FLA names are bold and italicized. FLAs are predicted to contain an N-terminal secretion signal (white), and 14 of the 21 FLAs have a C-terminal signal for addition of a GPI anchor (green with white arrow). Additional protein regions are shown in light gray (From Johnson *et al.,* 2003). **II**) Differential expression of FLA genes during 10 development stages in *A. thaliana* (Data from Genevestigator).

Currently, it is known that these domains can also be found in proteins from animals, yeast, bacteria, algae, lichens, and higher plants. They are 110 to 150 amino acids long and have low sequence similarity (Johnson *et al*., 2003). Previous studies showed a relation between proteins containing these domains and cell adhesion properties (Kim *et al*., 2000).

Twenty-one FLA genes were annotated in the *A. thaliana* genome. This highly variable group of proteins showed more heterogeneity than classical AGPs, with proteins possessing one to two AGP domains, one to two fasciclin (FAS) domains and only fourteen out of the twenty-one possessing a signal sequence for the addition of a GPI anchor at the C-terminus of the protein. This variety of differences led to the division of this family into 4 groups based on a phylogenetic analysis. The pair-wise sequence comparison that led to the formation of these groups is represented in Figure 3 I. Figure

3 II shows a differential expression of FLA genes during 10 development stages in *A. thaliana*.

From the current literature it is known that *AtFLA11* and *AtFLA12* and their orthologues from other plant species, were implicated with stem mechanical properties and growth, acting either indirectly as regulators or in a direct structural role in secondary cell walls (MacMillan *et al*., 2010; Xue *et al*., 2017). *FLA3*, a FLA specifically expressed in male tissues, was confirmed to be essential for microspore formation, because FLA3 RNA-interference transgenic plants showed 50% less male fertility due to severe defects in the intine layer of the pollen grain (Li *et al*., 2010). Based on the transcriptome from flowers of monoecious *Quercus suber*, *A.thaliana* orthologs of *AGP/FLA*-like genes that are differentially expressed in male or female flowers were identified (Rocheta *et al*., 2014). In a phylogenetic analysis carried out with 34 FLAs amino acid sequences from *A. thaliana* and *Q. suber*, it was revealed that Arabidopsis orthologs of *QsFLA1*, *QsFLA8, Qs.FLA4, QsFLA3 and QsFLA14* belonged to a clade of genes that are differentially expressed in the *Q. suber* male and female gametophyte (unpublished data). Their *A. thaliana* orthologs also belonged to the same clade, but currently there is no literature relating these genes with the reproductive mechanisms (Supplemental figure 1). The roles of the FLA family of genes on flowering plants are still mainly unknown which was one of the main reasons for choosing 3 FLA genes - *FLA1* (AT5G55730), *FLA4* (AT3G46550) and *FLA8* (AT2G45470) - as study subjects of this project.

FLA1 was shown to play a role in shoot regeneration, in a study using a T-DNA insertion line (whose insertion was found to be in the intron of *FLA1*) (Johnson *et al*., 2011), but there is currently no data connecting this gene to other possible functions. Presently, there is no literature regarding *FLA8*, however both *FLA8* and *FLA1* have been revealed to be highly expressed in stigma, ovaries and seeds (data obtained from eFP Browser - http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). On the other hand, *FLA4* has been studied by various research groups. FLA4 was implicated with a multitude of genetic functions and pathways such as root elongation, salt stress tolerance, ABA and ACC signalling, and seed coat mucilage composition, possibly acting in a linear genetic pathway with the leucine-rich repeat receptor-like kinase loci (LRR-RLK) AtFEI1and AtFEI2 (Griffiths *et al*., 2016; Xue *et al*., 2017). The role of *FLA4* in plant reproduction is closely associated with the development of the seed coat, which points to a possible connection with *STK*, demonstrated to be a master regulator of the development of this structure by Mizzotti *et al*. (2014), but there is conflicting data about the way this happens. The general

consensus is that *FLA4*, alongside the mentioned *FEI1* and *FEI2*, regulates the production of the cellulosic rays deposited across the inner adherent layer of seed coat mucilage (Harpaz-Saad *et al*., 2012; Griffiths *et al*., 2016). It was also proved that two AGP-specific galactosyltransferases, GALT2 and GALT5, act in the same pathway as FLA4, FEI1 and FEI2 in processes involving seed coat formation and salt stress tolerance. *galt2galt5* double mutants displayed impaired root growth and root tip swelling in response to salt, likely because of decreased cellulose synthesis, as well as reduced seed mucilage adherence. These results were the same as the ones registered for *sos5*(*fla4*) and *fei2* plants, which prompted the development of the quintuple mutants *galt2galt5sos5fei1fei2* that in turn provided evidence that these genes act in a single, linear genetic pathway (Basu *et al*., 2016). Ascorbate deficiency was also shown to affect the expression of *FLA1*, *FLA2* and *FLA8* in the leaf cell wall glycoproteome of *A.thaliana* (Sultana *et al*., 2015).

In conclusion plants have developed intricate systems of control to assure a correct development of the seeds and the embryos they protect. These systems are studied by numerous research laboratories to understand and decipher the relations between the plethora of genes required for the success of sexual reproduction in model plants such as *A. thaliana*, and then apply that knowledge into crops as a way to tackle food shortage, one of the most concerning challenges of current times.

1.6. Objectives

Thomas Robert Malthus changed the scientific landscape when, in 1798, he published his most acclaimed thesis called "*An Essay on the Principle of Population*". In this essay, Malthus exposed his concerns about overpopulation and implied that population growth is potentially exponential while the growth of the food supply is linear, causing severe problems within societies that could lead to the destruction of the society itself. This theory was severely criticized for its pessimist look about the future of mankind, but it helped in the development of scientific fields that work in order to create solutions for the lack of food, and other resources, that humanity may face in the near future. Plants were seen as an important and viable solution for these problems and many scientific consortiums were established. Increasing seed yield was conjectured to be an important case study, since seeds are an important part of a person's daily diet, so the study of the processes underlying seed formation have been studied ever since. This project focuses on the study of 3 FLA genes: *FLA1, FLA4* and *FLA8.* They were chosen as subjects for this study because *AtFLA1* and *AtFLA8* are orthologs of *Q. suber* genes that were shown to belong to a sub-family of chimeric AGPs differentially expressed in male and female flowers of this species, while *FLA4* was shown previously to be important in the development of seed tissues such as the seed coat. The work here presented will focus on establishing a relation between FLA genes and plant reproduction, through the phenotypical characterization of homozygous mutant lines for each of them. Several questions arise from this analysis. Are they important in the reproduction of plants? What role do they play in their development, more specifically in the development of seeds?

Not many functions have been attributed to *FLA1, FLA4* or *FLA8*, so this work could establish new insights into the *FLA* family one of the sub-family of chimeric AGPs.

2. Materials and Methods

2.1. Plant material and growth conditions

All genotypes of *Arabidopsis thaliana* (L.) Heynh. plants used in this work were of Columbia (Col-0) background. WT, *fla1* (SALK_058964.45.15), *fla4* (SALK_125874.44.30) and *fla8* (SALK_141852) seeds were obtained from the Nottingham Arabidopsis Stock Center (NASC), United Kingdom. Seeds were sown on soil and grown in a growth chamber under continuous light, at ≈ 18˚C and 60% relative humidity.

2.2. *In silico* analysis of the genes and mutant lines

The location of the T-DNA insertion site for the three mutant lines was acquired from the SeqViewer Tool (https://seqviewer.arabidopsis.org/) available on TAIR (https://www.arabidopsis.org/) (Figure 4). A qualitative assessment of the putative expression pattern of *FLA1* and *FLA8* was performed using the eFP Brower (http://bar.utoronto.ca/efp/cgibin/efpWeb.cgi) platform. Data regarding the expression pattern of *FLA4* was not available on this database. An analysis of the promoter of each gene was performed using both AthaMap (http://www.athamap.de/) and PlantPan2.0 (http://plantpan2.itps.ncku.edu.tw/). The results from the later confirmed the previous results and did not add new relevant information.

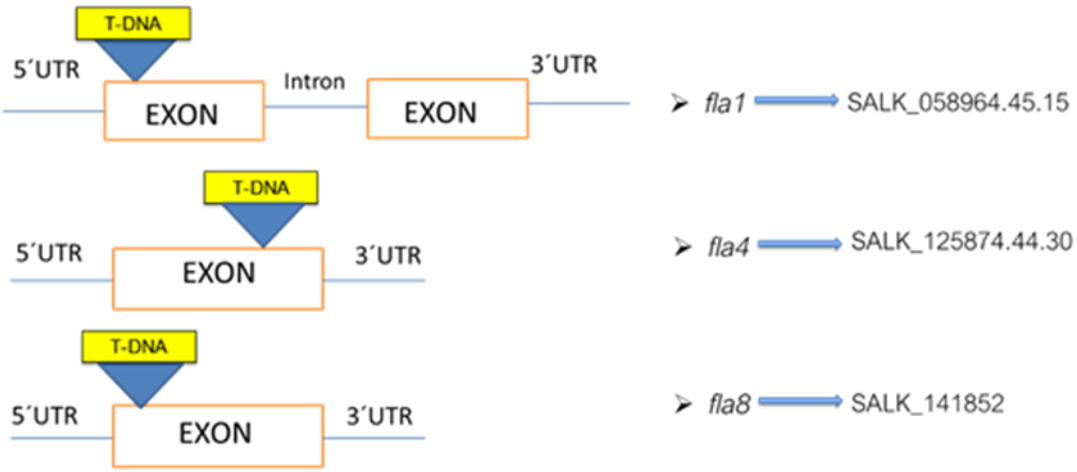


Figure 4. Graphical representation of the location of the T-DNA insertion in each mutant line. The T-DNA insertion for all mutant lines was shown to be located in the first exon of the corresponding gene. Each line was reported as having a *knock-out* of the corresponding gene by the supplier. 5´UTR- 5´untranslated region; 3´UTR- 3´untranslated region.

2.3. Genotyping

The genotypes of T-DNA insertion lines SALK_058964.45.15, SALK_125874.44.30 and SALK_141852 were confirmed by Polymerase Chain Reaction (PCR). The PCR was carried out using DreamTaq DNA Polymerase (Thermo Scientific). Two separate reactions were used for each mutant line as a way to identify homo- or heterozygous mutants. These reactions used two combinations of a set of three primers: Left primer (LP) and Right primer (RP) are primers that anneal to the forward or reverse strand of each gene, respectively, and a third primer called Border Primer (BP) that anneals in the forward strand of the T-DNA insertion border sequence. The reaction using LP and RP will amplify a section of the wild type genomic sequence of the gene, while a reaction using BP and RP will amplify a mutant fragment of the gene. The T-DNA insertion on each mutant line enlarges the size of the genomic sequence, which turns impossible an amplification using LP and RP to occur under the normal PCR extension conditions. By using BP instead of LP, it is possible to shorten the size of the amplification fragment and the mutant plants will be easily identified due to the size differences between the fragments obtained using the two combinations of primers. LBb1.3 is the BP usually used on genotyping SALK lines. The genotyping technique is shown in Figure 5. All primers used and the expected fragment size are shown in Table 1.

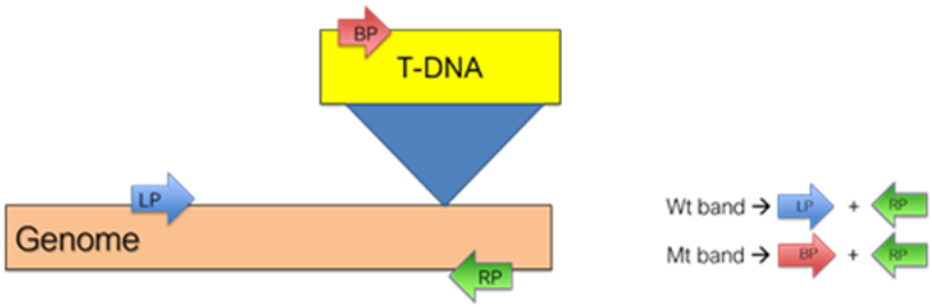


Figure 5. Scheme illustrating the genotyping technique. A wild type fragment is produced with an amplification between the left primer (LP) and the right primer (RP), which occurs when the gene is in its integral state. A mutant fragment would appear when the T-DNA insert is present. The border primer (BP) is then able to anneal and amplification with the RP can happen, forming a shorter band than the wt one.

2.3.1. Genomic DNA extraction – "Quick and Dirty"

DNA was extracted from a small leaf disk to isolate genomic DNA (gDNA) suitable for PCR amplification. The gDNA was extracted according to Edwards and associates (1991). The leaf disks were collected using eppendorf tubes and the vegetal material was macerated. 400µL of extraction buffer were added (composition is described in the

abovementioned article) and the solution was vortexed. The tubes were them centrifuged for 1 min at 12000g and 300µL of the supernatant were transferred to a new tube. 300µL of isopropyl alcohol were added and the solution was mixed and left 2 min at room temperature. Afterwards it was centrifuged for 5 min, at 12000g. The supernatant was discarded and the DNA sediment was allowed to dry and later on it was resuspended using 50µL of deionized water. The extracted gDNA was stored at 4ºC, prior to being used.

2.3.2. PCR reaction

The PCR was carried out using DreamTaq DNA Polymerase (Thermo Scientific).

Table 2 demonstrates the reagents and the quantities used in the PCR analysis. The primers used were: FLA1-LP+FLA1-RP for the amplification of FLA1, FLA4-LP+FLA4-RP for the amplification of *FLA4* and FLA8-LP+FLA8-RP for the amplification of *FLA8.* LBb1.3 was the BP primer used. The specific oligonucleotides and expected band sizes are listed on Table 1.

Table 1. List of all the primers used in PCR reactions. Adaptor sequences are underlined.

Name	Sequence (5´-3´)	Annealing temperature (˚C)	Product size (bp)	Application
FLA1-LP	GCATGTCTCACTTGTTTCATCC	58	FLA1-LP + FLA1-RP =1015	Genotyping *fla*1 (SALK_058 964)
FLA1-RP	TGAAGCTCCAGGGTTAGTGAG			
LBb1.3	ATTTTGCCGATTTCGGAAC		LBb1.3 + FLA1-RP ~800	
FLA4-LP	GAAACTGGGAATAACCTTCGG	58	FLA4-LP + FLA4-RP= 1051	Genotyping *fla*4 (SALK_125 874)
FLA4-RP	AGCTTCTCGAGACCAAACCTC			
LBb1.3	ATTTTGCCGATTTCGGAAC		LBb1.3 + FLA4-RP ~600	
FLA8-RP	CTTTGCCTCCTTTAAGATCGG	59	FLA8-LP + FLA84-RP= 1096	Genotyping *fla*8 (SALK_141 852)
FLA8-LP	ATGTAGAACATGAACGTCGGC			
LBb1.3	ATTTTGCCGATTTCGGAAC		LBb1.3 + FLA8-RP ~750	
pFLA1 ATTB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGAGTCCTGCGTCACCATTTT	59	pFLA1 ATTB1 + pFLA1 ATTB2 = 1293	pFLA1 amplification , Gateway
pFLA1 ATTB2	GGGGACCACTTTGTACAAGAAAGCTGGGTTGTTGGAGAGAGTGAAGAGTTTG			
pFLA4 ATTB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTCTTCATCGCCTCCTTGT	59	pFLA4 ATTB1+pFLA 4 ATTB2 = 726	pFLA4 amplification , Gateway
pFLA4 ATTB2	GGGGACCACTTTGTACAAGAAAGCTGGGTCGCCATTATTTTTTGAGTTAGG			
FLA1-RT FW	TCTCTCCCTCCACGTCCTTT	57	FLA1-RT FW + FLA1-RT RV = 185	RT-PCR fla1
FLA1-RT RV	ACAGATCTCCACCGTCAGGA			
FLA4-RT FW	CCTTCCAACGTTAACCTCCA	57	FLA4-RT FW + FLA4-RT RV = 213	RT-PCR fla4
FLA4-RT RV	CACACCCGAATTGATCGTTA			
FLA8-RT FW	CCTCTGCTCCACACTGACAC	57	FLA8-RT FW + FLA8-RT RV = 153	RT-PCR fla8
FLA8-RT RV	GGAGAATCAGCGAGGATTTG			
ACT8 FW	CTCAGGTATTGCAGACCGTATGAG	57	ACT8 FW+ ACT8 RV = 200	Pistil housekeeping gene
ACT8 RV	CTGGACCTGCTTCATCATACTCTG			

Table 2. PCR reaction mixture used with DreamTaq DNA Polymerase (Thermo Scientific).

Reagents	Quantity (µL)
10x Dream Taq Buffer	2.5
dNTP Mix (2 mM)	0.5
Forward Primer (20 nM)	1
Reverse Primer (20 nM)	1
Dream Taq DNA polymerase (5U/µL)	0.13
H_2O	18.9
DNA sample	1

A T3 Thermal Cycler (Biometra) was used and PCR conditions for each *FLA* mutant are presented on Table 3.

Table 3. PCR conditions used for each analysis with DreamTaq DNA Polymerase (Thermo Scientific).

		fla1	*fla4*	*fla8*
Step		Temperature / Time		
Initial denaturation		95ºC/ 2 min	95ºC/ 2 min	95ºC/ 2 min
35 cycles	Denaturation	95ºC/ 30 sec	95ºC/ 30 sec	95ºC/ 30 sec
	Annealing	58ºC/ 30 sec	58ºC/ 30 sec	59ºC/ 30 sec
	Extension	72ºC/ 1 min	72ºC/ 1 min	72ºC/ 1 min 10 sec
Final extension		72ºC/ 5 min	72ºC/ 5 min	72ºC/ 5 min

2.3.3. DNA gel Electrophoresis

All DNA fragment size observations were performed in 1% (w/v) agarose gel electrophoresis in 1X Sodium Borate (SB) buffer [47 g/L Boric Acid, 200 mM NaOH, pH ~8.2] and 0.5 mg/mL Ethidium Bromide was added before polymerization. 1X loading dye [10 mM Tris-HCl (pH 7.6), 0.15% orange G, 60% glycerol, 60 mM EDTA] was added to each sample prior to loading. GeneRuler DNA Ladder Mix (Thermo Scientific) was used a molecular weight marker. Fragment electrophoresis was performed in 1X SB running buffer and the electrophoretic separation was conducted at 200 V and non-limiting amperage. Ethidium bromide fluorescence allowed DNA visualization in a UV

transilluminator (302-365 nm) and images were acquired with a KODAK camera and analysed with the KODAK 1D Image Analysis software, version 3.5 (KODAK Scientific Imaging Systems).

2.4. Gene expression analysis

FLA1, *FLA4* and *FLA8* genes' expression level was evaluated as a way to confirm if the mutant lines possessed a complete *knock-out* of the corresponding gene, as was indicated by the supplier of these lines.

2.4.1. RNA extraction from Arabidopsis flowers

Complete inflorescences from *fla1*, *fla4*, *fla8* and WT plants were collected for RNA extraction. RNA extraction was performed using PureZOL™ RNA Isolation Reagent following the instructions manual (Bio-Rad). After each RNA extraction, all the samples were quantified and purity was verified using the µDropTM Plate in a Multiskan™ GO Microplate spectrophotometer, and results were analyzed with SkanIt™ software (Thermo Scientific). RNA samples were run in a RNAse free 1% (w/v) agarose gel (the electrophoresis tub had been thoroughly washed with oxygen peroxide and deionized water, prior to preparing the gel) in 1x SB buffer as a way to further evaluate the RNA integrity. RNA samples were stored at -75˚C before being used in a 1st strand cDNA synthesis reaction.

2.4.2. 1st strand cDNA synthesis

RNA samples were first treated with DNase I, RNase free (Thermo Scientific), according to the manufacturer's instructions, in order to remove any genomic DNA contamination. Ribolock RNase Inhibitor was applied prior to 1st strand cDNA synthesis to avoid RNA degradation. Reverse transcription was performed immediately after DNase treatment using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) and Oligo(dT)18 primers to initiate the reaction, according to the manufacturer's instructions (shown on Table 4).

Table 4. RT-PCR reaction mixture used with RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific).

Reagents	Quantity (µL)
RNA sample	1 µg
10x Reaction Buffer with $MgCl_2$	1 µL
DNase I, RNase free	1 µL (1 U)
Ribolock RNase Inhibitor (20 U/µL)	1µL
RevertAid M-MuLV RT (200 U/µL)	1 µL

2.4.3. Semi-quantitative RT-PCR

A semi-quantitative RT-PCR analysis was performed using the cDNA obtained previously from *fla1*, *fla4*, *fla8* and WT plants. Specific RT-PCR primers, shown in Table 1, were designed and used in a PCR reaction. The PCR conditions are shown in Table 5. *ACTIN8* (*ACT8*) was chosen as reference gene. WT cDNA was used to obtain the regular amplification fragment corresponding to FLA1, FLA4 and FLA8 cDNA. cDNA from the mutant plants was used to check if the insertion caused a complete *knock-out* of the genes. DreamTaq DNA Polymerase (Thermo Scientific) was used to obtain the amplification. The reagents and quantities used for PCR reaction were the same as shown previously in Table 2.

Table 5. PCR conditions used in semi-quantitative RT-PCR analysis of *FLA1*, *FLA4*, *FLA8* and *ACT8* expression in *fla1*, *fla4, fla8* and WT plants using DreamTaq DNA Polymerase (Thermo Scientific).

PCR Programme		
Step		Temperature / Time
Initial denaturation		95ºC/ 3 min
35 cycles	Denaturation	95ºC/ 30 sec
	Annealing	57ºC/ 30 sec
	Extension	72ºC/ 20 sec
Pause		8ºC / ∞

2.5. Marker lines production

p*FLA1:GUS* and p*FLA4:GUS* lines were designed to analyse the native expression pattern of *FLA1* and *FLA4*. p*FLA8:GFP* lines had already been designed in our laboratory, so the p*FLA8:GUS* line was not created. These lines were created using the Gateway® Gene Cloning technique (Thermo Fisher Scientific) (Hartley *et al.*, 2000).

2.5.1. Genomic DNA Extraction – CTAB Method

100 mg of WT plant leaf tissue were homogenized and 500 µL of CTAB Buffer (2% cetyl trimethylammonium bromide, 1% polyvinyl pyrrolidone, 100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA) were added. This solution was thoroughly mixed by vortexing and placed in a 65°C bath for 20 minutes with frequent inversions.

Following the incubation period, the homogenate was centrifuged for 5 minutes at 15500 g. The supernatant was removed to a new tube and 2.5 µL of RNase (10mg/mL) were added. The solution was then to incubated at 32°C for 20 min. After that, an equal volume of chloroform/isoamyl alcohol (24:1) was added and vortexed for 5 seconds. Then the sample was centrifuged for 1 minute at 15000 g to promote separation of the phases and the aqueous phase was transferred to a new tube. The DNA was precipitated by adding 0.7 volume of isopropanol, previously cooled to 4°C, and incubated at -20°C for 15 minutes. The gDNA was then stored at -20°C prior to being used.

2.5.2. Cloning the fragments of interest

The putative promoters of *FLA1* and *FLA4* were amplified from gDNA extracted from WT plants, as described in 2.5.1. Using primers designed to amplify the -720 to +6 region of the putative *FLA4* promotor and -1293 to -1 region of the putative *FLA1* promotor (distances relative to ATG, the translation initiation site). This amplification was made using DreamTaq DNA Polymerase (Thermo Scientific). The primers used for this amplification were: pFLA1ATTB1 and pFLA1ATTB2 for the pFLA1 amplification and pFLA4ATTB1 and pFLA4ATTB2 for the pFLA4 amplification (primer sequence specified in Table 1). The PCR conditions for the amplification are shown in Table 6.

Table 6. PCR conditions used for promoter amplification DreamTaq DNA Polymerase (Thermo Scientific).

	Step	p*FLA1* and p*FLA4*
Initial denaturation		95ºC/ 2 min
35 cycles	Denaturation	95ºC/ 30 sec
	Annealing	59ºC/ 30 sec
	Extension	72ºC/ 1 min 10 sec
Final extension		72ºC/ 5 min

The PCR products were separated on a 1% (w/v) agarose gel on 1x SB buffer and the desired fragments were excised from the agarose gel and purified using GeneJET Gel Extraction Kit (Thermo Scientific) according to the manufacturer's specifications. The extracted band was then diluted in 50µL H2O and stored at -20ºC. The length of the cloned sequences was decided based on the relative distances between the genes studied and their neighbouring genes, the intergenic regions, as shown in Supplemental Figure 2.

2.5.3. Electrocompetent *Escherichia coli* protocol

One colony of *E. coli* DH5α was inoculated in 5 mL of Luria-Bertani (LB) medium and was grown overnight (ON) at 37ºC, shaking at 180 rpm. 2mL of this culture were then transferred to 200 mL of LB and the bacteria were allowed to grow in the same conditions as previously mentioned, until they reached an optical density (OD) between 0,5 and 0,6 (2-3 hours of incubation). After reaching this OD value, the culture was placed in ice for 30 min. The 200 mL of bacteria culture were then divided in four portions of 50 mL each and centrifuged for 20 min at 4000 g, 4ºC. The supernatant was eliminated and the cell pellet was resuspended in 200 mL of cold H_2O. The bacteria were then centrifuged again, using the same parameters as before. The supernatant was disposed of and the cell pellet was resuspended in 100 mL of cold H_2O. The bacteria were centrifugated a final time for 20 min at 4000 g, 4ºC, and the cell pellet formed was resuspended in 2 mL of

10% glycerol previously placed in ice. The solution was then divided in 70 µL aliquots and frozen using liquid nitrogen. The bacteria were stored at -80ºC, until future use.

2.5.4. BP reaction and *E. coli* competent cells transformation

A BP reaction was performed according to the work of Hartley and associates (2000). This reaction was performed using 5 µL of the PCR product, for each gene, obtained previously, 2 µL of p*DONOR207*™ (Invitrogen) plasmid and 1 µL of H_2O. 2 µL of Gateway™ BP Clonase™ II Enzyme Mix (Thermo Fisher Scientific) were added and the mix was vortexed and left ON at 25ºC. 1 µL of proteinase K was added to inactivate the enzyme and the reaction was incubated at 37ºC for 10 minutes. Supplemental Figure 3 has a representation of p*DONOR207*.

A 70 µL shot of electrocompetent *E. coli* (DH5α strain) was used for transformation using p*DONOR207*+p*FLA1* and another was used for the transformation using p*DONOR207*+p*FLA4*. 2 µL of the BP reaction were added to the *E. coli* shot and gently mixed, after the bacteria thawed. The bacteria were then added to an electroporation cuvette and put in an electroporator. The *E. coli* electroporation program was used and immediately after electroporation occurred, 350 µL of LB medium were added to the cells. After being gently mixed, by pipetting up and down, the bacteria were transferred to a new eppendorf tube and kept for 45 min at 37ºC, without shaking. 100 µL of the bacteria were then plated in LB medium supplemented with agar 1,5% (w/v) and 15 µg/mL gentamycin. The bacteria were allowed to grow ON at 37ºC. A 5mL liquid culture of the colonies that grew was later prepared using LB with 15 µg/mL gentamycin and incubated ON at 37ºC, shaking at 180 rpm. The transformed plasmids were recovered using GeneJET Plasmid Miniprep Kit (Thermo Scientific). Bacteria stocks were created using glycerol 10% (v/v) and stored at -80ºC.

2.5.5. LR reaction and *Escherichia coli* competent cells transformation

After recovering the transformed plasmids (p*DONOR207*+p*FLA1* and p*DONOR207*+p*FLA4*) the LR reaction was made using Gateway™ LR Clonase™ II Enzyme Mix (Thermo Fisher Scientific), following the instructions of the provider. 3 µL of the transformed plasmids were mixed with 1 µL of a destination plasmid called p*BGWFS7* (Karimi *et al.*, 2002). 1 µL of Gateway™ LR Clonase™ II Enzyme Mix

(Thermo Fisher Scientific) was added and the mix was vortexed and left overnight (ON) at 25ºC. 1 µL of proteinase K was added to inactivate the enzyme and the solution was incubated for 10 minutes, at 37ºC.

From the previous reaction 2 µL were used for the transformation of a 70 µL shot of electrocompetent *E. coli* DH5α and the electroporation reaction occurred in the same conditions as the ones described in 2.5.2. 100 µL of the transformed *E. coli* were then grown on LB medium supplemented with agar 1,5% (w/v) and 50 µg/mL spectinomycin. The bacteria were allowed to grow ON at 37ºC and colony PCR was used to check for positive colonies, using the same set of primers that were used for the amplification of the promoter regions. The PCR conditions used in the colony PCR assessment are the same as displayed in Table 5, except for the initial denaturation step. The amount of time for denaturation was augmented to 5 minutes to disrupt the bacterial cell and free its DNA. Using a pipette tip, a bit of the chosen colonies was picked and diluted in 10 µL of H_2O prior to the PCR reaction. 1 µL of this dilution was used in the reaction as the template DNA source, and the rest of the reagents and the quantities used are the same as shown in Table 2. GeneJET Plasmid Miniprep Kit (Thermo Scientific) was later used to recover the transformed p*BGWFS7* plasmids (p*BGWFS7*+p*FLA1:GUS* and p*BGWFS7*+p*FLA4:GUS*) from 5 mL liquid cultures of the positive colonies (LB supplemented with 50 µg/mL spectinomycin), grown ON at 37ºC with a 180 rpm shaking. Supplemental Figure 4 reveals the graphical representation of vector p*BGWFS7* used for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*.

The plasmids were later sent for sequencing at STAB VIDA with the primers used for the extraction of the promoter fragment. All data received from the company was analysed using ApE Plasmid Program.

2.5.6. *Agrobacterium tumefaciens* transformation

The *Agrobacterium tumefaciens* strain used for transformation of *A. thaliana* was GV3101::p*MP90*. Electrocompetent *A.tumefaciens* GV3101::p*MP90* were graciously donated by Rómulo Sobral, Univesidade do Minho. The same guidelines for *E. coli* transformation were used, except for a few differences due to the dissimilar characteristics of each species. 2 µL of the transformed plasmids recovered from *E. coli* were added to a 70 µL shot of electrocompetent *A. tumefaciens* GV3101::p*MP90*, after the bacteria thawed. The bacteria were then added to an electroporation cuvette and put in an electroporator, where the Agro program was chosen (2.20 kV, 1 pulse).

Immediately after electroporation occurred, 350 µL LB medium were added to the cells. After being gently mixed, by pipetting up and down, the bacteria were transferred to a new eppendorf tube and kept for 4 hours at 28ºC, without shaking, to allow the bacteria to recover. 100 µL of the bacteria solution were then plated in LB medium supplemented with agar 1,5% (w/v) and 10 µg /mL rifampicin, 50 mg/mL spectinomycin and 30 µg/mL gentamycin. They were then allowed to grow for 2 days at 28ºC.

2.5.7. Floral Dip mediated transformation of *A. thaliana*

A. thaliana plants were transformed using the floral dip technique following the protocol designed by Clough and Bent (1988). Isolated colonies from *A. tumefaciens* GV3101::p*MP90* harbouring the transformed plasmids (p*BGWFS7*+p*FLA1:GUS* and p*BGWFS7*+p*FLA4:GUS*) were picked and inoculated in 5 mL of LB media supplemented with 10 µg /mL rifampicin, 50 mg/mL spectinomycin and 30 µg/mL gentamycin. Then, they were incubated for two days at 28ºC, with shaking at 180 rpm. The inoculum was then added to 250 mL of LB medium supplemented with the same antibiotics and incubated ON at 28ºC, with shaking at 180 rpm. Later, the culture was centrifuged for 20 minutes, at 1250 g, at room temperature and the cell pellet was resuspended in 250 mL H_2O with 5% (m/v) sucrose and 0.05% (v/v) Silwett-77 (reduces water surface tension so that the buds could be immersed). Siliques and opened flowers of WT plants were removed prior to transformation. The plants were dipped twice in the solution described above, for 2 minutes. They were later covered with a plastic bag and left in the greenhouse until the next day. The bag was then removed and the plants were allowed to grow. The seeds are still to be recovered and tested to see if transformed plants were obtained. Transformant lines are selected using BASTA (as a selection agent (transformed plants are granted higher resistance to this herbicide due to a gene present in the *pBGWFS7* vector). Seedlings are pulverized 3 times with a 25 µL/mL Basta solution, with a time interval of 2 days between pulverizations.

2.5.8. Histochemical detection of GUS (β-glucuronidase) activity

The fresh plant material – flowers and siliques – were collected and placed in tubes, properly identified, with about 700 µL of 90% acetone (fixative) for 2 hours at -20°C. The samples were washed 2 times with phosphate/NaPi [1:1] buffer for 5-10 minutes. Following that, the buffer was replaced with a solution of X-Gluc (5-bromo-4-chloro-3-indolyl glucuronide) and left in the incubator at 37°C, overnight. The next steps of this

technique were always performed protected from the light. The X-Gluc solution was removed and washed with 90% ethanol for 10 min, followed by 70% ethanol for 10 min. The ethanol was discarded and a solution of chloral hydrate was added and left at 4°C overnight. The results were observed using a ZEISS Axio lmager AZ microscope with differential interference contrast (DIC) optics and were photographed with a ZEISS Axiocam MRc3 camera, using the Zen Imaging acquisition software. This protocol was used to detect the expression pattern of *FLA14* using a previously transformed line designed by Ferreira MJ and Amorim MI (2017), unpublished. Meanwhile this protocol was not applied yet to the expression lines developed in this study.

2.6. Seed set analysis

The seed set of mutant plants was compared with the seed set of WT in order to observe any phenotypical differences caused by the insertion of T-DNA in single mutant lines for *FLA1*, *FLA4* and *FLA8.* In this analysis, a representative sample of 10 siliques was chosen from 2 months old plants, both from WT and mutant lines. 2 to 3 plants of each mutant line and WT were chosen for each seed set analysis and the analysis was repeated 3 different times throughout this study. The siliques were measured and, afterwards, dissected using tweezers as a way to reveal the development state of the seeds in its interior. These observations were made under a dissecting microscope. A Student's t-test analysis was used to determine if the differences in the amount of green seeds, white seeds, aborted seeds, aborted ovules, and silique length, between WT and the three mutant lines, are significant.

2.7. Germination assay

A germination assay was performed to check for defects in germination between mutant lines and WT. Approximately the equivalent to 10 µL of Arabidopsis seeds were put in a 1,5 mL eppendorf tube and 1 mL of 70 % ethanol was added. The tubes were vortexed for 3 minutes and the seeds were transferred to a previously sterilized filter paper inside of a flow hood. After the ethanol completely evaporated, the seeds were transferred to 0,7% agar plates with ½ Murashige and Skoog (MS) medium, pH=5,7 (adjusted using 1M KOH), using a sterilized toothpick. The plates were sealed using micropore tape and stored for 3 days at 4 ºC in obscurity. After vernalization the plates were placed in a growth chamber under continuous light, at ≈ 18˚C and 60% relative humidity. The

germination rate was measured after 4 days under these conditions. The emergence of the radicle from the seed was determined as the point where the seed was considered to have started the germination process. This analysis was performed twice, using seeds from the same batch. A Student's t-test analysis was used to determine if the differences observed in germination rate were significant.

2.8. Seed morphology assay

Mature seeds from homozygous *fla1*, *fla4* and *fla8* were observed under a dissecting microscope as a way to compare their morphology to the morphology of WT seeds. Three major phenotypical classes can be detected: round seeds (regular seeds, without defects), shrivelled seeds (phenotype resulting from failure in the separation between endosperm and embryo, restricting embryo expansion - described by Yang *et al*., 2007) and *flat* seeds (completely collapsed seeds). This analysis was performed twice using seeds from the same batch. A Student's t-test analysis was used to determine if the differences detected in seed morphology were significant.

3. Results

3.1. Bioinformatics analysis

3.1.1. Gene expression patterns in *Arabidopsis thaliana*

The putative expression patterns of *FLA1* and *FLA8* are available in the eFP Browser platform (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi), in which microarray data was compiled to study the expression pattern of *A.thaliana* proteins. Data regarding the expression of *FLA4* is absent from this platform, so the Klepikova Arabidopsis Atlas eFP Browser (Klepikova *et al*., 2016) tool was used to analyse this gene´s expression pattern. *FLA1* and *FLA8* have distinct expression patterns in *A. thaliana*. *FLA8* is moderately expressed in flowers from stages 9 to 12, in both petals and carpels, but in later stages the expression levels in petals are reduced and this gene appears only to be expressed in the carpels (floral stages determined by Smyth *et al*., 1990). This gene is expressed in all stages of rosette development and is also highly expressed in several types of root cells, ranging from meta and protophloem cells to xylem pole pericycle cells. *FLA8* is also expressed in reproductive tissues, where high levels of gene activity were recorded in the seed coat of seeds containing pre-globular and globular embryos and in the ovaries present in pistil tissues (Figure 6). Like *FLA8, FLA1* is also moderately expressed in flower from stages 9 to 12 in both petals and carpels, and also appears to be highly expressed in several types of root cells and in the pistils. The main differences between the two expression patterns are a severely diminished expression of *FLA1* in rosette leaves and the different seed development stages in which they appear. *FLA1* seems to be highly expressed in seeds containing torpedo stage embryos, appearing in the cotyledons of the embryos, in this developmental stage, and also in the peripheral endosperm of developed seeds (Figure 7). Supplemental Figure 5 shows the expression pattern of these 2 genes.

On the other hand, it is impossible to analyse the expression pattern of *FLA4* as thoroughly as the previous two, due to the limitations of the Klepikova Arabidopsis Atlas when compared to eFP Browser. This platform reveals that FLA4 is present in every stage of flower development and exists in all floral organs. Interestingly, the higher levels of expression in floral tissues were registered in the stamen filaments, while the anthers presented the lower levels. *FLA4* is also shown to be expressed in young seeds and in siliques. The expression levels of this gene were also measured during the germination

process, in which a gradual increase in expression was quantified as the seedling emerged from the seed.

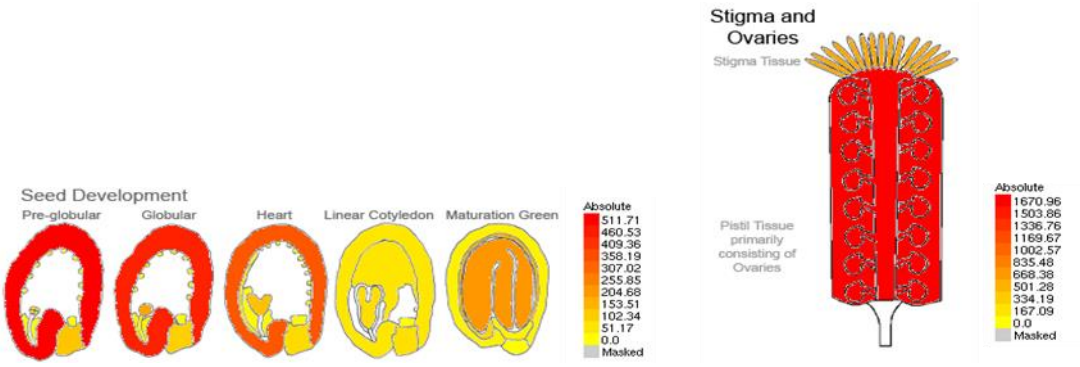


Figure 6. Expression of FLA8 in the stigma, ovaries and seed of *A. thaliana*. Adapted from eFP Browser

Figure 7. Expression of FLA1 in the stigma, ovaries and seed of *A. thaliana*. Adapted from eFP Browser

3.1.2. Promoter analysis for each gene

An analysis of the promoters of *FLA1*, *FLA4* and *FLA8* was performed in order to identify potential transcription factor (TF) binding sites in *Arabidopsis thaliana*. AthaMap was the main tool used in this task. Data available in AthaMap is based on published TF binding specificities available as alignment matrices or experimentally determined single binding sites. Therefore, this platform has a limited pool of TF and TF binding motifs that it can identify (221 TF and their binding motifs, to be exact). The PlantPan2.0 tool was then used to fill in for some of the gaps created by using AthaMap. The results from this analysis did not provide substantial data to add to the data provided with AthaMap, so they were not added to this report. The genomic sequence analysed was different for each gene due to their proximity to the other upstream genes, so that it would only include intergenic regions between adjacent genes and the first fifty bps after the transcription start site (TTS). The region analysed for *FLA1* was -1100 bp to +50 bp, -700 bp to +50 bp for the *FLA4* gene and -1200 bp to +50 bp for the region analysed for

the *FLA8* gene (all distances are relative to the TTS). Not all families of TF were selected for this study. Only TF with functions that were either predicted or shown to be important for plant reproduction and seed development were chosen. From the pool of TF and TF families of *A. thaliana* available on AthaMap, the TF chosen were from the MADS, MYB, C2C2(Zn) DOF (DNA-binding One Zinc Finger, belongs to the C2C2 zinc finger family), SBP (Squamosa promoter binding protein), ABI3/VP1 and SRS (SHI-related sequence) families. Tables 7, Table 8 and Table 9 represent TF that are predicted, by AthaMap, to be able to bind to the promoters of *FLA8*, *FLA1* and *FLA4*, respectively. There are some differences in both the type of TF that can bind to the promoters of each gene as well as the number of possible binding site identified in their promoter sequence. All the analysed promoter sequences are shown to possess binding sites for AGAMOUS-LIKE 1 (AGL1), AGL2, AGL3, AGL15, and MYB61. The promotor of *FLA8* possesses more binding sites for MADS-box proteins than any of the other two *FLA* genes and is the only promotor in which a binding site for AGAMOUS (AG) was detected. On the other hand, only the promoters of both *FLA1* and *FLA4* were shown to possess binding sites for SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 8 (SPL8) and LEAFY COTYLEDON 2 (LEC2). Binding sites for DOF AFFECTING GERMINATION 2 (DAG2) were marked in *FLA1* and *FLA8´*s promoter sequences. STYLISH 1 (STY1) was demonstrated to be able to bind to the regulatory region of *FLA1* and SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 3 (SPL3) was only shown to be able to bind to the promoter of *FLA4*.

Table 7. List of TFs that can potentially bind to the promotor region of *FLA8*.

Gene	Description	Family	Potential binding sites	Function	Reference
At4g18960	AGAMOUS (AG)	MADS	2	Floral homeotic protein. Encoded by class-C gene required for the development of stamens and carpels.	Coen *et al*., 1991
At3g58780	AGAMOUS-LIKE 1(AGL1); SHATTERPROOF 1 (SHP1)		2	Floral homeotic protein. Encoded by class-D gene required for seed coat development and endosperm formation.	Ehlers *et al*., 2016
At5g15800	AGAMOUS-LIKE 2 (AGL2); SEPALLATA 2 (SEP2)		4	Floral homeotic proteins. Encoded by class-E genes. Class-E genes are required for the formation of sepals, petals, stamens, carpels and ovules.	Huang *et al*., 1996
At2g03710	AGAMOUS-LIKE 3 (AGL3); SEPALLATA 4 (SEP4)		4		
At5g13790	AGAMOUS-LIKE 15 (AGL15)		10	Preferentially expressed during embryogenesis. Prevents premature abscission and seed desiccation. Upregulates the expression of *LEC2* and other genes related to embryogenesis	Perry *et al*., 1999; Fernandez *et al*., 2000; Zheng *et al*., 2009
At1g09540	MYB61	MYB	3	Important for deposition of mucilage during development of the seed coat epidermis. Important in mucilage extrusion from the seeds during imbibition.	Stracke *et al*., 2001; Arsovski *et al*., 2009
At2g46590	DOF AFFECTING GERMINATION 2 (DAG2)	C2C2(Zn) DOF	4	Maternal exclusive factor that positively regulates light-mediated seed germination.	Santopolo *et al*., 2015

Table 8. List of TFs that can potentially bind to the promotor region of *FLA1*.

Gene	Description	Family	Potential binding sites	Function	Reference
At3g58780	AGAMOUS-LIKE 1(AGL1); SHATTERPROOF 1 (SHP1)	MADS	1	Floral homeotic protein. Encoded by class-D gene required for seed coat development and endosperm formation.	Ehlers *et al*., 2016
At5g15800	AGAMOUS-LIKE 2 (AGL2); SEPALLATA 2 (SEP2)		1	Floral homeotic proteins. Encoded by class-E genes. Class-E genes are required for the formation of sepals, petals, stamens, carpels and ovules.	Huang *et al*., 1996
At2g03710	AGAMOUS-LIKE 3 (AGL3); SEPALLATA 4 (SEP4)		2		
At5g13790	AGAMOUS-LIKE 15 (AGL15)		4	Preferentially expressed during embryogenesis. Prevents premature abscission and seed desiccation. Upregulates the expression of *LEC2* and other genes related to embryogenesis	Perry *et al*., 1999; Fernandez *et al*., 2000; Zheng *et al*., 2009
At1g09540	MYB61	MYB	1	Important for deposition of mucilage during development of the seed coat epidermis. Important in mucilage extrusion from the seeds during imbibition.	Stracke *et al*., 2001; Arsovski *et al*., 2009
At2g46590	DOF AFFECTING GERMINATION 2 (DAG2)	C2C2(Zn) DOF	6	Maternal exclusive factor that positively regulates light-mediated seed germination.	Santopolo *et al*., 2015
At1g02065	SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 8 (SPL8)	SBP	1	Required for ovule differentiation, pollen production, seed formation, siliques elongation and polen sac development	Unte *et al*., 2003; Zhang *et al*., 2007
At3g51060	SHI-RELATED SEQUECE 1(SRS1); STYLISH 1 (STY1)	SRS	2	Promotes expression of genes affecting stamen development, cell expansion and timing of flowering	Ståldal *et al*., 2012
At1g28300	LEAFY COTYLEDON 2 (LEC2)	ABI3/VP1	2	*LEC2* RNA accumulates during seed development. Required for the maintenance of suspensor morphology, cotyledon identity, progression through the maturation phase, and suppression of premature germination.	Tsukagoshi *et al*., 2007

Table 9. List of TFs that can potentially bind to the promotor region of *FLA4*.

Gene	Description	Family	Potential binding sites	Function	Reference
At3g58780	AGAMOUS-LIKE 1(AGL1); SHATTERPROOF 1 (SHP1)	MADS	1	Floral homeotic protein. Encoded by class-D gene required for seed coat development and endosperm formation.	Ehlers *et al*., 2016
At5g15800	AGAMOUS-LIKE 2 (AGL2); SEPALLATA 2 (SEP2)		1	Floral homeotic proteins. Encoded by class-E genes. Class-E genes are required for the formation of sepals, petals, stamens, carpels and ovules.	Huang *et al*., 1996
At2g03710	AGAMOUS-LIKE 3 (AGL3); SEPALLATA 4 (SEP4)		1		
At5g13790	AGAMOUS-LIKE 15 (AGL15)		3	Preferentially expressed during embryogenesis. Prevents premature abscission and seed desiccation. Upregulates the expression of *LEC2* and other genes related to embryogenesis	Perry *et al*., 1999; Fernandez *et al*., 2000; Zheng *et al*., 2009
At1g09540	MYB61	MYB	1	Important for deposition of mucilage during development of the seed coat epidermis. Important in mucilage extrusion from the seeds during imbibition.	Stracke *et al*., 2001; Arsovski *et al*., 2009
At2g33810	SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 3 (SPL3)	SBP	1	Promotes both vegetative phase change and flowering	Cardon *et al.*, 1998
At1g02065	SQUAMOSA PROMOTER-BINDING-LIKE PROTEIN 8 (SPL8)	SBP	1	Required for ovule differentiation, pollen production, seed formation, siliques elongation and polen sac development	Unte *et al*., 2003; Zhang *et al*., 2007
At1g28300	LEAFY COTYLEDON 2 (LEC2)	ABI3/VP1	2	*LEC2* RNA accumulates during seed development. Required for the maintenance of suspensor morphology, cotyledon identity, progression through the maturation phase, and suppression of premature germination.	Tsukagoshi *et al*., 2007

3.1.3. *stk* RNA-seq analysis

STK is a MADS-box TF known to be a master regulator of fertilization and seed development (Mizzotti *et al.*, 2012; Mizzotti *et al.*, 2014). The analysis of transcriptome data regarding this gene is, therefore, an essential tool in identifying the possible functions of genes that are highly expressed in reproductive tissues.

Publicly available *stk* RNA-seq data (Mizzotti *et al.*, 2014) was analysed to check if the expression of *FLA1*, *FLA4* and *FLA8* was affected by the mutation in this gene. According to this data, the *knock-out* mutation that stopped *STK* expression did not affect the expression of either of these three FLA genes, although other FLA genes where shown to be differently expressed in *stk* mutant plants, such as AT4G31370/FLA5, a gene that is essentially expressed in flowers at stage 6-8 according to Klepikova Arabidopsis Atlas (Klepikova *et al.,*2016).

3.2. Genotyping of mutant lines

A PCR based approach was used to identify homozygous mutant plants, containing the T-DNA insert, that would later be used to evaluate the phenotype of these mutant plants. Two different reaction mixes were used. In one of them the added primers are the Border Primer (BP), that binds to the border of the T-DNA insertion, and the Right Primer (RP), which binds to the forward strand of the target gene. Amplification between BP and RP produces a mutant (mt) band. The other reaction mix has the Left Primer (LP), which binds to the reverse strand of the target gene, in conjunction with RP. Amplification between LP and RP generates a wild-type (wt) band. The presence of absence of mt or wt bands, as seen through an electrophoretic analysis, allow for the determination of the genotypes of the different plants. Band sizes and the electrophoretic analysis are shown in Figure 8.

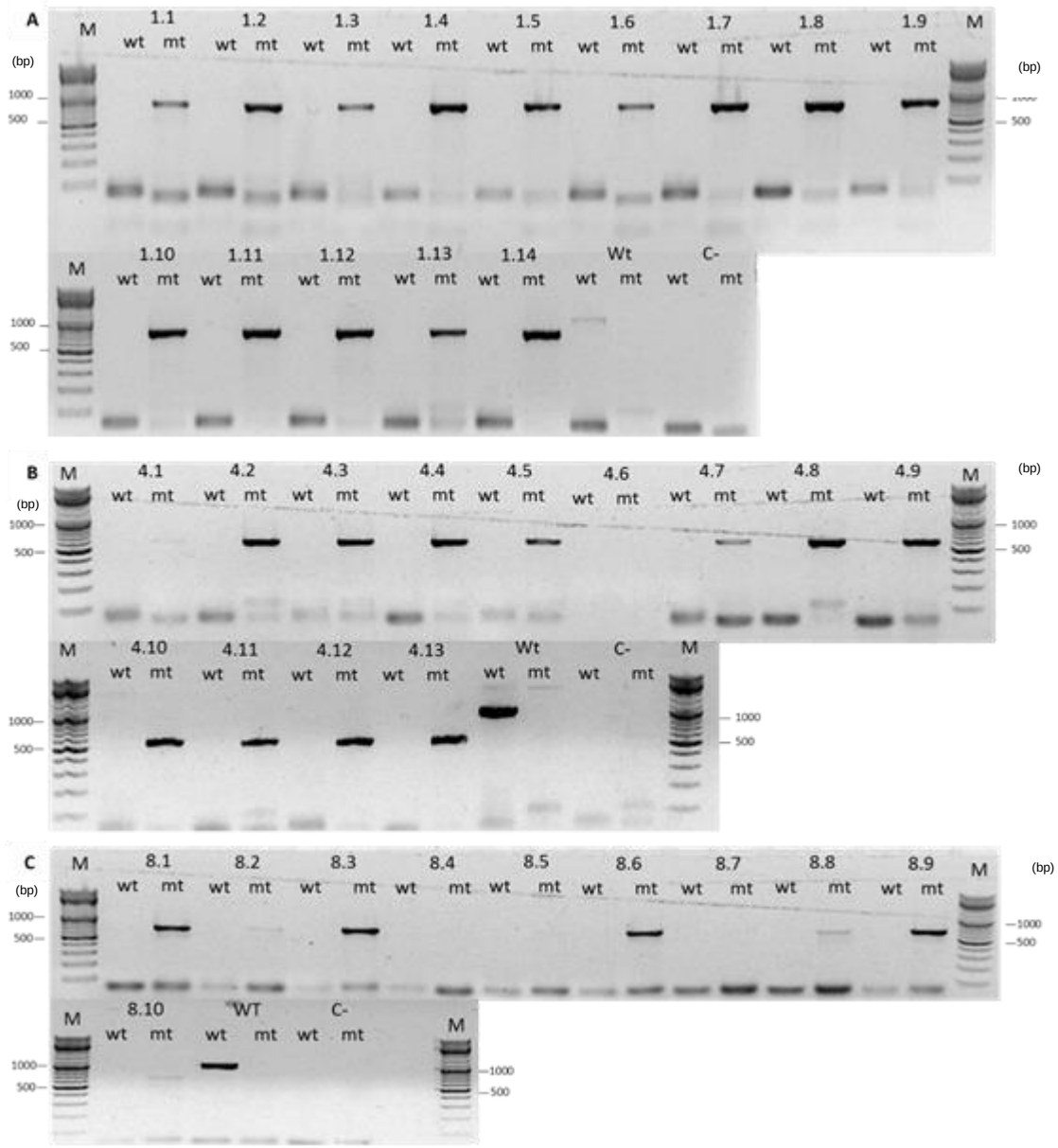


Figure 8. Electrophoresis of DNA fragments corresponding to *fla1*, *fla4* and *fla8* genotyping, on 1% (w/v) agarose gel. Amplification between BP and RP produces a mutant (mt) band. Amplification between LP and RP generates a wild-type (wt) band. A) Genotyping of 14 *fla1* plants ;expected wt band size = 1015 bp ; expected mt band size ~800 bp. B) Genotyping of 13 *fla4* plants ;expected wt band size = 1051 bp; expected mt band size ~600 bp. C) Genotyping of 10 *fla8* plants ;expected wt band size = 1096 bp; expected mt band size ~750 bp. WT lane: gDNA from WT plant was used as control, only a fragment between LP and RP is amplified, no mt band is formed. C- lane: no DNA was added therefore no specific amplification is expected. (M): Molecular size ladder.

3.3. Checking the nature of the mutant lines *fla1*, *fla4* and *fla8*

The expression levels of *FLA1*, *FLA4* and *FLA8* were detected using a semi-quantitative RT-PCR analysis using RNA extracted from complete inflorescences of plants from SALK_058964.45.15 (*fla1*), SALK_125874.44.30 (*fla 4*) and SALK_141852 (*fla 8*) T-DNA insertion lines. This RNA was converted to cDNA that was then used in this analysis. Figure 9 represents the RT-PCR analysis. This data show that each *FLA* gene is only being expressed in the WT plants and is not being expressed in *fla1*, *fla4* and *fla8* plants. *ACT8* was used as reference gene and, taken together, this data reveals that the three mutant lines are likely to possess a complete knock-out of each *FLA* gene.

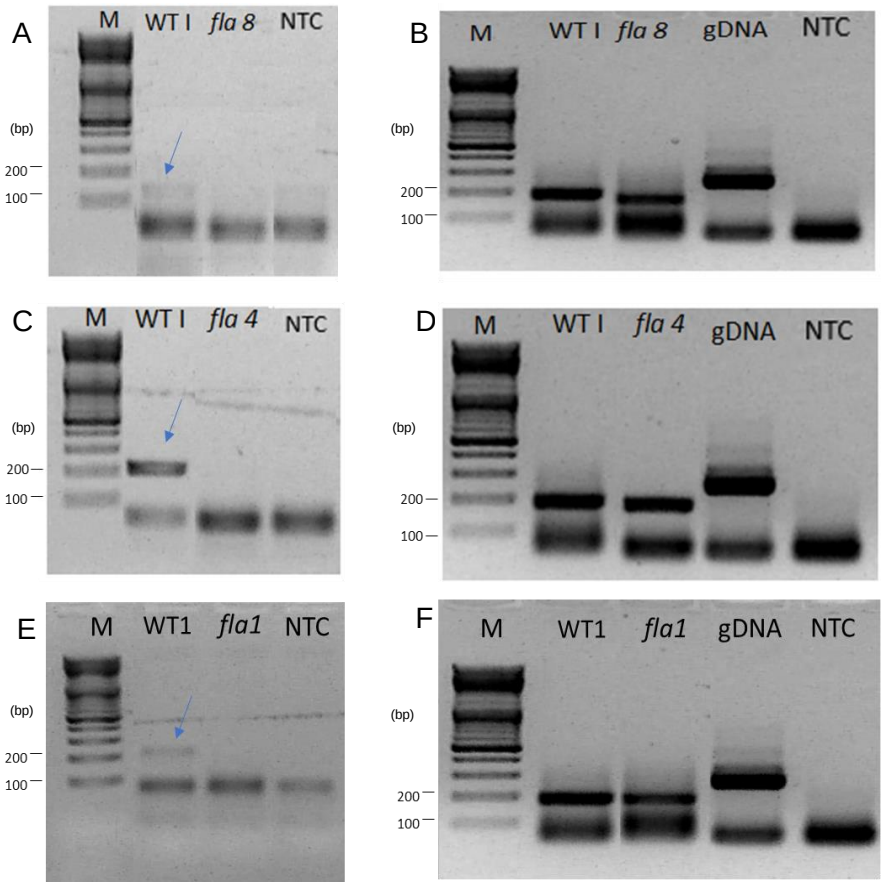


Figure 9. RT-PCR analysis of cDNA obtained using RNA from complete inflorescences of *fla1*, *fla4* and *fla8* and WT plants. A) Analysis of the expression of *FLA8* in WT and *fla8*; B) Expression of *ACT8* in WT and *fla8*; C) Analysis of the expression of *FLA4* in WT and *fla4*; D) Expression of *ACT8* in WT and *fla4*; E) Analysis of the expression of *FL18* WT and *fla1*; F) Expression of *ACT8* in WT and *fla1*. Expected band sizes and primers are displayed in Supplemental Table 1. gDNA from a WT plant was used to distinguish the band size obtained using cDNA or gDNA in the PCR protocol; NTC is the no template control; blue arrows point to the gene fragment obtained with the primers for each specific *FLA* gene.

3.4. Design and construction of transgenic lines

As a way of studying the expression pattern of *FLA1* and *FLA4*, transgenic marker lines of p*FLA1:GUS* and p*FLA4:GUS* were designed using the Gateway® Gene Cloning technique (Thermo Fisher Scientific). WT plants were transformed using *A. tumefaciens* GV3101::p*MP90* transformed with the recombinant plasmid, p*BGWFS7*+p*FLA1:GUS* and p*BGWFS7*+p*FLA4:GUS*. Prior to transformation of the plants, the presence of the transformed plasmid was confirmed via a colony PCR assay (represented in Figure 10). So far, the seeds are yet to be recovered and analysed to see if they were transformed with the plasmids. The transformed plants will have their genotypes determined using a PCR approach and a GUS detection assay will be performed later on. A previously obtained p*FLA14:GUS* plant line was already analysed. The third generation of this plant line was used to determine the putative expression pattern of *FLA14* , via a GUS staining assay (this line was designed and obtained by Ferreira MJ and Amorim MI, 2017, unpublished). This gene was shown to be mainly expressed in anthers containing bicellular/tricellular pollen, on flower development stage 12, and also with some residual expression observed in mature pollen grains and in pollen tubes. The expression pattern is represented in Figure 11.

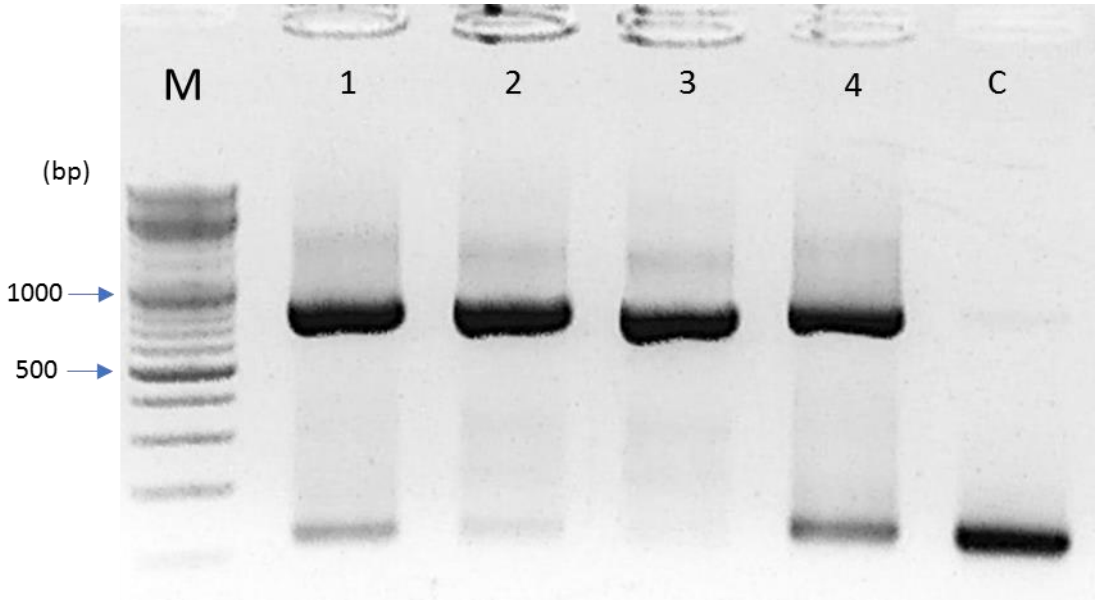


Figure 10. **Colony PCR of 4 *E.coli* colonies transformed using p*BGWFS7*+p*FLA4:GUS*.** The band size is approximately 726 bp. M is the molecular size ladder. C is the negative control.

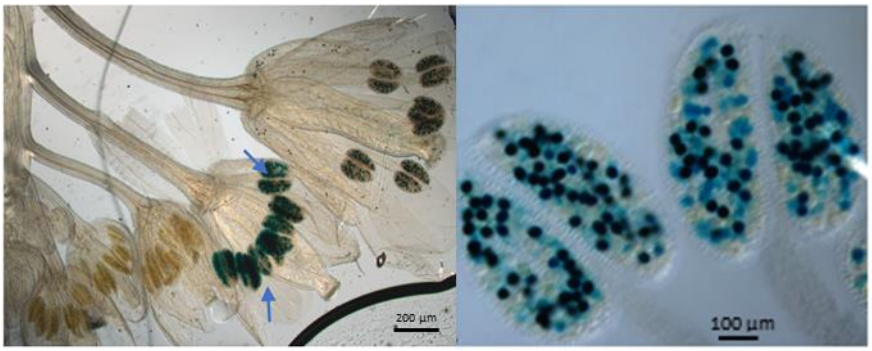


Figure 11. Putative gene expression of FLA14 in *A.thaliana* inflorescences and anthers.

Col-0 wild-type plants were transformed using *A. tumefaciens* containing p*FLA14:GU*S and a GUS staining assay was performed in order to deduce the putative gene expression pattern of FLA14. This gene was shown to be mainly expressed in anthers containing tricellular pollen because GUS staining is intense in anthers with pollen in this development state (blue arrows), with some residual expression also being observed in mature pollen grains and in pollen tubes.. Plants containing p*FLA1:GUS* and p*FLA4:GUS* constructs were also obtained, but the expression pattern of these genes is yet to be analysed with a GUS staining protocol.

3.5. Seed set analysis of mutant lines

The siliques of two months old *fla1*, *fla4*, *fla8* and WT plants were analysed, and the seed set of these siliques was registered. Ten siliques, in similar development stages, from each plant were chosen, as a way to represent the plant as a whole. Four major phenotypes can be determined from analysing the ovules and seeds inside the siliques: green seeds, resultant from a correct development of the embryo and the endosperm; white seeds, corresponding to seeds containing embryos whose development was arrested at the pre-globular stage; aborted ovules, a result of unfertilized ovules and lastly aborted seeds, which are the outcome of seeds whose embryo did not mature beyond the pre-globular stage and later on suffered desiccation, acquiring a brown colour (further description in http://seedgenes.org/Tutorial.html). The analysis is shown in Figure 12. The comparison with WT siliques showed that, in average, *fla8* plants possessed shorter siliques and a reduced number of green seeds. The siliques of *fla8* plants measured, on average, around 1,30 cm and possessed 40 green seeds, while WT siliques measured 1,46 cm and possessed 50 green seeds. The siliques of *fla8* plants had 3 times the number of aborted ovules and also displayed a decreased number of white seeds, when compared to the siliques of WT plants. On average, *fla1* mutant

plants possessed 2 aborted seeds per silique, while WT siliques possessed less than 1 aborted seed. The siliques of *fla4* were similar do those of WT in every parameter evaluated, except on the total number of seeds in each silique. This last parameter revealed 2 different variations from the norm. *fla1* and *fla4* plants possessed, on average, siliques with 58 and 52 seeds, respectively. WT siliques have 49 seeds and *fla8* have siliques with 44 seeds.

Images of siliques and the different phenotypes of seeds and ovules are presented in Figure 13.

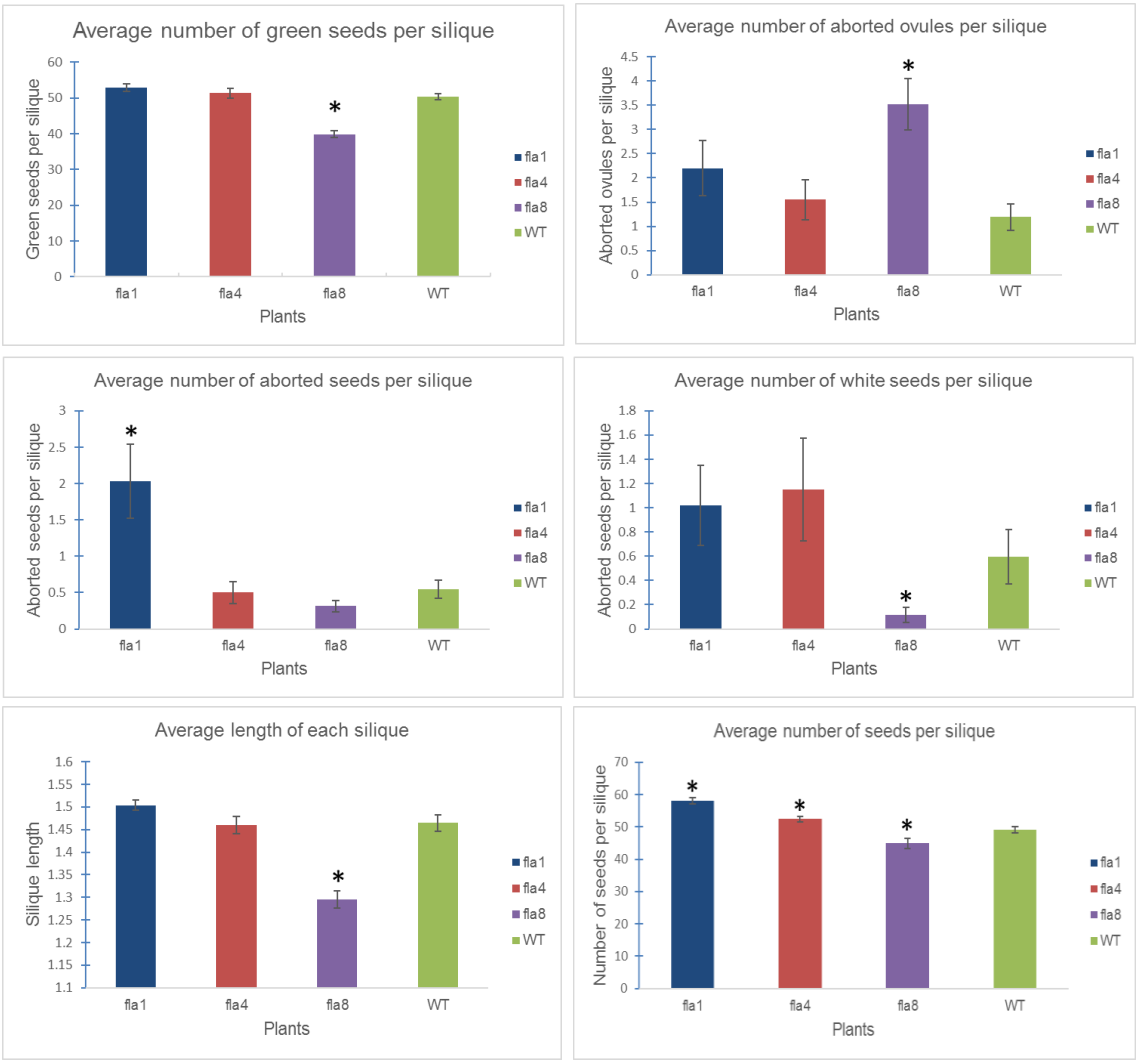


Figure 12. **Seed set analysis of *fla1, fla4, fla8* and WT plants**. 5 different parameters were evaluated in this analysis: the average number of green seeds, white seeds, aborted seeds, aborted ovules and the average length of the siliques. The average number of seed per silique was also calculated. Error bars are the (±) standard error. A Student's t-test was performed, α = 0.05. * indicates statistical differences when compared to WT.

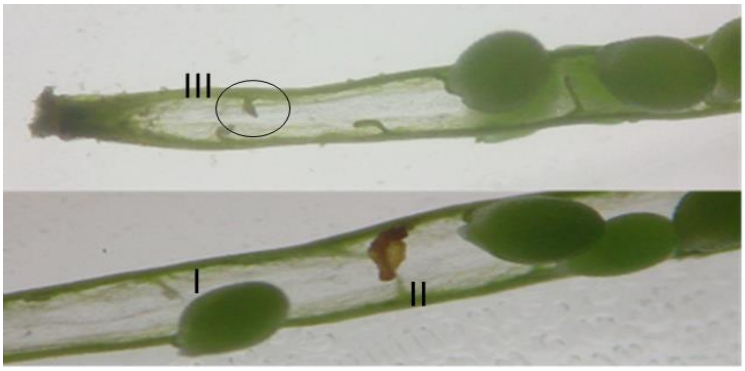


Figure 13. Image depicting a silique and its seed set from a *fla4* plant. Siliques were harvested and seed set was analysed by assessing the number of green seeds (I), aborted seeds (II), aborted ovules (III) and white seeds (not depicted here).

3.6. Germination assay

A germination assay was performed to check for anomalies in the germination rate that could be caused by the T-DNA insertion on each mutant line. The germination rate was measured 96 hours after the seeds were placed in the greenhouse and were allowed to grow. The data regarding this assay is displayed in Figure 14. Every situation displayed a germination rate above 88 %, but no statistically significant differences were determined between the germination rate of WT plants and the germination rate of the mutant plants.

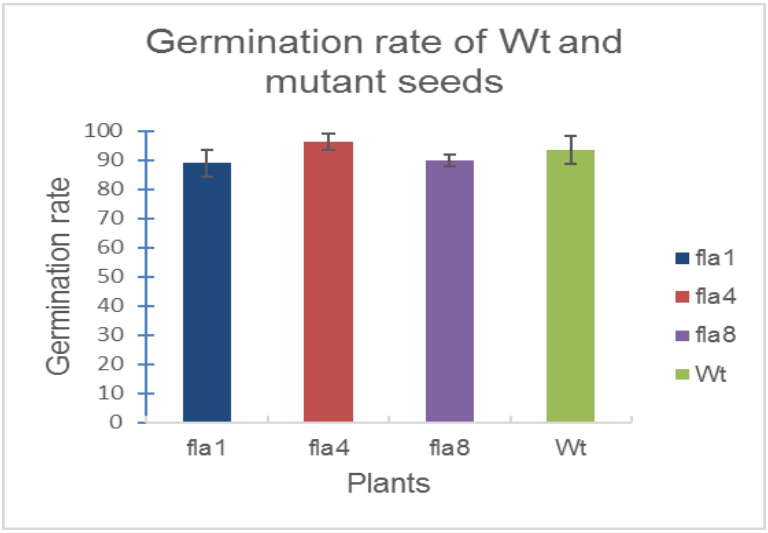


Figure 14. Germination rate measured 96 hours after the seeds were placed in the greenhouse. No statistical differences were measured between WT and mutant plants. Error bars are the (±) standard error. A Student's t-test was performed, α = 0.05.

3.7. Seed morphology assessment

The seed morphology assessment was performed to identify possible anomalies in the morphology of the seeds that could be caused by the T-DNA insertion on the mutant lines. The phenotypes of mature seeds observed can be catalogued into 3 major types: round seeds, shrivelled seeds and flat seeds. A sample of over 100 mature seeds of each mutant line, previously collected from homozygous mutant plants, was observed under a dissection microscope and the percentage of occurrence of each phenotype was documented. This test was performed twice and a Student´s t-test was used to reveal that the differences observed in the percentage of round, shrivelled and flat seeds were

not statistically significant. This analysis is depicted in figure 15. The three phenotypes of seeds are depicted in Figure 16.

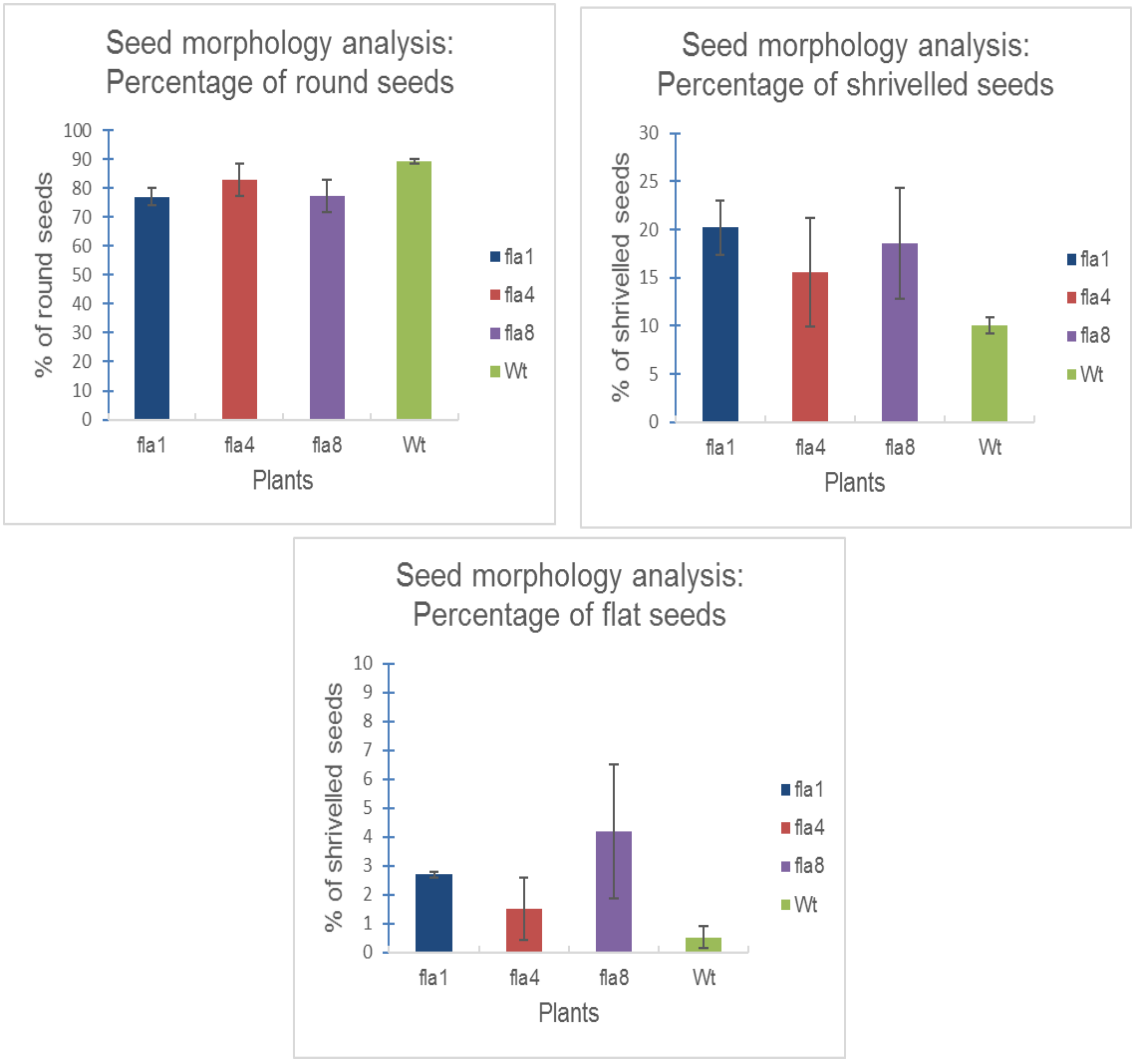


Figure 15. Seed morphology analysis. The percentage of occurrence of three seed phenotypes were recorded. There were no significant differences between the percentage of round, shrivelled or flat seeds between WT and the three mutant lines. Error bars are the (±) standard error. A Student's t-test was performed, α = 0.05.

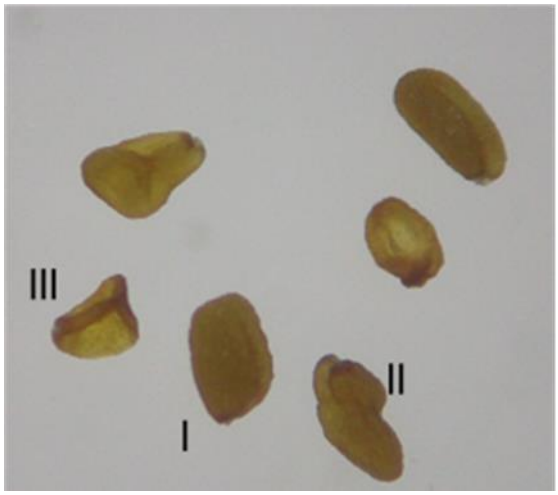


Figure 16. Seed morphology: analysis of the phenotypes. Image depicting the three major seed phenotypes observed. I: Round (Wt) seed; II: Shrivelled seed; III: Flat seed.

4. Discussion

4.1. Bioinformatics analysis

The expression pattern of *FLA1*, *FLA4* and *FLA8* in *A. thaliana* (analysed using both eFP Browser and Klepikova Arabidopsis Atlas eFP Browser) revealed interesting results. *FLA1* and *FLA8* were shown to be expressed in various tissues, but with some common patterns. Both are expressed in flowers from stages 9 to 12, in both petals and carpels, and both appear to be highly expressed in several types of root cells and in the pistils. The main difference between the two putative expression patterns is their distribution in seed tissues. *FLA8* is expressed in the seed coat of seeds containing pre-globular and globular embryos, while *FLA1* seems to be highly expressed in seeds containing torpedo stage embryos, appearing in the cotyledons of the embryos and also in the peripheral endosperm of developed seeds. *FLA4´*s expression information is not available on eFP Browser, therefore it is impossible to thoroughly compare the putative expression pattern of this gene with the other two FLA genes studied. Information regarding the expression of this gene is only available on Klepikova Arabidopsis Atlas, a less detailed platform that does not include information such as the specific seed tissue in which this gene is being expressed.

The differences in the expression pattern can be attributed to the different kinds of TF that bind to their promoter, promoting their expression in various tissues. In order to establish a connection between the expression pattern and the TFs that can bind to the promoters of these genes, an analysis using AthaMap was performed. This platform is able to identify up to 221 TFs and their binding motifs, using previously documented information extracted from literature regarding these proteins. The small pool of binding sites registered in the AthaMap database restricts the amount of information that can be obtained regarding the total amount of TFs that bind to the promoters of the target genes. Therefore, it is impossible to obtain enough information regarding the transcriptional control of the *FLA* genes, when using only this tool.

The TFs binding sites analysis was restricted to families of *A. thaliana* TFs known to be important during reproduction related processes in this species. Taken this information into account, it was observed that the promoters of the *FLA* genes studied were rich in MADS-box proteins binding sites. These TFs are known to have essential functions, ranging from root to flower and fruit development (Becker and Theißen *et al*., 2003). MADS-box TFs share a common DNA-binding domain (the MADS-box) and recognize

similar target DNA sequences (Ng and Yanofsky ,2001). The pool of MADS-box proteins whose recognition sites are registered in AthaMap is very restricted. This information, coupled with the wildly reported fact that many MADS-box proteins act redundantly in many processes related to reproduction (even when they belong to completely different clades of the MADS box phylogenetic tree), hinders the ability to associate the presence of a MADS-box binding site to a specific function during plant reproduction (Gregis *et al*., 2006; 2008).

FLA8 was the only gene, from the three studied here, whose promoter possessed the binding sequence recognized by *AG*. This C-class floral homeotic gene has been extensively studied and has been shown to be essential in floral meristem determinacy (when forming tetramers with SEPALATTA 3), control of sepal senescence and abscission and other functions (Jibran *et al*., 2017; Hugouvieux *et al*., 2018). All three genes studied also possessed binding sites for other homeotic flower genes such as SHP1,SEP2 and SEP4, as well as potential binding sites for AGL15, a MADS-box TF known to be involved in the transition from seed maturation to vegetative growth and a positive regulator of LEC2 and other genes related to embryogenesis (Zheng *et al*., 2009; Chen *et al*., 2018). LEC2 has the potential of regulating the expression of both *FLA1* and *FLA4,* and this TF is known to directly control a transcriptional program involved in the maturation phase of seed development (Braybrook *et al*., 2006; Tsukagoshi *et al*., 2007).

The other TFs found to be able to bind to the promoter region of the *FLA* genes studied are MYB61 (a TF known to be important for deposition of mucilage during development of the seed coat epidermis), DAG2 (a Dof zinc finger protein known to be involved in seed germination), STY1 (involved in stamen development and flowering), SPL3 (involved in vegetative change and flowering) and SPL8 (TF known to be required for ovule differentiation, pollen production, seed formation, siliques elongation and pollen sac development) (Cardon *et al*., 1998; Zhang *et al*., 2007; Arsovski *et al*., 2009).

Current literature regarding the binding potential of AGL15 did not identify any of the *FLA* genes studied here as being potentially regulated in a direct matter by this TF. This was achieved through a ChiP-seq analysis performed by Wang and associates (2002). These contradictory results demonstrate that a bioinformatics analysis by itself is not sufficient to understand the intricate relations between TF and the genes whose activity they control.

4.2. Knock-out of STK does not affect the expression of *FLA1*, *FLA4* and *FLA8*

The analysis of publicly available stk RNA-seq data (Mizzotti et al., 2014) determined that the expression of *FLA1*, *FLA4* and *FLA8* was not affected by the mutation in this gene. This result can be interpreted in two different ways. The first interpretation is that STK is not involved in the transcriptional regulation of these three *FLA* genes, therefore the lack of *STK* expression did not affect the expression of either of them. The second possible interpretation is related to the interchangeability between STK and the other D-class flower homeotic genes, SHP1 and SHP2. Since SHP1 can bind to the promoter regions of all three *FLA* genes studied, it is possible that SHP1 could substitute the defective STK protein in the tetrameric formation that would bind to the promoter of either one of the *FLA* genes, regulating their expression. This occurrence could reduce any potential changes in gene expression and therefore no alterations in *FLA1*, *FLA4* and *FLA8* expression would be detected.

4.3. *fla8* plants show defects in seed set

A semi-quantitative RT-PCR analysis determined that all *fla* mutant lines utilized could be considered single knock-out lines for each corresponding *FLA* gene. This analysis will be repeated in the future to consolidate this result. A qRT-PCR analysis can also be performed later on to acquire a better result regarding the expression of the *FLA* genes in these mutants.

After confirmation of the mutant nature of these lines, the analysis of the seed set of homozygous mutant plants could be used to deduce possible functions of *FLA1*, *FLA4* and *FLA8*. The *fla8* mutants were the ones demonstrating more alterations on certain traits (when compared to WT) such as, smaller siliques and a lesser amount of both green and white seeds per silique. Furthermore, *fla8* siliques had an increased number of aborted ovules, when comparing to WT siliques. These changes indicate that *FLA8* may have an important role in both the formation of the siliques and in the maturation of seeds. The larger numbers of aborted unfertilized ovules could also point out to a role of *FLA8* during the processes underlying pollen tube guidance. Also, the decreased number of white seeds in siliques may point to a role of the native FLA8 protein during the

development of the embryo, since *FLA8* is expressed in the general seed coat of seeds containing embryos in the pre-globular and globular state.

fla1 homozygous plants contained more aborted seeds than WT plants. This phenotype shows that FLA1 knock-out probably triggered an arrest in the development of the embryos, preventing them from advancing beyond the pre-globular stage. This alteration was unexpected according data obtain in eFP Browser, since *FLA1* was shown to be mainly expressed in seeds in more advanced stages, such as the torpedo stage.

The total number of seeds per silique revealed that *fla1* and *fla8* plants had siliques with a greater number of seeds, when compared to the WT situation. Since the average length of each silique was not altered, it is possible that these mutant plants also possessed seeds with smaller sizes, a phenotype which would be caused by the mutation in these *FLA* genes. *fla8* plants on the other hand had a lesser number of seeds and also smaller siliques, which may indicate that the average size of each individual silique was not affected.

4.4. Seed germination was not affected in the mutant plants

The seed germination assay performed revealed that the germination rate of seeds from *fla1*, *fla4* and *fla8* homozygous plants was not hindered, when compared to the germination rate of WT plants. FLA4 was previously shown to be important for seed coat mucilage composition, possibly acting in a linear genetic pathway with the leucine-rich repeat receptor-like kinase loci (LRR-RLK) AtFEI1and AtFEI2 (Griffiths et al., 2016; Xue et al., 2017). This function, apparently, does not affect the germination rate of plants with a knock-out of the *FLA4* gene. This information may indicate that *FLA4* is not important for the germination process. On the other hand, *FLA1* and *FLA8* were never associated with the germination processes by previously published studies.

4.5. *fla* mutant plants do not shown defects in seed morphology

Mature seeds from homozygous *fla1*, *fla4* and *fla8* were observed under a dissecting microscope. No changes in the percentage of round seeds were detected in either of the seeds from mutant homozygous plants and WT seeds. The percentage of aberrant

seeds was also not affected by the mutated genes, which may indicate that the endosperm of the mutant seeds was also unaffected.

This lack of differences may also be due to redundancy in the function of the *FLA* genes studied. This could explain the reason why single mutants of these genes do not shown severe phenotypical changes in the seed morphology and in germination rate. Cross-fertilization between *fla1* and *fla8* homozygous plants, per example, could shed new light on the functions of these closely related genes.

4.6. Creation of marker lines

Marker lines for p*FLA1:GUS* and p*FLA4:GUS* are close to being analysed. So far, WT plant were transformed using *A. tumefaciens* bacteria containing the transformed plasmids with p*FLA1:GUS* and p*FLA4:GUS* constructs. The histochemical detection of β-glucuronidase activity will allow for the determination of a clear expression pattern of *FLA1* and *FLA8* in the reproductive tissues of *A. thaliana*, which will in turn help in the discovery of the role of these genes during reproduction. Col-0 wild-type plants transformed using *Agrobacterium tumefaciens* containing p*FLA14:GUS* displayed the putative gene expression pattern of FLA14. This gene was shown to be mainly expressed in anthers containing tricellular pollen, with some residual expression also being observed in mature pollen grains and in pollen tubes. This analysis was important because it confirmed the microarray data available on eFP Browser about this gene. This restricted expression pattern indicates that this gene may have important functions in the control of pollen development and its function should be studied using mutants and other available tools.

5. Conclusion

The objective of this work was to uncover the role of FLA1, FLA4 and FLA8 in the processes related to the reproduction in *A. thaliana*. The bioinformatic analysis performed demonstrated that the expression of the genes encoding the three FLA proteins is controlled by a large number of TF known to be important during essential steps of plant reproduction, therefore indicating a possible role of FLA1, FLA4 and FLA8 in the same processes. The phenotypical analysis of T-DNA insertion lines for each gene revealed that *fla8* and *fla1* showed defects related to silique formation and seed maturation. *fla4* plants did not show the same defects that were detected in the other mutant plants. These results point to a possible function of *FLA1* during early steps of embryo development and to a possible function of *FLA8* in both the determination of silique formation and during early stages of embryo development. No significant differences in seed morphology and germination were observed in any of the plants from the T-DNA insertion lines, which leads to the conclusion that these genes are probably not involved in the control of seed germination and may not affect the morphology of the seed and the development of the endosperm.

The results obtained will require further validation. Firstly, a qRT-PCR analysis will be performed to definitively determine if the mutant lines have a complete knock-out of the genes of interest. Secondly, the mutant phenotypes have to be confirmed using a phenotype complementation assay, in which the mutant plants will be transformed with a non-mutant form of the gene in order to determine if the phenotypic differences that here detected in this study were a product of the mutation of the *FLA* genes.

There are still many questions to be answered regarding the function of these proteins in the reproductive tissues of *A. thaliana.* The expression pattern of these proteins in the *A. thaliana* tissues will be further analysed using plants expressing p*FLA1:GUS* and p*FLA4:GUS* constructs that were created during this project. The histochemical detection of GUS activity will help determine the tissue specific pattern of expression of the native *FLA1* and *FLA4* genes. The p*FLA8:GFP* line is already available and can also be used for the same objective.

6. References

Amici G (1824) Observations microscopiques sur diverse espèces de plantes. Annales des Sciences Naturelles: Botanique, 2: 41-70, 211-248

Amici G (1830) Note sur le mode d'action du pollen sur le stigmate. Extrait d'une lettre d'Amici à Mirbel. Annales des Sciences Naturelles: Botanique, 21: 239-332

Amici G (1847) Sur la fécondation des Orchidées. Annales des Sciences Naturelles: Botanique, 7/8: 193-205

Arsovski A, Villota M, Rowland O, Subramaniam R, Western T (2009). MUM ENHANCERS are important for seed coat mucilage production and mucilage secretory cell differentiation in *Arabidopsis thaliana*. Journal of Experimental Botany, 60(9):, 2601–2612

Basu D, Tian L, Debrosse T, Poirier E, Emch K, Herock H, Travers A, Showalter M (2016) Glycosylation of a fasciclin-like arabinogalactan-protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) pathway in *Arabidopsis*. PloS one: 11(1)

Battaglia R, Vittoria B, Colombo L (2008) Morphological analysis of female gametophyte development in the *bel1 stk shp1 shp2* mutant. Plant Biosystems, 142: 643-649

Becker A, Theißen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molecular Phylogenetics and Evolution,. 29 (3) : 464-489

Boavida L, Vieira A, Becker J, Feijo J (2005) Gametophyte interaction and sexual reproduction: how plants make a zygote. International Journal of Developmental Biology, 49: 615-632

Borner GHH, Sherrier DJ, Weimar T, Michaelson LV, Hawkins ND, MacAskill A, Napier JA, Beale MH, Lilley KS, Dupree P (2005) Analysis of detergent-resistant membranes in *Arabidopsis*. Evidence for plasma membrane lipid rafts. Plant Physiology, 137: 104-116

Braybrook S, Stone S, Park S, Bui A, Le B, Fischer R, Goldberg R, Harada J (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis Proceedings of the National Academy of Sciences, 103 (9) : 3468-3473

Cardon G, Höhmann S, Nettesheim K, Saedler H, Huijser P (1997) Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: a novel gene involved in the floral transition. The Plant Journal,. 12(2) : 367-77

Chen N, Veerappan V, Abdelmageed H, Kang M, Allen R (2018) HSI2/VAL1 silences AGL15 to regulate the developmental transition from seed maturation to vegetative growth in *Arabidopsis*. Plant Cell, 30(3): 600-619

Chow C-N, Zheng H-Q, Wu N-Y, Chien C-H, Huang H-D, Lee T-Y, Chiang-Hsieh Y-F, Hou P-F, Yang T-Y, Chang W-C (2016) PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. Nucleic Acids Research, 44: D1154-D1160

Clough SJ and Bent AF (1998) Floral Dip: A Simplified Method for *Agrobacterium*-Mediated Transformation of *Arabidopsis thaliana*. The Plant Journal, 16: 735-743

Coen S, Meyerowitz M (1991) The war of the whorls: Genetic interactions controlling flower development. Nature, 353: 31–37

Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG (2009) Pollen grain development is compromised in Arabidopsis *agp6 agp11* null mutants. Journal of Experimental Botany, 60 (11): 3133–3142

Coimbra S, Jones BJ, Pereira LG (2008) Arabinogalactan proteins (AGPs) related to pollen tube guidance into the embryo sac in *Arabidopsis*. Plant Signaling & Behavior, 3: 455–456

Costa LM, Gutierrez-Marcos JF, Dickinson HG (2004) More than a yolk: the short life and complex times of the plant endosperm. Trends in Plant Science, 9: 507-514

Costa M, Nobre M, Becker J, Masiero S, Amorim MI, Pereira LG, Coimbra S (2013) Expression-based and co-localization detection of Arabinogalactan protein 6 and Arabinogalactan protein 11 interactors in Arabidopsis pollen and pollen tubes. BioMed Central Plant Biology, 13: 7

Costa M, Sobral R, Ribeiro Costa MM, Amorim MI, Coimbra S (2015) Evaluation of the presence of arabinogalactan proteins and pectins during *Quercus suber* male gametogenesis. Annals of Botany, 115: 81-92

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

Ehlers K, Bhide AS, Tekleyohans DG, Wittkop B, Snowdon RJ, Becker A (2016) The MADS Box gGenes ABS, SHP1, and SHP2 aAre essential for the coordination of cell divisions in ovule and seed coat development and for endosperm formation in *Arabidopsis thaliana*. PLoS One, 11

Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-proteins: key regulators at the cell surface? Plant Physiology, 153: 403-419

Fan YF, Jiang L, Gong HQ, Liu CM (2008) Sexual reproduction in higher plants I: Fertilization and the initiation of zygotic program. Journal of Integrative Plant Biology, 50: 860-867

Fernandez D, Heck G, Perry S, Patterson S, Bleecker A, Fang S. (2000) The embryo MADS domain factor AGL15 acts post embryonically. Inhibition of perianth senescence and abscission via constitutive expression. Plant Cell, 12: 183-198

Franklin-Tong N (2010) Plant fertilization: bursting pollen tubes! Current Biology, 20(16): 681-683

Gregis V, Sessa A, Colombo L, Kater M (2006) AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis. Plant Cell, 18 : 1373–1382

Gregis V, Sessa A, Colombo L, Kater M (2008) AGAMOUS-LIKE 24 and SHORT VEGETATIVE PHASE determine floral meristem identity in *Arabidopsis*. The Plant Journal,. 66 : 891–902

Griffiths JS, Crepeau M-J, Ralet M-C, Seifert GJ, North HM (2016) dissecting seed mucilage adherence mediated by FEI2 and SOS5. Frontiers in Pplant Sscience, 7: 1073-1073

Guignard L (1899) Sur les antherozoides et la double copulation sexuelle chez les végétaux angiosperms. Comptes Rendus de l''Académie des Sciences Paris, 128: 864-871

Harpaz-Saad S, Western TL, Kieber JJ (2012) The FEI2-SOS5 pathway and CELLULOSE SYNTHASE 5 are required for cellulose biosynthesis in the Arabidopsis seed coat and affect pectin mucilage structure. Plant Signaling & Behavior, 7(2): 285-288

Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome Research, 10(11): 1788-1795

Hijazi M, Velasquez SM, Jamet E, Estevez JM, Albenne C (2014) An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. Frontiers in Plant Science, 5:395

Huang H, Tudor M, Su T, Zhang Y, Hu Y, Ma H (1996) DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. Plant Cell 8(1) : 81-94.

Hugouvieux V, Silva CS, Jourdain A, ... Zubieta C (2018) Tetramerization of MADS family transcription factors SEPALLATA3 and AGAMOUS is required for floral meristem determinacy in Arabidopsis. Nucleic Acids Research,. 46(10): 4966-4977

Jibran R, Tahir J, Cooney J, Hunter DA, Dijkwel PP (2017) Arabidopsis AGAMOUS regulates sepal senescence by driving jasmonate production. Frontiers in Plant Science,. 8 : 2101

Johnson KL, Jones BJ, Bacic A, Schultz CJ (2003) The fasciclin-like arabinogalactan proteins of *Aarabidopsis*. A multigene family of putative cell adhesion molecules. Plant Physiology, 133: 1911-1925

Kaufmann K, Melzer R, Theißen G (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Genes, 347: 183–198

Kieliszewski MJ (2001) The latest hype on Hyp-O-glycosylation codes. Phytochemistry, 57: 319-323

Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS (2000) Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, beta ig-h3. Journal of Biological Chemistry, 275: 30907-30915

Kishor PBK, Kumari PH, Sunita MSL, Sreenivasulu N (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. Frontiers in Plant Science, 6

Klepikova A, Kasianov A, Gerasimov E, Logacheva M, Penin A (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. The Plant Journal, 88 (6): 1058-1070

Li J, Yu M, Geng L-L, Zhao J (2010) The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of *Arabidopsis*. The Plant Journal, 64: 482-497

Lopes AL, Costa ML, Sobral R, Costa MMR, Amorim MI, Coimbra S (2016) Arabinogalactan proteins and pectin distribution during female gametogenesis in *Quercus suber* L. Annals of Botany 117: 949-961

Ma Y, Chenchao Y, Huimin L, Wentao W, Yaxue L, Yuqian W, Qin C, Haoli M (2017) Bioinformatics prediction and evolution analysis of arabinogalactan proteins in the plant kingdom. Frontiers in Plant Science, 8: 66

MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus*. The Plant Journal, 62: 689-703

Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, Grossniklaus U, Kater MM, 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: 1702-1715

Mizzotti C, Ezquer I, Paolo D, Rueda-Romero P, Guerra RF, Battaglia R, Rogachev I, Aharoni A, Kater MM, Caporali E, Colombo L (2014) SEEDSTICK is a master regulator of development and metabolism in the *arabidopsis Arabidopsis* seed coat. Plos Genetics, 10

Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. The Plant Journal, 70: 409-420

Nawaschin S (1898) Resultate einer Revision der Befruchtungsvorgänge bei Lilium martagon und Fritillaria tenella. Bulletin de l'Académie Impériale des Sciences de St. Pétersbourg, 9: 377-382

Ng M, Yanofsky M (2001) Function and evolution of the plant MADS-box gene family. Nature Reviews Genetics, 2(3) : 186–195

Pereira AM, Pereira LG, Coimbra S (2015) Arabinogalactan proteins: rising attention from plant biologists. Plant Reproduction, 28: 1-15

Perry S, Lehti M,Fernandez D (1999) The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. Plant Physiology,. 120: 121-130

Pettolino F, Liao ML, Ying Z, Mau SL, Bacic A. (2006) Structure, function and cloning of arabinogalactan-proteins (AGPs): an overview. Food and Food Ingredients Journal of Japan, 211: 12–25

Rocheta M, Sobral R, Magalhães J, Amorim MI, Ribeiro T, Pinheiro M, Egas C, Morais-Cecílio L, Costa MRR (2014) Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. Frontiers in Plant Science 5: 599

Saedler H, Becker A, Winter KU, Kirchner C, Theissen G (2001) MADS-box genes are involved in floral development and evolution. Acta Biochimica Polonica, 48(2): 351-358

Santopolo S, Boccaccini A, Lorrai R, Ruta V, Capauto D, Minutello E, … Vittorioso P (2015) DOF AFFECTING GERMINATION 2 is a positive regulator of light-mediated seed germination and is repressed by DOF AFFECTING GERMINATION 1. BioMed Central Plant Biology, 15: 72

Schneitz K, Hulskamp M, Pruitt RE (1995) Wwild-type ovule development in *Arabidopsis thaliana* - a light-microscope study of cleared whole-mount tissue. The Plant Journal, 7: 731-749

Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. In Annual Review of Plant Biology, Vol 58, pp 137-161

Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiology, 153: 485-513

Smyth DR, Bowman JL, Meyerowitz EM (1990) Eearly flower development in *Aarabidopsis*. Plant Cell, 2: 755-767

Ståldal V, Cierlik I, Chen S, Landberg K, Baylis T, Myrenås M, Sundström JF, Eklund DM, Ljung K, Sundberg E (2012) The *Arabidopsis thaliana* transcriptional activator STYLISH1 regulates genes affecting stamen development, cell expansion and timing of flowering. Plant Molecular Biology,. 78(6) : 545-59

Stracke R, Werber M, Weisshaar, B (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. Current Opinion in Plant Biology, 4(5): 447-456

Su S, Higashiyama T (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. Plant Reproduction, 31: 67-75.

Sultana N, Florance H, Johns A, Smirnoff N (2015) Ascorbate deficiency and cell wall glycoproteome. Plant, Cell & Environment, 38: 375-384

Theißen G, Melzer R, Rümpler F (2016) MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development, 143: 3259-3271

Tsukagoshi H, Morikami A, Nakamura K (2007) Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in *Arabidopsis* seedlings. Proceedings of the National Academy of Sciences of the United States of America, 104(7) : 2543-2547

Tucker MR, Koltunow AMG (2014) Traffic monitors at the cell periphery: the role of cell walls during early female reproductive cell differentiation in plants. Current Opinion in Plant Biology, 17: 137-145

Tucker MR, Okada T, Hu Y, Scholefield A, Taylor JM, Koltunow AMG (2012) Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in *Arabidopsis*. Development, 139: 1399-1404

Ulrike Unte U, Anna-Marie S, Paolo P, Madhuri G, Dario L, Heinz S, Peter H (2003) SPL8, an SBP-Box gGene that affects pollen sac development in *Arabidopsis*. The Plant Cell, 15(4) : 1009-1019

Velasquez M, Salter JS, Dorosz JG, Petersen BL, Estevez JM (2012) recent advances on the posttranslational modifications of exts and their roles in plant cell walls. Frontiers in Plant Science, 3: 93

Wang H, Tang W, Zhu C, Perry S (2002) A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos. The Plant Journal, 32: 831-843

Xue H, Veit C, Abas L, Tryfona T, Maresch D, Ricardi MM, Manuel Estevez J, Strasser R, Seifert GJ (2017) *Arabidopsis thaliana* FLA4 functions as a glycan-stabilized soluble factor via its carboxy-proximal Fasciclin 1 domain. The Plant Journal, 91: 613-630

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

Yang J, Sardar HS, McGovern KR, Zhang Y, Showalter AM (2007) A lysine-rich arabinogalactan protein in *Arabidopsis* is essential for plant growth and development, including cell division and expansion. The Plant Journal, 49: 629-640

Yariv J, Lis H, Katchalski E (1967) Pprecipitation of arabic acid and some seed polysaccharides by glycosylphenylazo dyes. Biochemical Journal, 105 (1): 1C-2CC1-+

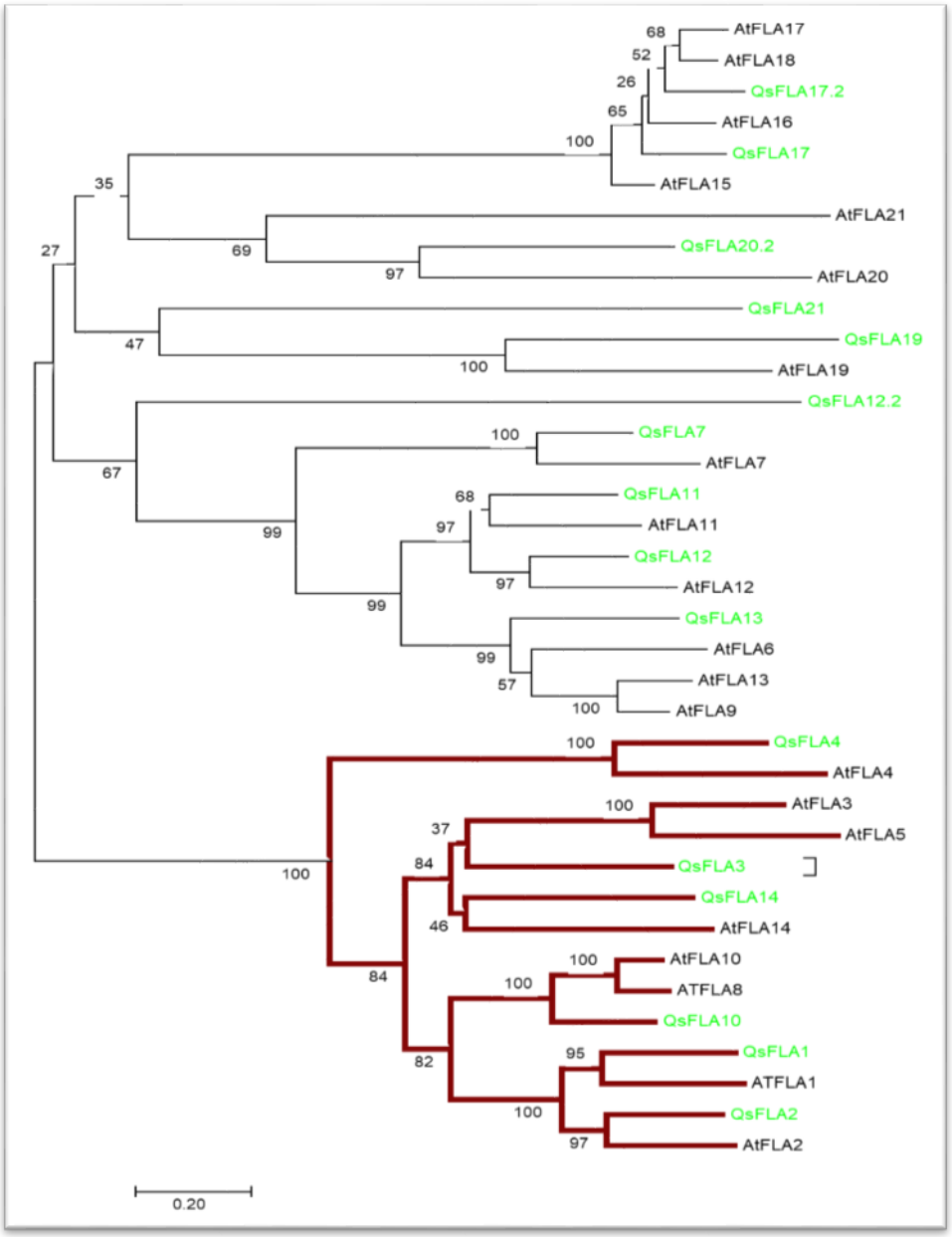
Zhang XL, Ma HL, Qi HD, Zhao J (2014) Roles of hydroxyproline-rich glycoproteins in the pollen tube and style cell growth of tobacco (*Nicotiana tabacum* L.). Journal of Plant Physiology, 171: 1036-1045

Zhang Y, Schwarz S, Saedler H, Huijser P (2007) SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in *Arabidopsis*. Plant Molecular Biology, 63(3): 429-39

Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE (2009) Global identification of targets of the *Arabidopsis* MADS domain protein AGAMOUS- Like15. Plant Cell, 21: 2563–2577

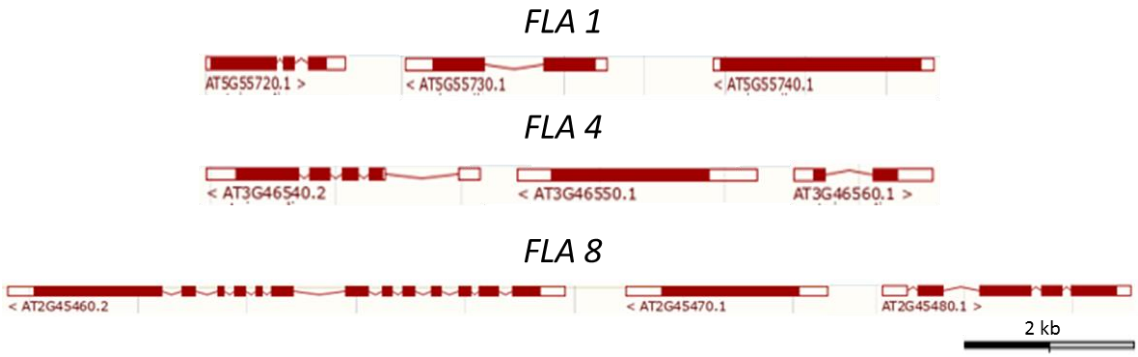
Zinn K, McAllister L, Goodman CS (1988) Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. Cell, 53: 577–587

7. Supplemental material



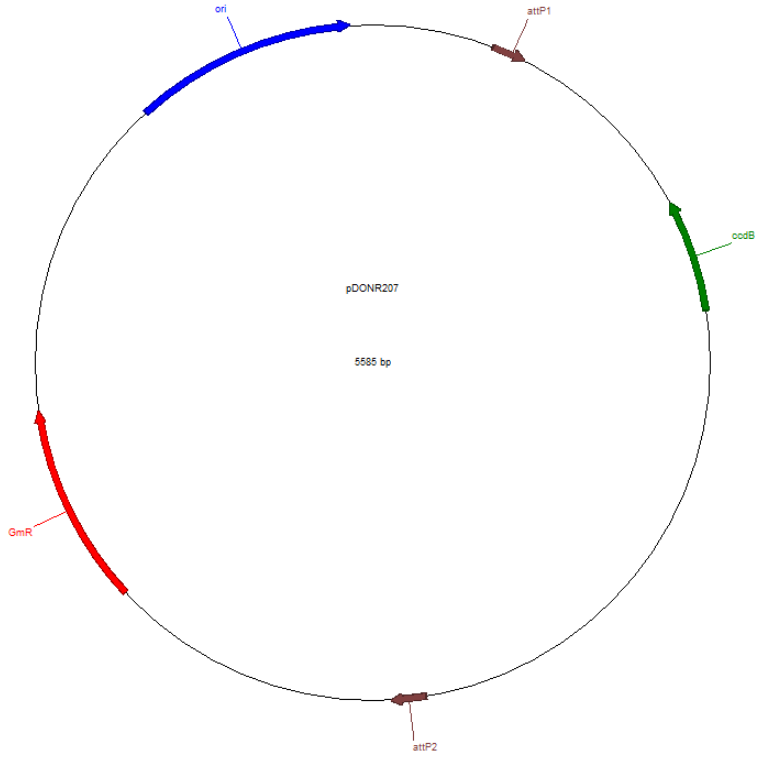
Supplemental Figure 1. Identification of putative *Arabidopsis* orthologous AGP genes of *Quercus suber.*

Neighbour-Joining tree based on 34 FLAs amino acid sequences of *A. thaliana* and *Q. suber*. The length of the branches refers to the amino acid variation rates. The numbers indicate the percentage of 1000 bootstrap re-samplings that support the inferred topology. The scale bar is an indicator of genetic distance based on branch length.



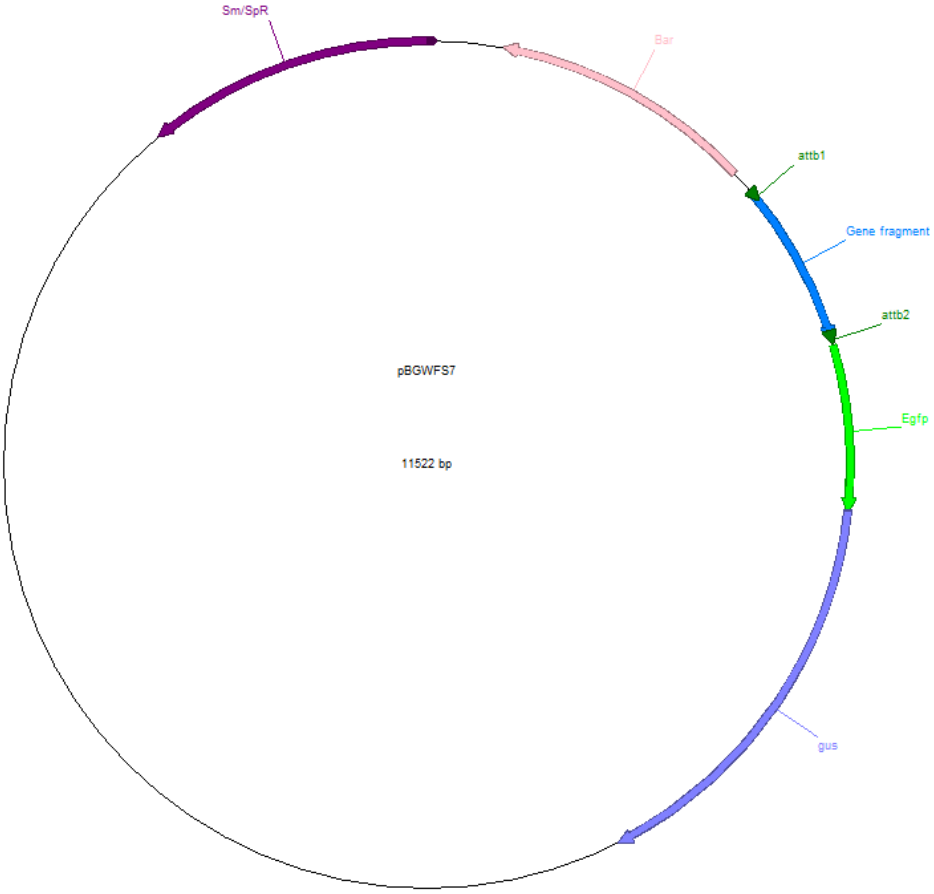
Supplemental Figure 2. Relative position of *FLA1*, *FLA4* and *FLA8* in relation to their neighbouring genes.

The position of *FLA1*, *FLA4* and *FLA8* was considered when deciding the length of the promoter region to be cloned into the plasmids used for the design of marker lines. The objective was to include as much as possible of the intergenic regions, without including parts of the neighbouring genes. UTR regions are in white; red regions represent the exons. Image taken from Ensembl Plants (http://plants.ensembl.org/index.html).



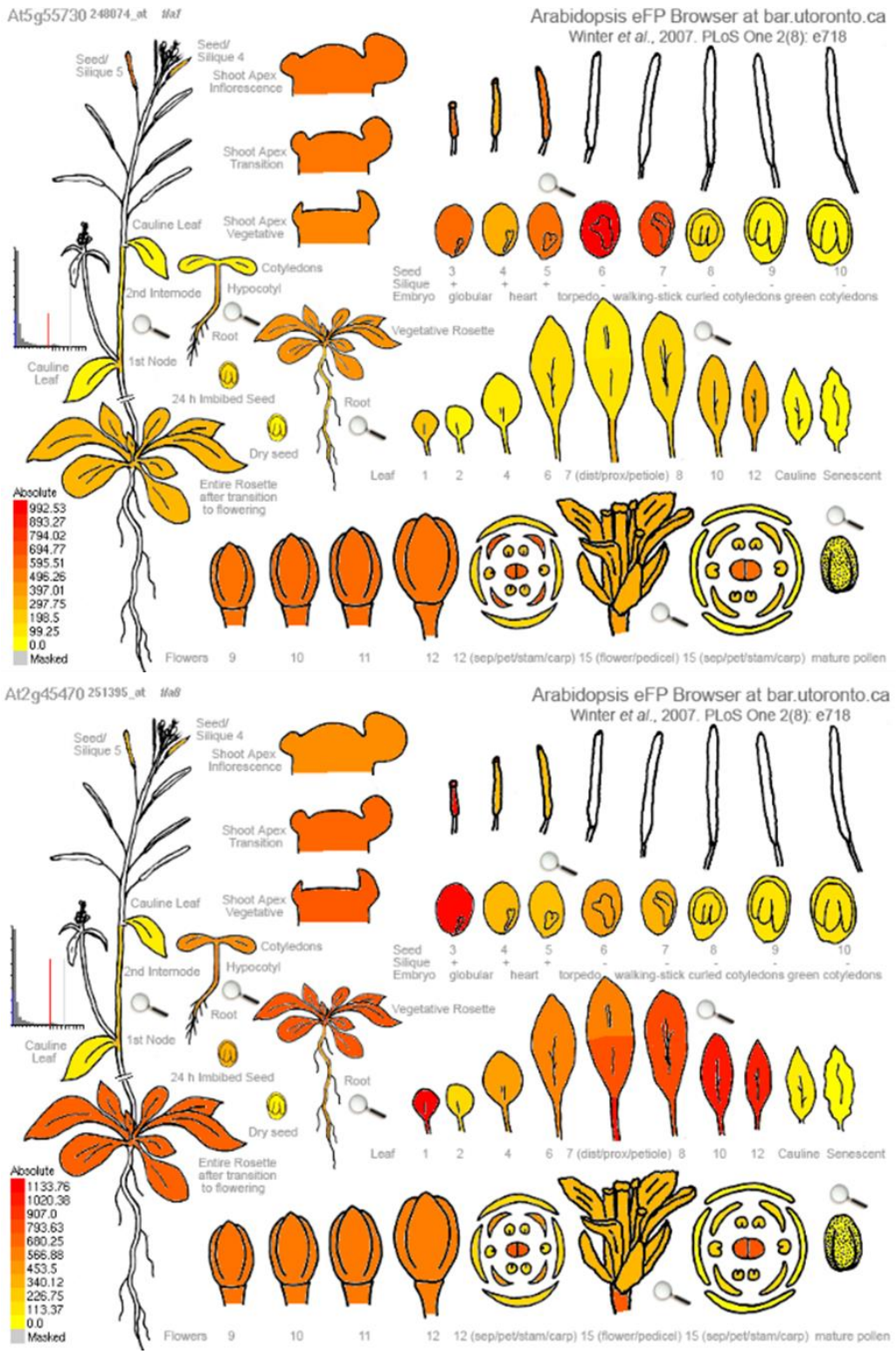
Supplemental Figure 3. Graphical representation of vector p*DONOR207* used for transformation of *A. thaliana.*

att1P1/attP2 - recombination sites; GmR - gene which confers resistance to gentamycin; ori - origin of replication; ccdB – bacterial toxin that poisons DNA gyrase. The cloned gene fragment (p*FLA1* and p*FLA4*) will replace the sequence between the two recombination sites of the plasmid during the BP reaction step of the Gateway recombination. Image obtained using ApE - A plasmid Editor program (v. 2.0.51).



Supplemental Figure 4. Graphical representation of vector p*BGWFS7* used for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*.

attB1/attB2 - recombination sites; Bar - gene which confers resistance to the herbicide BASTA; gus - β-glucuronidase gene; Gene fragment – the promoter regions of *FLA1*, *FLA4* and *FLA14* were added between the recombination sites; Egfp - encodes green fluorescent protein (already tested and shown to not be working); Sm/SpR - streptomycin-spectinomycin resistance cassette. Image obtained using ApE - A plasmid Editor program (v. 2.0.51).



Supplemental Figure 5. Expression patterns of *FLA8* and *FLA1* in *A.thaliana.*

Expression pattern in the various organs of *A.thaliana* during the different developmental stages of this species. Data from eFP Browser.