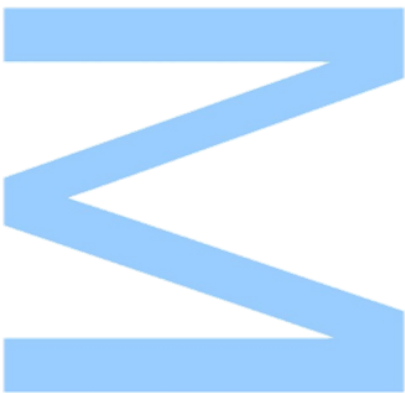
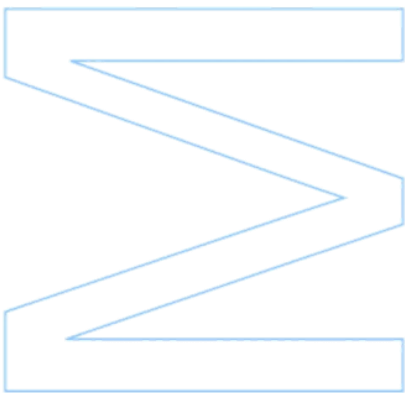
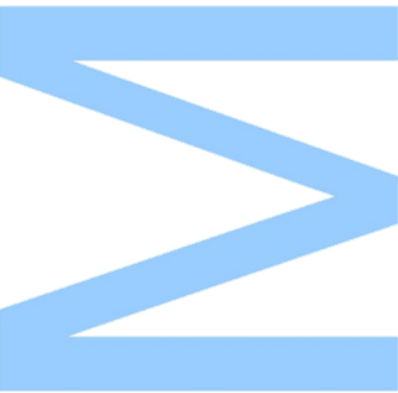


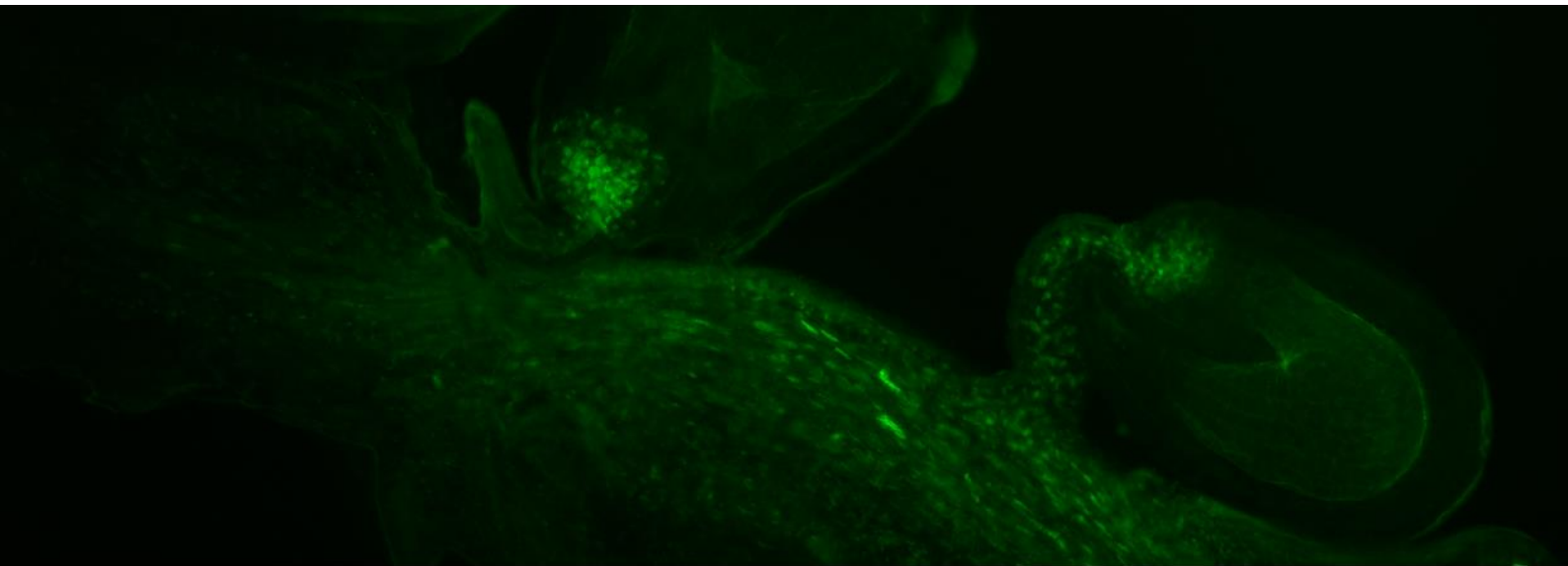


STK target genes crosstalk in *Arabidopsis* *thaliana* ovules and seed development

Ricardo André Campos Ferraz

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Área Científica
2020





STK target genes crosstalk in *Arabidopsis thaliana* ovules and seed development

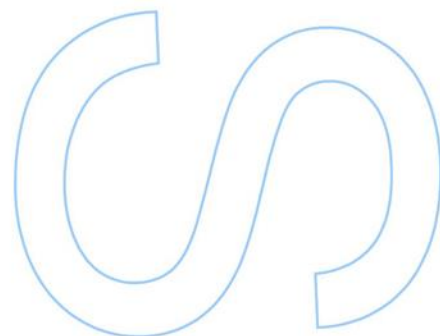
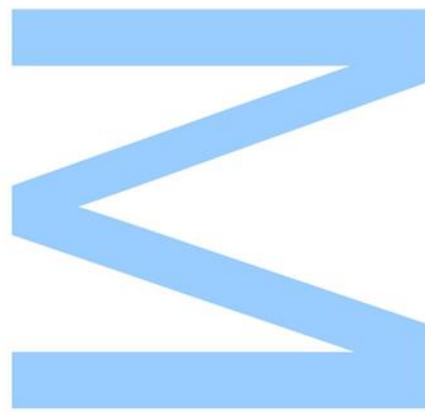
Ricardo André Campos Ferraz
Biologia Funcional e Biotecnologia de Plantas
Departamento de Biologia
2020

Orientador

Sílvia Coimbra, Professora, FCUP

Coorientador

Tetsuya Higashiyama, Professor, Nagoya University





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

O Presidente do Júri,

Assinado por : **MARIA FERNANDA DA SILVA**

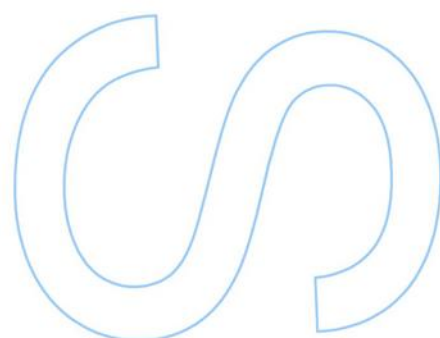
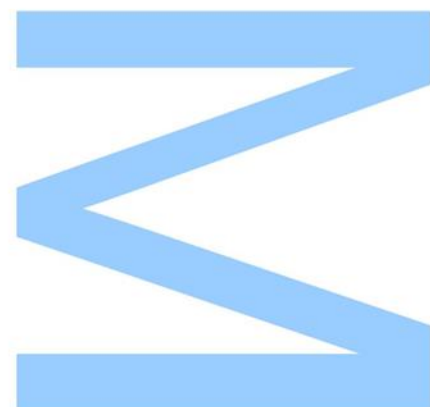
FIDALGO FERRO DE BEÇA

Num. de Identificação: B1037944649

Data: 2020.12.13 20:39:45+00'00'



CARTÃO DE CIDADÃO



Acknowledgments

Queria agradecer à Prof. Dr^a. Sílvia Coimbra por todas as oportunidades que me proporcionou desde o 2º ano de licenciatura, quando me aceitou no seu laboratório. O projeto SEXSEED foi uma incrível experiência para mim que me deu muitas bases para o futuro. Desde o conhecimento científico à experiência profissional, o SPReD lab foi realmente uma verdadeira aventura incrível que me ajudou a crescer muito profissionalmente e como ser humano. Professora, muito obrigado por toda a sua dedicação e empenho. Obrigado por ter sempre os seus estudantes como prioridade e por ter acreditado em mim e no meu trabalho.

Queria agradecer também à Ana Lopes. Ana, és a maior! A ajuda incondicional e sempre presente que me deste foi crucial. Mas foi muito mais do que isso. A tua disponibilidade para ajudar e apoio psicológico não tem limites. Ajudaste-me a ser mais organizado no laboratório e a não desistir. Mesmo à distância, enviaste-me sempre muita energia e carinho. Obrigado pela paciência e por toda a ajuda quando a coisa estava um pouco mais complicada, neste ano particularmente difícil! Tenho a certeza de que serás sempre grande, como já és! Quero dizer grande coisa, porque até és um pouco baixinha ;)! Gosto muito de ti e espero que corra tudo bem contigo, o Artur e o Eduardo.

Obrigado a todos os membros do SPReD lab por me terem sempre ajudado e me aturado quando tive dúvidas e quando fazia asneirinhas! Um agradecimento para a Maria João que me ajudou muito a orientar-me na Universidade de Nagoya e naquele país longínquo a que chamam de Japão. Obrigado pela paciência, sei que de vez em quando fui muito chato. Desejo-te tudo de bom Maria e continua sempre com o teu espírito aventureiro. Ah e obrigado por todas as ideias de sítios onde comer doces, isso é importante ;)!

I would also like to thank Dr. Tetsuya Higashiyama for accepting me in his laboratory, one of the best in the world in its field. It was a spectacular experience, both professionally and culturally. Thank you Dr. Hidenori Takeuchi for all the help and patience and for teaching me important techniques and skills. Thank you Bob! You are incredible! I know I asked you a lot of things, but you were always a good friend and helped me in everything I needed. And I would also like to thank all team members of Dr. Higashiyama lab for welcoming me so well and for all the parties!

Queria agradecer à minha namorada. Obrigado Bárbara por todo o apoio mesmo à distância. Sei que não foi fácil, mas estiveste sempre lá para me ajudar! O apoio psicológico e o incentivo que me deste foram muito importantes. Fui muito chato de vez em quando, mas deste-me sempre muito amor e carinho.

Muito obrigado aos meus dois melhores amigos, Francisca e André. Obrigado por todo o apoio nas horas mais complicadas e pelas palavras de força. Acima de tudo, obrigado pelo exemplo de trabalho e esforço que são e pela grande amizade que sempre tiveram comigo!

Finalmente queria agradecer a toda a minha família por todo o apoio incondicional desde sempre! Obrigado pais por acreditarem em mim, me apoiarem e terem confiança em mim e no meu trabalho. Tenho muito orgulho em vocês e são os melhores pais do mundo. Obrigado Miguel por me teres dado ânimo sempre que estive em casa com as tuas brincadeiras. E obrigado Rosa por aturares este sucateiro!

Resumo

A reprodução sexuada em plantas é um sistema complexo que envolve várias etapas. O fator de transcrição SEEDSTICK (STK) pertence à família MADSbox e está envolvido em muitos mecanismos moleculares relacionados com o desenvolvimento do óvulo e da semente. Análises recentes de CHIP-seq identificaram algumas proteínas arabinogalactânicas (AGPs) como possíveis genes alvo diretos de STK. Neste trabalho, os fenótipos de óvulos e sementes de plantas mutantes *stk* foram descritos. Sementes de *stk* apresentam um tamanho menor e uma forma mais arredondada do que sementes WT. Além disso, a zona de abscisão do funículo não constringe nesses mutantes. Quanto à produção de sementes e comprimento das silíquas, as plantas *stk* produzem menos sementes verdes maduras por silíqua e silíquas mais curtas. Os níveis de expressão dos genes *AGP9*, *AGP21*, *AGP24* e *AGP31*, potenciais genes alvos de STK, foram quantificados em plantas WT e *stk*. Os padrões de expressão desses genes foram descritos usando linhas repórter. Os padrões de expressão génica de *AGP24* e *AGP31* também foram observados em *stk*. A expressão de *AGP9* foi observada no pólo calázico de óvulos e sementes e nos tecidos vasculares do tecido de transmissão e do funículo. A expressão de *AGP21* foi detetada no septo e pólo micropilar de óvulos. A expressão de *AGP24* foi observada no pólo micropilar e calázico de óvulos nos estágios iniciais do desenvolvimento da flor e uma expressão intensa no endosperma em estágios posteriores. Em relação ao padrão de expressão do gene *AGP31*, este foi detetado no pólo calázico de óvulos e no tecido de transmissão. Mesmo que os resultados dos níveis de expressão de Real-time PCR não tenham sido conclusivos e necessitem de ser melhorados, eles apontam para uma regulação negativa do gene *AGP31* nas plantas mutantes *stk*, o que é sustentado pela ausência de expressão desse gene nas flores dos mutantes *stk*. A ausência de expressão no mutante *stk* foi também verificada para o gene *AGP24*, mas estes resultados precisam de ser validados. Além disso, foram iniciados os ensaios de desvio de mobilidade eletroforética (EMSA) e de transativação de protoplastos (PTA), com o objetivo de verificar se STK se liga diretamente a estes potenciais genes alvo. Finalmente, a técnica CRISPR-Cas9 foi realizada para criar mutantes duplos para *STK* e para os potenciais genes alvo. Embora ainda nenhuma planta mutante tenha sido obtida por esta técnica, o protocolo utilizado para a montagem dos plasmídeos recombinantes usados nessa técnica mostrou-se eficiente e rápida. O presente trabalho pretende aumentar o conhecimento existente sobre as redes de regulação génica presentes no desenvolvimento de óvulos e sementes.

Palavras-chave

A. thaliana, reprodução sexuada em plantas, redes de regulação génica, SEEDSTICK, proteínas arabinogalactânicas, EMSA, PTA, RT-qPCR, CRISPR-Cas9

Abstract

Sexual plants reproduction is a complex system and involves several steps. The SEEDSTICK (STK) transcription factor (TF) belongs to the MADSbox family and is involved in many molecular mechanisms related to ovule and seed development. Recent ChIP-sequencing analyses identified certain arabinogalactan proteins (AGPs) as possible direct target genes of STK. In this work, the phenotypes of *stk* mutant plants' ovules and seeds were described. Seeds from *stk* present a smaller size and a rounder shape than WT seeds. Moreover, the funiculus abscission zone does not constrict in these mutants. As for seed production and siliques' length, *stk* plants produce less mature green seeds per silique and shorter siliques. The expression levels of *AGP9*, *AGP21*, *AGP24* and *AGP31*, putative STK target genes, were quantified in WT and *stk* plants. The expression patterns of these genes were described using reporter lines. *AGP24* and *AGP31* gene expression patterns were also observed in *stk*. *AGP9* expression was observed in the ovules' and seeds' chalazal pole and in the vascular tissues of the transmitting tissue and funiculus. *AGP21* gene expression was detected in the septum and micropylar pole of ovules. The *AGP24* expression was observed in the ovules' micropylar and chalazal pole at early stages of flower development and an intense expression in the endosperm at later stages. In what concerns the expression pattern of *AGP31* gene, it was detected in ovules' chalazal pole and in the transmitting tissue. Even though the results of the expression levels by Real-time PCR were not clear and need to be improved, they point towards a down-regulation of the *AGP31* gene in the *stk* mutant plants, which is supported by the absence of an expression of this gene in the *stk* mutant flowers. This absence of expression in the *stk* mutant was also verified for the *AGP24* gene, but these results needs validation. Furthermore, Electrophoretic Mobility Shift Assay (EMSA) and Protoplast Transactivation Assay (PTA) techniques were initiated, intending to verify if STK binds directly to these potential target genes. Finally, the CRISPR-Cas9 technique was performed to create double mutants for *STK* and the putative target genes. Although no mutant plant has yet been obtained by this technique, the protocol used to assemble the recombinant plasmids employed in this technique proved to be efficient and fast. The present work intends to increase the existing knowledge related to the gene regulatory networks present in ovule and seed development.

Key words

A. thaliana, plant sexual reproduction, gene regulatory network, SEEDSTICK, arabinogalactan proteins, EMSA, PTA, RT-qPCR, CRISPR-Cas9

Index

List of figures and tables	IX
List of abbreviations and acronyms.....	XIV
1.Introduction	1
1.1. <i>Arabidopsis thaliana</i> as a model organism	1
1.2. <i>Arabidopsis thaliana</i> reproduction process	2
1.3. <i>Arabidopsis thaliana</i> seed development	3
1.4.SEEDSTICK control of ovule development.....	4
1.5.STK targets: crosstalks behind <i>Arabidopsis</i> reproduction	9
1.5.1.STK and cytokinin oxidase/dehydrogenase enzymes	9
1.5.2.STK and arabinogalactan proteins	12
1.6.Gene expression networks in plant reproduction.....	16
1.6.1.Study TFs direct regulation of target genes expression <i>in vitro</i>	16
1.6.2.Study TFs direct regulation of target genes expression <i>in</i> <i>planta</i>	18
1.7.Study gene function: genome editing through CRISPR-CAS9.....	20
2.Objectives	23
3.Materials and methods	24
3.1.Plant material and growth conditions	24
3.2.Genotyping.....	24
3.3.Gene expression analysis through RT-qPCR	25
3.3.1.RNA extraction	25
3.3.2.RNA quantity and integrity analysis.....	26
3.3.3.cDNA synthesis.....	26
3.3.4.Semi-quantitative and quantitative RT-PCR	27
3.4. <i>In silico</i> analyses.....	28
3.4.1.CArG boxes and <i>cis</i> elements prediction	28
3.4.2.sgRNAs design for the CRISPR-Cas9 technique	30
3.5.Electrophoretic Mobility Shift Assay	31
3.5.1.Introduction to the technique	31
3.5.2.STK coding sequence amplification	32
3.5.3.Plasmid and amplified fragment double digestion and plasmid dephosphorylation.....	33
3.5.4.Ligation and restriction analysis	34
3.6.Protoplast Transactivation Assay	35

3.6.1.STK gene cloning.....	35
3.6.2.cis elements identification.....	37
3.6.2.1.Amplify putative cis elements	37
3.6.2.2.Plasmid miniprep and restriction analysis	38
3.7.Creation of knockout mutant plants using the CRISPR-Cas9 technique	39
3.7.1.Assembly of constructs with two sgRNA cassettes for plant transformation	39
3.7.1.1.Bacterial strains and vectors used.....	39
3.7.1.2.Growth media and conditions	39
3.7.1.3.Acquisition of sgRNA cassettes within an amplified template.....	39
3.7.1.4.DNA fragments extraction from gel and purification.....	42
3.7.1.5.Golden Gate Assembly	42
3.7.1.6. <i>E. coli</i> Mach1 and <i>A. tumefaciens</i> transformation.....	44
3.7.1.7.Colony-PCR	44
3.7.1.8.Plasmid DNA extraction	45
3.7.1.9.Sequencing	45
3.7.2.Plant transformation by floral dip mediated by <i>A. tumefaciens</i>	46
3.7.3.Transformed plants selection and identification of genome edited plants...	47
3.8.Controlled crosses	48
3.9.Phenotype characterisation	49
3.9.1.Sample preparation for microscopy.....	49
3.9.2.Clear <i>A. thaliana</i> flowers and siliques and ovule microscopic observation .	49
3.9.3.Aniline blue staining of <i>A. thaliana</i> PTs.....	49
3.9.4.Histochemical detection of GUS activity.....	50
3.9.5.Observation of <i>A. thaliana</i> reproductive tissues in a fluorescence microscope	50
3.9.6.Seed set analysis.....	50
4.Results	52
4.1.Genotyping.....	52
4.2.Gene expression analysis: from RNA extraction to RT-qPCR	53
4.3.STK direct interaction with potential target genes	59
4.3.1.Electrophoretic Mobility Shift Assay	59
4.3.2.Protoplast Transactivation Assay	61
4.4.Creation of knockout mutant plants using the CRISPR-Cas9 technique	65
4.4.1.Assembly of constructs for plant transformation	65
4.4.1.1.DNA template amplification.....	65

4.4.1.2.Colony-PCR	66
4.4.1.3.Sequencing	67
4.4.2. <i>Agrobacterium tumefaciens</i> and plant transformation	67
4.5.Phenotype characterisation	70
4.5.1.WT and <i>stk</i> phenotypes	70
4.5.1.1.Arabidopsis flowers and siliques clearing	70
4.5.1.2. <i>In vivo</i> PT growth assay	72
4.5.1.3.Seed set analysis	73
4.5.2.Putative AGPs target genes expression pattern	76
4.5.3.Putative AGPs target genes expression pattern in <i>stk</i> mutants	80
5.Discussion	83
5.1.Characterise SEEDSTICK functions in ovule and seed development	83
5.2.Identify STK target genes	85
5.3.STK direct interaction with the putative AGPs target genes.....	88
5.4.Create double mutants for <i>STK</i> and each one of its putative AGPs target genes	90
6.Conclusion	92
7.References	94
8.Annexes.....	102
Annex I	102
Annex II	103
Annex III	104
Annex IV.....	105
Annex V.....	106
Annex VI.....	107
Annex VII.....	109
Annex VIII.....	112

List of figures and tables

Figure 1. Schematic representation of the PT growth through the pistil's tissues in <i>A. thaliana</i>	3
Figure 2. Development and biological functions of Arabidopsis seeds' sub-regions.....	4
Figure 3. Arabidopsis <i>STK</i> characterisation.	7
Table 1. <i>STK</i> proposed interactions and functions throughout ovule and seed development.	8
Figure 4. Localisation of <i>AtCKX6:GUS</i> activity.	10

Figure 5. Gus activity in <i>CKX7:GUS</i> Arabidopsis transgenic plants.	11
Figure 6. Evidences of STK directly regulation of <i>CKX7</i> expression and the influence of this gene in silique elongation.	11
Figure 7. Schematic representation of an Arabidopsis stamen and ovule, showing the different AGPs present along the PT path, in the pollen grain and in the endosperm.	15
Figure 8. Autoradiography of a polyacrylamide gel that resulted from an EMSA.	17
Figure 9. Results from Arabidopsis protoplast transactivation.	20
Figure 10. Mechanism of action of the CRISPR-Cas9 technique.	22
Table 2. List of the sequences of the primers used in the genotyping PCR.	25
Table 3. Genotyping PCR mixture composition.	25
Table 4. Conditions applied in the genotyping PCR with WT and <i>stk</i> mutant lines and GUS reporter lines.	25
Table 5. Genomic DNA degradation and reverse transcription reactions mixture composition.	27
Table 6. List of the sequences of the primers used in the semi-quantitative RT-PCR.	27
Table 7. Semi-quantitative RT-PCR mixture composition.	27
Table 8. Conditions applied in the semi-quantitative RT-PCR.	28
Table 9. List of the sequences of the primers used in the RT-qPCR.	28
Table 10. Conditions applied in the RT-qPCR.	28
Figure 11. Schematic representation of the position of the twelve selected CARG boxes (C1 to C12) in a section of the promoter region near the ATG (1000 bp upstream of the ATG) of the <i>AGP9</i> (a)), <i>AGP24</i> (b)) and <i>AGP31</i> (c)) genes.	29
Table 11. Sequences of the twelve selected CARG boxes (C1 to C12 in Annex I, Supplemental table 1) present in the <i>AGP9</i> , <i>AGP24</i> and <i>AGP31</i> genes.	30
Table 12. List of the sgRNAs specific sequences selected for the <i>AGP1</i> , <i>AGP4</i> , <i>AGP9</i> , <i>AGP15</i> , <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes.	31
Table 13. Sequences of the primers designed for the <i>STK</i> coding sequence amplification and cloning into the pET28a (+) expression vector.	32
Table 14. Conditions applied in the PCR for the <i>STK</i> coding sequence amplification.	33
Table 15. Restriction reaction mixture composition.	34
Table 16. Ligation reactions mixtures composition.	34
Table 17. Restriction analysis mixture composition.	34
Table 18. Restriction reaction mixture composition.	35
Table 19. Sequences of the primers designed for the <i>STK</i> genomic sequence amplification and cloning into the MU1110 vector.	36

Table 20. PCR mixture composition for the amplification of the <i>STK</i> genomic sequence.	36
Table 21. Conditions applied in the PCR for the <i>STK</i> genomic sequence amplification.	36
Table 22. Ligation reaction mixture composition.	37
Table 23. Restriction analysis mixture composition.	37
Table 24. Sequences of the primers designed for the <i>AGP24</i> and <i>AGP31</i> genes predicted cis elements amplification and cloning into pBluescript II SK (+).	38
Table 25. Reaction mixture composition and conditions of the restriction analysis of the plasmid DNA using KpnI restriction enzyme (Thermo Scientific™).	39
Table 26. Sequences of the primers designed for the amplification of the DNA template containing sgRNAs.	40
Figure 12. Schematic representation of: a) the forward primer (1.) and the reverse primer (2.); b) the region of the template that was amplified; c) the whole amplified fragment (Wang <i>et al.</i> , 2015).	41
Table 27. PCR mixture composition for the amplification of the DNA template containing sgRNAs.	41
Table 28. PCR conditions for the amplification of the DNA template containing sgRNAs and expected band size.	42
Table 29. Golden Gate assembly reaction composition.	43
Table 30. Golden Gate assembly reaction conditions.	43
Figure 13. Schematic representation explaining the Golden Gate assembly.	43
Table 31. Sequences of the primers for colony-PCR.	45
Table 32. Colony-PCR mixture composition.	45
Table 33. Colony-PCR conditions.	45
Table 34. Primers used in the sequencing reaction.	46
Table 35. Sequencing reaction mixture composition.	46
Table 36. Sequencing reaction conditions.	46
Table 37. Sequences of the primers used in the amplification of the target gene fragment that includes the sgRNAs specific cassettes.	48
Table 38. PCR conditions used for the amplification of the target gene fragment that includes the sgRNAs specific cassettes.	48
Figure 14. Electrophoresis of the PCR products that correspond to the genotyping of 7 WT plants (1 to 7) and 7 <i>stk</i> mutant plants (8 to 14).	52
Figure 15. Electrophoresis of the PCR products that correspond to the genotyping of 6 transgenic plants for the pAGP21:GUS construct (1 to 6).	53

Figure 16. Electrophoresis of the seed RNA samples from 2 WT plants (WTI and WTII) and 2 <i>stk</i> mutant plants (<i>stkI</i> and <i>stkII</i>).	54
Figure 17. Electrophoresis of the semi-quantitative RT-PCR products using cDNA samples from 2 WT (WI, WII, wI and wII) plants and 2 <i>stk</i> mutant plants (SI, SII, sI and sII).	54
Figure 18. Relative normalised expression of the <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes in WTI, <i>stkI</i> and <i>stkII</i> seeds obtained in the first RT-qPCR, defining WTI sample as the control sample.	55
Figure 19. Relative normalised expression of the <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes in WTII, <i>stkI</i> and <i>stkII</i> seeds obtained in the first RT-qPCR, defining WTII sample as the control sample.	55
Figure 20. Relative normalised expression of the <i>AGP9</i> , <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes in WTI, <i>stkI</i> and <i>stkII</i> seeds obtained in the second RT-qPCR, defining WTI sample as the control sample.	56
Figure 21. Relative normalised expression of the <i>AGP9</i> , <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes in WTII, <i>stkI</i> and <i>stkII</i> seeds obtained in the second RT-qPCR, defining WTII sample as the control sample.	56
Figure 22. Electrophoresis of the seed RNA samples from 1 WT plant (WT) and 3 <i>stk</i> mutant plants (<i>stkI</i> , <i>stkII</i> and <i>stkIII</i>).	57
Figure 23. Electrophoresis of the semi-quantitative RT-PCR products using cDNA samples from 1 WT plant (WT) and 3 <i>stk</i> mutant plants (<i>stkI</i> , <i>stkII</i> and <i>stkIII</i>).....	58
Figure 24. Relative normalised expression of the <i>AGP9</i> , <i>AGP21</i> , <i>AGP24</i> and <i>AGP31</i> genes in WT, <i>stkI</i> and <i>stkIII</i> samples obtained in the third RT-qPCR, defining the WT sample as the control sample.	58
Figure 25. Electrophoresis of the PCR product that amplified the <i>STK</i> coding sequence (<i>STK</i>).....	59
Figure 26. Electrophoresis of the products of the pET28a plasmid and <i>STK</i> coding sequence restriction.	60
Figure 27. Electrophoresis of the PCR product that amplified the <i>STK</i> genomic sequence (<i>STK</i>).....	61
Figure 28. Electrophoresis of the restriction reactions' products that digested the MU1110 vector (MU) and the <i>STK</i> genomic sequence (<i>STK</i>).	62
Figure 29. Electrophoresis of the PCR products of the putative <i>AGP24</i> gene <i>cis</i> elements (promoter and 5' UTR) amplification reactions.	63
Figure 30. Electrophoresis of the PCR products of the putative <i>AGP31</i> gene <i>cis</i> elements (promoter, 5' UTR and 3' UTR) amplification reactions.....	64

Figure 31. Electrophoresis of the products of the pBluescript II SK (+) plasmid restriction analysis.....	65
Figure 32. Electrophoreses of the PCR products that resulted from the amplification of the DNA templates.	66
Figure 33. Electrophoreses of the products of the colony-PCR performed to identify the <i>E. coli</i> clones transformed with the vectors assembled for the knockout of: a) <i>AGP1</i> and <i>AGP15</i> ; b) <i>AGP24</i> and <i>AGP31</i> (1st pair of sgRNAs specific cassettes [<i>AGP31</i> (1)]); c) <i>AGP31</i> (2nd pair of sgRNAs specific cassettes [<i>AGP31</i> (2)]); <i>AGP4</i> and <i>AGP9</i> (1st pair of sgRNAs specific cassettes [<i>AGP9</i> (1)]); d) <i>AGP9</i> (2nd pair of sgRNAs specific cassettes [<i>AGP9</i> (2)]) and <i>AGP21</i>	67
Figure 34. Electrophoresis of the products of the colony-PCR performed to identify the <i>Agrobacterium</i> clones transformed with the vectors assembled for the knockout of <i>AGP9</i> (9.1 and 9.2), <i>AGP21</i> (21.1 and 21.2), <i>AGP24</i> (24.1 and 24.2) and <i>AGP31</i> (31.1 and 31.2).	68
Figure 35. Electrophoreses of the products of the amplification reactions that used genomic DNA from the germinated plants in order to amplify fragments that comprise the sequences homologous to the four sgRNAs specific cassettes designed for the <i>AGP9</i> (a) 9.1 to 9.5), <i>AGP21</i> (b) 21.1 and 21.2) and <i>AGP31</i> (c) 31.1 and d) 31.2) genes.	70
Figure 36. Clearing with chloral hydrate of WT (a) and c)) and <i>stk</i> (b), d) – f)) ovules. .	71
Figure 37. Clearing with chloral hydrate of seeds of WT (a) and b)) and <i>stk</i> (c) and d)) siliques extracted around stage 15 of flower development (b) and d)) and stage 17 of flower development (a) and c)).	72
Figure 38. Aniline blue staining of PTs in WT (a) – e)) and <i>stk</i> (f) – h)) pistils.	73
Figure 39. Graphics showing the results of seed set and silique length when using ten random siliques from a group of WT plants and another ten random siliques from a group of <i>stk</i> plants as the samples being studied.	74
Figure 40. Graphics showing the results of seed set and silique length when using three WT plants and three <i>stk</i> mutant plants, analysing ten random siliques from each plant.	75
Figure 41. Fluorescence microscopy of ovules and seeds of plants transgenic for the <i>pAGP9:NLS:3GFP</i> construct.	77
Figure 42. Histochemical detection of GUS activity regulated by the <i>AGP21</i> gene promoter in ovules and seeds.	78
Figure 43. Histochemical detection of GUS activity regulated by the <i>AGP24</i> gene promoter in ovules and seeds.	79

Figure 44. Histochemical detection of GUS activity regulated by the <i>AGP31</i> gene promoter in ovules and seeds.	80
Figure 45. Histochemical detection of GUS activity regulated by the <i>AGP24</i> gene promoter in ovules and seeds of <i>stk</i> mutant plants.	81
Figure 46. Histochemical detection of GUS activity regulated by the <i>AGP31</i> gene promoter in ovules and seeds of <i>stk</i> mutant plants.	82
Figure 47. Histochemical detection of GUS activity regulated by the <i>AGP31</i> gene promoter in nectaries of <i>stk</i> mutant plants.	82
Supplemental table 1. Sequences of the oligonucleotides designed for the EMSA technique containing the CArG boxes of the <i>AGP9</i> , <i>AGP24</i> and <i>AGP31</i> genes.....	102
Supplemental figure 1. Schematic representation of the pET28a (+) vector.	103
Supplemental figure 2. Schematic representation of the MU1110 vector.....	104
Supplemental figure 3. Schematic representation of the pBluescript II SK (+) vector.	105
Supplemental table 2. PCR conditions applied in the amplification of the <i>AGP24</i> (promoter and 5' UTR) and <i>AGP31</i> (promoter, 5' UTR and 3' UTR) predicted <i>cis</i> elements.	106
Supplemental figure 4. Schematic representation of the template plasmid pCBC-DT1T2 (Xing <i>et al.</i> , 2014).	107
Supplemental figure 5. Schematic representation of the backbone vector pHEE401E+Blii.	108
Supplemental figure 6. Results of the sequencing of the plasmid DNA extracted from colonies transformed with the recombinant plasmid transformed with the vectors assembled for the knockout of the <i>AGP1</i> (a) – d)), <i>AGP4</i> (e) – h)), <i>AGP9</i> (i) – l)), <i>AGP15</i> (m) – p)), <i>AGP21</i> (q) – t)), <i>AGP24</i> (u) – x)) and <i>AGP31</i> (y) – ab)) genes.....	109
Supplemental figure 7. Sequencing of fragments that comprise the sequences homologous to the four sgRNAs specific cassettes designed for the <i>AGP9</i> (a) to e)) and <i>AGP31</i> (f) to g)) genes, amplified using genomic DNA of the plants that germinated following the transformation processes.....	112

List of abbreviations and acronyms

6× His – Hexa-Histidine tag	AG – <i>Agamous</i>
a – antipodals	AG – Arabinogalactans
<i>aadA1</i> – <i>Aminoglycoside</i> (3")	AG peptides – Arabinoglactan peptides
<i>adenylyltransferase</i>	AGP – Arabinogalactan proteins
ABS – <i>Arabidopsis B Sister</i>	<i>AmpR</i> – <i>Ampicillin resistance</i>
ACT8 – <i>Actin 8</i>	ANOVA – Analysis of Variance

ANT – AINTEGUMENTA

az – seed abscission zone

BAN – BANYLUS

BEL1 – BELL 1

bp – base pair

Cas9 – CRISPR associated protein 9

cc – central cell

cDNA – complementary DNA

CFP – Cyan Fluorescent Protein

ch – chalazal pole

ChIP-seq – Chromatin

Immunoprecipitation sequencing

CK – Cytokinin

CKX – Cytokinin

Oxidase/Dehydrogenase

Col-0 – Columbia 0 variety

CRISPR – Clustered Regularly

Interspaced Palindromic Repeats

crRNA – CRISPR RNA

CZE – Chalazal Endosperm

CZSC – Chalazal Seed Integument

DAB – Decolourised Aniline Blue

DIC – Differential Interference Contrast

DNA – Deoxyribonucleic acid

DSB – Double-Strand Break

ec – egg cell

EC1.1p – Egg Cell specific promoter

EC1.2en – Egg Cell specific enhancer

EDTA – Ethylenediamine Tetraacetic Acid

EMSA – Electrophoretic Mobility Shift Assay

EP – Embryo Proper

FBP – FLOWER BINDING PROTEIN

FISH – Fluorescence *in situ*

Hybridisation

FLA – Fasciclin-Like AGP

fu – funiculus

GAL4 AD – GAL4 Activation Domain

GFP – Green Fluorescent Protein

GPI – Glycosylphosphatidylinositol lipidic anchor

GUS – β -glucuronidase

HA tag – human influenza hemagglutinin tag

HDR – Homology Directed Repair

Hyg – Hygromycin

i, in, it – integument

INO – INNER NO OUTER

IPTG – Isopropyl β -D-1-thiogalactopyranoside

KanR, *KmR* – Kanamycin Resistance

LAB – Lithium Acetate Borate buffer

lacI – *lac* repressor

LB medium – Luria Bertani medium

LRE – LORELEI

Lys-AGP – Lysine-rich AGP

MCE – Micropylar Endosperm

MCS – Multiple Cloning Site

MES – 4-morpholineethanesulfonic acid

mi – micropyle

MMC – Megaspore Mother Cell

MS medium – Murashige and Skoog medium

MYB98 – MYB DOMAIN PROTEIN 98

NHEJ – Non Homologous End-Joining

NLS – Nuclear Localisation Signal

NOST – *Nopaline Synthase* Transcription Termination

ON – overnight

ori – standard origin of replication

ov – ovule

PA – pro-anthocyanidins	sy – synergids
PAM – Protospacer Adjacent Motif	TAE – Tris-acetate-EDTA buffer
PCR – Polymerase Chain Reaction	T-DNA – Transfer DNA
PEN – Peripheral Endosperm	TEAMP – Transient Expression in
pg – pollen grain	Arabidopsis Mesophyll Protoplast
<i>PME16 – PECTIN METHYLESTERASE</i>	TF – transcription factor
<i>INHIBITOR 6</i>	tracrRNA – trans-activating crRNA
Pol – polymerase	tt – transmitting tissue
PT, pt – pollen tube	U6-26p – U6 small nucleolar RNA 26
PTA – Protoplast Transactivation Assay	promoter
rbcS-E9t – <i>rbcS E9 (Pea sativum ninth</i>	U6-26t – U6 small nucleolar RNA 26
<i>element of the Rubisco small subunit)</i>	terminator
terminator	U6-29p – U6 small nucleolar RNA 29
RBS – Ribosome-Binding Site	promoter
<i>REM – REPRODUCTIVE MERISTEM</i>	UTR – Untranslated Region
RNA – Ribonucleic acid	v – vasculature
<i>ROP – REPRESSOR OF PRIMER</i>	VAL – VALKYRIE
RT-PCR – Reverse transcriptase PCR	VDD – VERDANDI
RT-qPCR – Quantitative reverse	WOX8 – WUSCHEL-RELATED
transcriptase PCR	HOMEBOX 8
<i>RUB1 – Ubiquitin-Related Protein 1</i>	WT – Wild Type
SAT2 – STREPTOTHRICIN	Y1H – Yeast One-Hybrid
ACETYLTRANSFERASE 2	YFP – Yellow Fluorescent Protein
SB – Sodium Borate buffer	zCas9 – <i>Zea mays</i> codon-optimised
SC – Distal Seed Integument	Cas9
sc – stigmatic cell	
SDS – Sodium Dodecyl Sulfate	
SEP – SEPALLATA	
sgRNA – single guide RNA	
<i>SHP – SHATERPROOF</i>	
SOC medium – Super Optimal Broth	
medium	
sp – septum	
st – style	
STK – SEEDSTICK	
SUS – Suspensor	

1.1. Introduction

The dissertation project here presented is integrated in the SEXSEED project, funded by the EU Horizon 2020 Research and Innovation Staff Exchange programme. This project intends to unravel some of the mechanisms and gene interactions controlling sexual plant reproduction and seed development, more specifically the ones related to SEEDSTICK (STK), a MADSbox TF highly involved in these processes.

1.1. *Arabidopsis thaliana* as a model organism

The ovule and seed development in *Arabidopsis thaliana* is well described physiologically. However, the regulatory networks involved still need to be completely unravelled. Part of the Brassicaceae family and with its origin in West Eurasia, *Arabidopsis thaliana* (L.) Heynh. is characterised by being a long-day facultative, self-pollinating, small plant and by presenting a leaf rosette in the base as well as a corymb inflorescence above it, whose both pedicels and axis contain leaves that present branched trichomes. As for the *Arabidopsis* flower, it presents four green sepals, four white petals, six stamens and a pistil which will form *Arabidopsis* fruit, the silique. This species presents a short life cycle that only lasts for six weeks (Blamey and Grey-Wilson, 1989; Meinke *et al.*, 1998).

Apart from its own characteristics, such as being a simple, fast growing organism and a great seed producer, the scientific community chose *A. thaliana* as a model organism for genetic studies in Angiosperms largely due to many advances in the study of this organism, mainly: i) the completion of the whole genome sequencing in the year 2000 ii) the rise of efficient mutagenesis, crossing and DNA sequence insertion methods; iii) the existence of a highly broad mutants compilation for many phenotypes and of T-DNA insertion transgenic lines throughout the whole genome; iv) the easy obtention of a high number of homozygotic seeds for a recessive allele by self-fertilisation; v) the existence of chromosomes' genetic maps containing the identification of several molecular markers; vi) the ever increasing number of different genetic analyses made in *A. thaliana* that go from epigenetics to reverse genetics (Meinke *et al.*, 1998). Furthermore, being *Arabidopsis* a model organism, the discoveries made in this organism can be easily extrapolated to other plant species, such as economically relevant cereal species, therefore the use of *Arabidopsis* in many genetic studies (Serino *et al.*, 2018).

1.2. *Arabidopsis thaliana* reproduction process

Multicellularity makes fertilisation in angiosperms one of the most complex fertilisation processes. With regards to the formation of the male gametophyte, inside the anther locus, each microspore mother cell goes through meiosis and forms four haploid spores, each of which undergoes a mitotic division, originating a vegetative cell and a generative cell, being the second cell enveloped by the first. This process is followed by a second mitosis in the generative cell, which creates two male gametes, the sperm cells. Altogether, the two sperm cells and the vegetative cell form the male gametophyte, the pollen grain (McCormick, 2004; Dresselhaus and Coimbra, 2016).

While the male gametophyte develops within the anthers, the female gametophyte, the embryo sac, forms inside the ovule. The ovules arise from the internal surface of the carpel, in the placenta, emerging as a small, fingertip-like structure, presenting three regions: the funiculus, which connects the ovule to the placenta, the chalaza, the central zone where the integuments are originated, and the nucellus, that contains the megaspore mother cell (MMC). Both the inner and the outer integuments grow and surround the nucellus, protecting it throughout the ovule development, carrying nutrients from the mother plant to the embryo sac (Schneitz *et al.*, 1995). In order to originate the female gametophyte, initially the MMC forms four haploid megaspores, from which three degenerate and only one survives, the functional megaspore. This megaspore goes through three successive mitosis, giving rise to eight nuclei organised in seven cells, that compose the female gametophyte: two synergids and an egg cell in the micropylar pole, a binucleate central cell and three antipodals in the chalazal pole that ultimately degenerate in *Arabidopsis* (Schneitz *et al.*, 1995; Dresselhaus and Coimbra, 2016). The synergid cells present several wall invaginations at the micropylar pole, called the filiform apparatus, whose functions are involved in pollen tube (PT) attraction and reception (Takeuchi and Higashiyama, 2012).

In Angiosperms, two fertilisation events constitute a process called double fertilisation. After the release of the pollen grain by the anther and its arrival at the stigma, the pollen grain becomes hydrated and germinates forming the PT. This structure grows through the stigma and the style in order to reach the transmitting tissue. This growth occurs only at the extracellular matrix of these tissues. When arriving at the placenta next to a funiculus, the PT curves and grows in the surface of the funiculus towards the micropyle, in order to enter the embryo sac through one of the synergid cells (Figure 1). At this moment, the PT bursts, ejecting the two sperm cells. While one of the sperm cells merges with the egg cell forming the embryo, the other sperm cell combines with the central cell diploid nucleus, originated by the fusion of its two previous nuclei, creating the first

nucleus of the triploid endosperm, which will supply the embryo with nutrients along its development (Berger, 2003; Dresselhaus and Coimbra, 2016). As for synergids, after fertilisation, the penetrated one degenerates even before PT arrival and the other synergid goes into a programmed cell death process (Higashiyama *et al.*, 2001; Dresselhaus and Franklin-Tong, 2013).

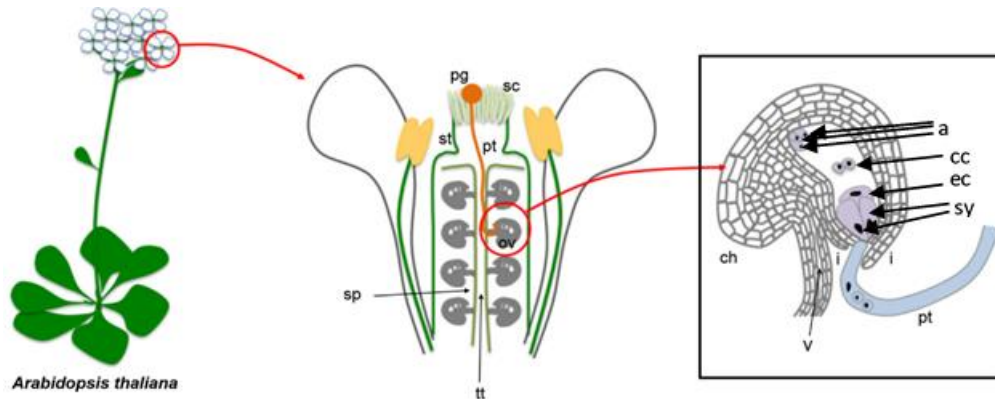


Figure 1. Schematic representation of the PT growth through the pistil's tissues in *A. thaliana*. After the pollen grain lands on the stigma, it germinates and forms the PT, which grows along the style and the transmitting tract until reaching near an ovule. There, the PT curves and grows upon the funiculus towards the micropyle to penetrate one of the two synergids and release the sperm cells. a, antipodals; cc, central cell; ch, chalazal pole; ec, egg cell; i, integuments; ov, ovule; pg, pollen grain; pt, pollen tube; sc, stigmatic cell; sp, septum; st, style; sy, synergids; tt, transmitting tissue; v, vasculature. (Adapted from Pereira *et al.*, 2014).

1.3. *Arabidopsis thaliana* seed development

Arabidopsis seed starts to be formed after the double fertilisation and it is composed by the embryo and the ovule sporophytic maternal tissues. During seed development, two main phases can be distinguished: a first one during which the embryo establishes structures that will generate the future plant, followed by a period when the embryo cells differentiate and expand, at the same time the embryo produces storage products.

In what concerns the gametophytic tissues, the endosperm firstly goes through a coenocytic phase followed by a cellularisation stage, being finally reabsorbed during seed maturation. This tissue can be divided in micropylar endosperm (Figure 2, MCE), located near the embryo, peripheral endosperm (Figure 2, PEN) and chalazal endosperm (Figure 2, CZE), at the chalazal pole (Becker *et al.*, 2014; Berger, 2003; Mansfield and Briarty, 1991).

Embryo development starts with the polarisation of the zygote, creating an apical region with the nucleus and most of the cytoplasm, and a basal region highly vacuolised. This non-uniform distribution of the embryo initial cell cytoplasm leads to the first asymmetric cell division of the embryo. From this asymmetric cell division, two different cells arise, the apical cell, which will give rise to the embryo proper (Figure 2, EP) and a basal cell

that will originate the suspensor and the hypophysis (Figure 2, SUS). During embryo development, the first cell divisions produces a globular shape embryo. Furthermore, at this stage the embryonic tissues also become determined and start to differentiate.

Subsequently, protoderm cell divisions increase in the area where the cotyledons will arise leading to the formation of the cotyledon initials, giving the embryo a heart shape. Procambium starts developing at this stage.

At the last stage of development, the embryo elongates and expands, the primary vascular tissues are established, being surrounded by parenchymatous cell layers, and the embryo becomes greener, due to the development of chloroplasts in the photosynthetic tissues, mainly in the cotyledons. As a result of the embryo elongation, it gets a torpedo shape. As the cotyledons develop, they curve and grow towards the chalazal pole. (Mansfield and Briarty, 1991).

The ovule sporophytic tissues, inner and outer integuments, lead to the formation of the seed integument, which in turn is divided in two regions, the chalazal integument (Figure 2, CZSC) and the distal integument (Figure 2, SC) (Becker *et al.*, 2014).

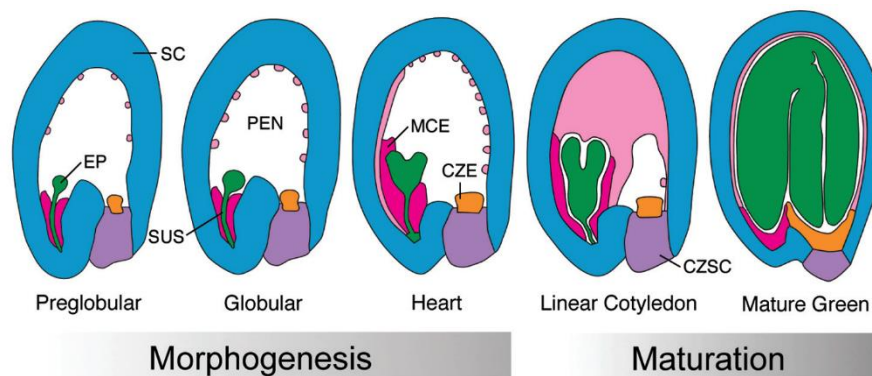


Figure 2. Development and biological functions of *Arabidopsis* seeds' sub-regions. After passing through several divisions at the globular stage, the embryo develops the cotyledon initials at the heart stage. Finally, the cotyledons grow and curve in the last phase. The endosperm is divided into three different zones depending where it is formed and the same is applied to both regions of the seed integument. Green, embryo proper (EP) and suspensor (SUS); dark pink, micropylar endosperm (MCE); light pink, peripheral endosperm (PEN); orange, chalazal endosperm (CZE); purple, chalazal seed integument (CZSC); blue, distal seed integument (SC). (Adapted from Becker *et al.*, 2014).

1.4. SEEDSTICK control of ovule development

Relatively to the genes involved in ovule and seed development, SEEDSTICK is maybe one of the most important ones. Its expression pattern has already been reported by Mizzotti *et al.* (2014), using a STK-GFP reporter line: as the ovules start to develop, *STK* is expressed in the placental tissues and throughout the ovule primordia. Later, at the beginning of the integuments' development, its expression appears exclusively in the nucellus and funiculus. Finally, when the seed reaches a mature stage, *STK* is still

detected in the funiculus, but additionally in the outer integument and in the outer layer of the inner integument. Pinyopich *et al.* (2003) also verified these results (Figure 3. **a** – **e**)).

Together *STK*, *SHATTERPROOF* (*SHP*) and *AGAMOUS* (*AG*) constitute a MADSbox genes clade, contributing redundantly to the ovule's identity (Pinyopich *et al.*, 2003). The MADS TFs assemble themselves in dimers which subsequently form tetramers. These tetramers connect to two CArG boxes, DNA-binding motifs present in the target gene promoter responsible for the binding to these TFs, forming a DNA loop (Yan *et al.*, 2016). *STK* is crucial in ovule and seed development, interacting with many other genes to do so. Some of these crosstalks are listed below. Firstly, relatively to ovule identity, despite *shp1 shp2* double mutants and *stk* single mutants present no alterations in the ovules' identity, *stk shp1 shp2* triple mutants' ovules stop their development or are transformed into leaf-like or carpel-like structures (Pinyopich *et al.*, 2003). Furthermore, this complex also interacts with other genes. Analyses of the quadruple mutant *bel1 stk shp1 shp2* phenotype, along with outcomes of yeast-three-hybrid and pull-down assays, made it possible to infer that *STK* is part of an ovule identity complex that stabilises another complex, *BEL1-AG-SEP3*, contributing to the regulation of the outer integument development. This is because this last complex activates *INNER NO OUTER* (*INO*) (Baker *et al.*, 1997), a relevant gene for the outgrowth of the outer integument (Brambilla *et al.*, 2007; Battaglia *et al.*, 2008; Brambilla *et al.*, 2008). When combined with *Arabidopsis* B Sister (*ABS*), *STK* also controls the formation of the innermost layer of the inner integument, the seed endothelium. The *abs* single mutant already shows an abnormal endothelium whose cells are flatter (Nesi *et al.*, 2002). Nonetheless, the *abs stk* double mutant does not even contain an endothelium, leading to an abrupt decrease in seed production (Mizzotti *et al.*, 2012). Moreover, *STK* and *ABS* are also involved in the pro-anthocyanidins (PAs) accumulation in the endothelium. In this accumulation process, the gene *BANYULS* (*BAN*) is responsible for coding the anthocyanidin reductase, a crucial enzyme for the synthesis of PAs (Debeaujon *et al.*, 2003). *BAN* expression is activated by *ABS*, which, in turn, is repressed by *STK*, that also directly represses *BAN*. So, it is possible to conclude that this crosstalk controlled by *STK* is responsible for regulating PAs accumulation in the innermost layer of the seed coat (Mizzotti *et al.*, 2014). With regards to the identity of the female gametophyte cells and in the PT-synergids interaction, also here *STK* has a fundamental role. The heterotetramer formed by two *STK-SEP3* dimers promotes the expression of *VERDANDI* (*VDD*) and *VALKYRIE* (*VAL*), which are *REPRODUCTIVE MERISTEM* (*REM*) TFs (Matias-Hernandez *et al.*, 2010; Mendes *et al.*, 2013; Mendes *et al.*, 2016). While *VDD*

regulates the identity of the antipodal and synergid cells (Matias-Hernandez *et al.*, 2010), VAL together with VDD control the bursting of the PT and synergid cell death (Mendes *et al.*, 2016). STK has an interference even in the mucilage release and seed germination mechanisms, since it directly activates *PMEI6*, which is involved in the biogenesis of the cellulose-pectin matrix of the cell wall, being *stk* mutants defective in this process (Ezquer *et al.*, 2016).

Still, STK is considered a master TF that regulates ovule and seed development due to many other reasons. Favaro *et al.* (2003) demonstrated that STK ectopic expression leads to the formation of carpeloid organs containing ovules instead of sepals in *Arabidopsis* plants (Figure 3. **f**). This was also verified when overexpressing *FLOWER BINDING PROTEIN 11 (FBP11)* in *Petunia hybrida* (Colombo *et al.*, 1995). *FBP11* and *FBP7* are STK orthologues genes (Rounsley *et al.*, 1995) and were the first ovule identity genes discovered (Colombo *et al.*, 1995), which is why STK was proposed to regulate ovule-specific gene expression. Considering all the previous data presented in this section, it can be concluded that STK is necessary to stimulate normal ovule development. Moreover, this TF also interferes in seed formation and dehiscence. In wild-type (WT) *Arabidopsis* plants, seed dehiscence happens when the funiculus experiences a constriction in the abscission zone, because the cells present in this zone differentiate and become smaller (Figure 3. **h**, **i**). However, in *stk* mutants, this zone contains more cells and these cells are bigger (Figure 3. **m** – **p**). As a result, there is an abnormal growth of the funiculus, which prevents seed dehiscence (Figure 3. **j**, **k**), as well as decreases the available space for each seed to grow (Figure 3. **l**). STK is thus, responsible for the funiculus growth regulation, controlling cell division and expansion in funiculus abscission zone (Pinyopich *et al.*, 2003). Through the analysis the *stk shp1 shp2 ant* quadruple mutant, whose funiculus cells are much larger than the observed in WT plants, Losa *et al.* (2010) concluded that the *AINTEGUMENTA (ANT)* gene, together with the ovule identity genes of this complex, also regulates the size of the funiculus cells. Furthermore, *stk* mutants' siliques are smaller than WT *Arabidopsis* siliques (Figure 3. **g**). As for the seed size and shape, *stk* mutants' seeds are smaller and rounder than the WT ones (Figure 3. **q**) (Pinyopich *et al.*, 2003). All the STK functions mentioned are present in Table 1.

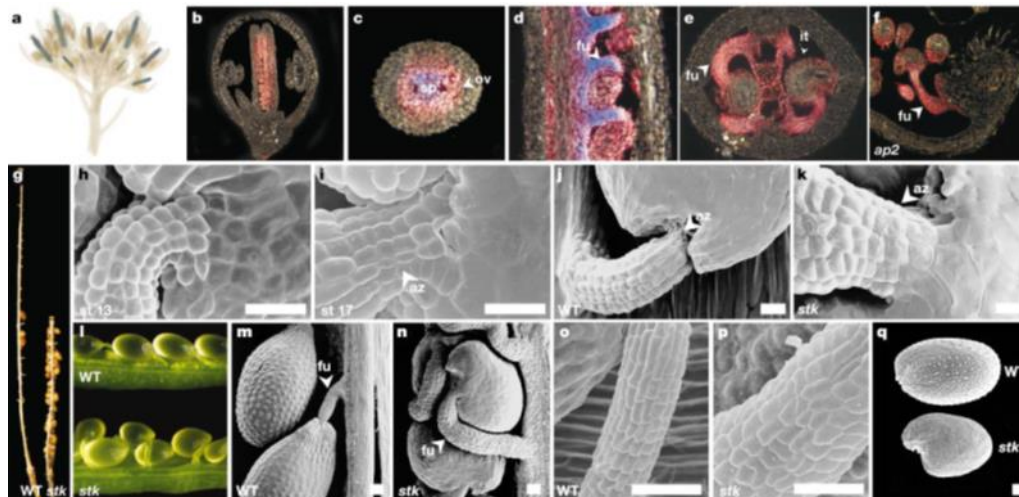


Figure 3. Arabidopsis *STK* characterisation. **a**, histochemical detection of β -glucuronidase activity (GUS) reveals a strong *STK* expression in the pistils; **b**, GUS activity is initially detected in the placenta and in the ovule primordia (stage 8); **c**, transversal section of a pistil, showing uniform GUS expression in the ovule primordia and in the septum; **d**, during stages 11 and 12, GUS expression is stronger in the funiculus (blue) and weaker in the rest of the ovule primordia (pink); **e**, at stage 13, the GUS activity is restricted to the funiculus and to the integument; **f**, GUS activity is present in ectopic ovules originated in organs of the first whorl of *ap2* mutants; **g**, WT plants siliques (WT, left) are bigger than *stk* mutants siliques (*stk*, right) at stages 19 and 20. The majority of the seeds remain in the silique in *stk* mutants; **h** and **i**, in WT Arabidopsis plants, the differentiation of the abscission zone is observed at stage 17 (**i**), but not at stage 13 (**h**); **j** and **k**, seed abscission occurs in WT plants (**j**), but not in *stk* mutants (**k**); **l**, irregular seed disposition in a *stk* mutant silique (*stk*) relatively to seed disposition in WT plant silique (WT); **m** and **n**, at stage 17, WT plants funiculus (**m**) is smaller than *stk* mutants funiculus (**n**); **o** and **p**, funiculus cells in WT plants (**o**) are smaller than in *stk* mutants (**p**); **q**, *stk* plants mature seeds (*stk*) are smaller than WT plants seeds (WT). az, seed abscission zone; fu, funiculus; it, integument; sp, septum; ov, ovule. Scale bars: 20 μ m (**h** – **k**), 50 μ m (**m** – **q**). (Flower developmental stages according to Smith *et al.*, 1990; Adapted from Pinyopich *et al.*, 2003).

Table 1. STK proposed interactions and functions throughout ovule and seed development. Question marks indicate a still unknown target gene or member of the complex where STK is involved. *, *BAN* is a target gene of the *STK-SHP-SEP* complex and of *ABS*; +, *INO* is a target gene of the *BEL1-AG-SEP* complex, which is a target of the *STK-SHP-SEP* complex.

Complex	Target genes		Mechanism influenced	Article
STK-SHP-AG-SEP	?		Ovule identity	Favaro <i>et al.</i> , 2003
STK-SHP-SEP-?	<i>ABS</i>		Endothelium formation	Mizzotti <i>et al.</i> , 2012
STK-SHP-SEP-?	<i>ABS</i>	<i>BAN</i> *	PAs accumulation in the endothelium	Mizzotti <i>et al.</i> , 2014
STK-SHP-SEP	<i>BEL1-AG-SEP</i>	<i>INO</i> +	Outer integument development	Brambilla <i>et al.</i> , 2007 Brambilla <i>et al.</i> , 2008 Battaglia <i>et al.</i> , 2008
STK-SHP-ANT-SEP	?		Funiculus development	Pinyopich <i>et al.</i> , 2003 Losa <i>et al.</i> , 2010
STK-SEP3-STK-SEP3	<i>VDD</i>	Synergid and antipodals cells identity		Matias-Hernandez <i>et al.</i> , 2010
	<i>VAL</i>	Synergid programmed cell death	PT burst	Mendes <i>et al.</i> , 2013 Mendes <i>et al.</i> , 2016
STK-?-?-?	<i>PMEI6</i>		Mucilage release and seed germination	Ezquer <i>et al.</i> , 2016
STK-?-?-?	?		Seed shape and size Silique size	Pinyopich <i>et al.</i> , 2003

1.5. STK targets: crosstalks behind Arabidopsis reproduction

1.5.1. STK and cytokinin oxidase/dehydrogenase enzymes

Recent ChIP-sequencing analyses identified certain arabinogalactan proteins (AGPs) and the enzymes cytokinin oxidase/dehydrogenase 6 and 7 (CKX6 and CKX7) as STK targets, indicating a possible relationship between these players (unpublished data). What is interesting to notice is that, such as STK, some AGPs (Lalanne *et al.*, 2004; Tucker *et al.*, 2012; Pereira *et al.*, 2014; Tucker and Koltunow, 2014; Pereira *et al.*, 2016a; Pereira *et al.*, 2016b), as well as CKX6 and CKX7 (Werner *et al.*, 2003; Köllmer *et al.*, 2014) also have a role in reproduction and ovule development.

Phytohormones are important molecules, interfering in a variety of processes, many of them related to plant development. Within phytohormones, cytokinins (CKs) are involved in many significant mechanisms namely cell division and differentiation, seed, fruit and root development, stem growth, plant nutrition, senescence and response to biotic and abiotic signals. Many processes where cytokinin is involved depend on its concentration level, which, in turn, is regulated by mechanisms of cytokinin biosynthesis, activation, reactivation, reversible inactivation, irreversible inactivation and degradation (Jameson and Song, 2016). Crucial to CK degradation, cytokinin oxidase/dehydrogenase (CKX) catalyses the reaction that transforms the isopentenyladenine molecule into 3-methyl-2-butenal and adenine. This reaction is essential to regulate CK levels and the distribution of this hormone in the plant (Galuszka *et al.*, 2000). *CKX* gene family is composed by seven genes in *A. thaliana*, some of them already described by Werner *et al.* (2003) and Bartrina *et al.* (2011), showing their interference in development processes. Through the analysis *Arabidopsis* plants overexpressing *CKX1* and *CKX3*, Werner *et al.* (2003) observed a decrease in seed production in these individuals relatively to WT *Arabidopsis* plants. Moreover, studying the same species, Bartrina *et al.* (2011) reported an increase in flower and meristem size as well as in the production of meristematic cells, ovules, seeds, flowers and siliques in the *ckx3 ckx5* double knockout mutant plants relatively to the WT ones. With these observations and taking into account that CKX is essential for CK degradation and that CKs delay the exit of cells from a meristematic phase, it can be assumed that a loss of function of the CKX3 and CKX5 enzymes leads to an increase in the CK level which, in turn, increases the number of placenta cells and, consequently, the number of ovules and seeds formed, as well as the flower size and the production of flowers and siliques. In the opposite sense, if *CKX1* and *CKX3* are overexpressed, the CK level decreases resulting in a decrease in the number of placenta cells arrested at

the meristematic stage and less seeds are formed, which explains the results obtained by Werner *et al.* (2003).

Nevertheless, there are more indications of the influence of CKX enzymes in ovule and seed development. *CKX6* expression was detected in the vasculature of the cotyledons, leaves, roots and apical parts of the stem, but also in the flower vasculature, in the transmitting tissue and in the funiculus (Figure 4). This might indicate that CK regulation by *CKX6* interferes on nutrient translocation to the embryo, since this translocation goes through the funiculus (Werner *et al.*, 2003). Köllmer *et al.* (2014) studied the *CKX7* enzyme and observed its activity in the vasculature of the root, hypocotyl, cotyledons and young leaves, in the transmitting tissue of the pistil, in synergids and egg cell, where its activity decreases after pollination (Figure 5). The detection of *CKX7* activity in specific locations of the embryo sac might suggest an involvement of *CKX7* in the female gametophyte development. Recently and in what concerns *CKX7* function, through a ChIP assay, observing siliques' size in *stk* and *ckx7* mutant lines and analysing CK signalling and *CKX7* expression in *stk* mutants (Figure 6), Di Marzo *et al.* (2020) concluded that CKs action opposes *Arabidopsis* fruit growth and STK controls the crosstalk that leads to fruit elongation, directly activating *CKX7* expression. At this point, nothing else is known about these two enzymes regarding ovule and seed development. Therefore, some questions must be answered in order to fully understand STK and CKX role in these processes: how does *CKX6* interact with STK? Are there other players that “crosstalk” with STK, *CKX6* and *CKX7*?

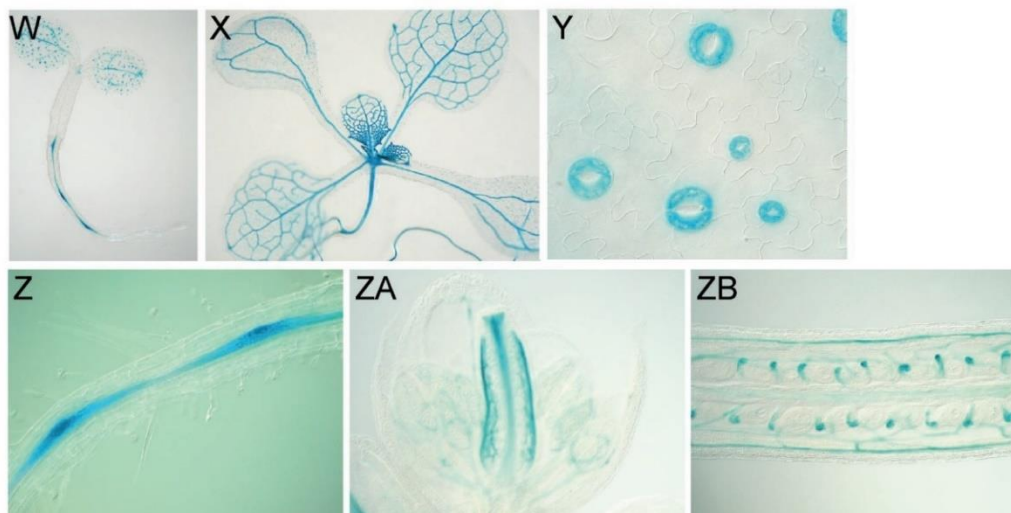


Figure 4. Localisation of *AtCKX6:GUS* activity. Expression at the aerial part is confined mainly to young tissues' vascular system (for example, cotyledons (**W**) and growing leaves (**X**) and to young leaves' developing stomata guard cells (**Y**). In post-embryonic roots this expression is detected for the first time two days after germination in the vasculature, with maximum expression around lateral roots primordia (**W** and **Z**). In flowers, the expression is detected in the gynoecium (**ZA**) and in the funiculus (**ZB**). (Adapted from Werner *et al.*, 2003).

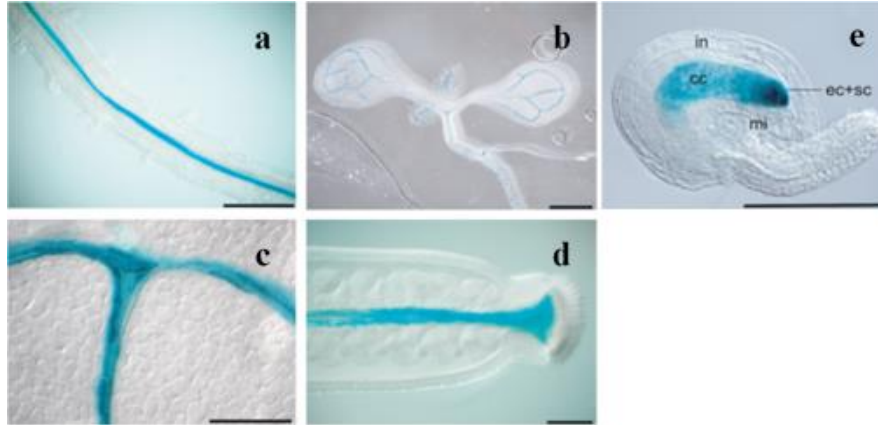


Figure 5. Gus activity in *CKX7:GUS* *Arabidopsis* transgenic plants. In plantlets, *CKX7:GUS* is active in the primary root vasculature (a) and in the aerial organs (b); This expression is also found in vascular cells in the distal tip of the third leaf of *Arabidopsis* rosette nine days after germination (c); During the late developmental stages, GUS activity is also detected in the transmitting tract (d) and in the mature embryo sac, being higher in the synergids and egg cell (e). Scale bars: 100 μ m (a, c, d and e) and 1 mm (b). cc, central cell; ec, egg cell; in, integument; mi, micropyle; sc, synergids. (Adapted from Kölmer *et al.*, 2014).

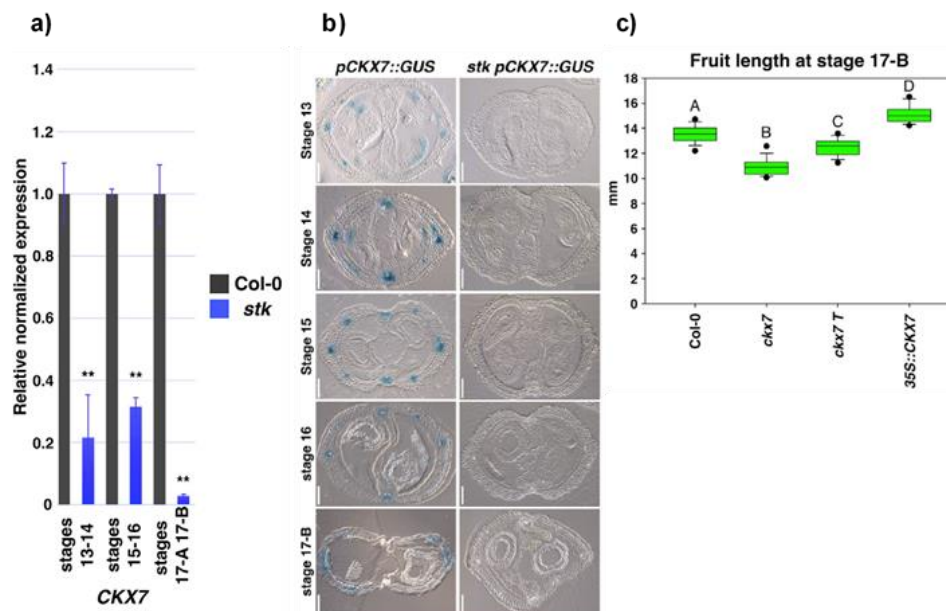


Figure 6. Evidences of STK directly regulation of *CKX7* expression and the influence of this gene in silique elongation. **a)** expression levels of *CKX7* in *stk* pistil and fruit are much lower than in WT at stages 13 – 14, 15 – 16, 17-A and 17-B (statistical analysis was performed using Student's t test (** $p < 0.01$) and gene expression was normalised to that of *ACTIN* and *UBIQUITIN*. Histograms show a representative experiment in which error bars represent the standard deviation for three technical replicates. Three qPCR (biological replicates) were performed and generated similar results); **b)** there is no expression of *pCKX7::GUS* in cross sections of *stk* pistil and fruit from stage 13 to stage 17-B, contrary to what happens in WT pistil and fruit at the same stages (scale bar: 100 μ m); **c)** fruit length in *ckx7* mutants (*ckx7* – T-DNA insertion mutant; *ckx7 T* – mutant generated using CRISPR-Cas9 technology) is lower than fruit length in WT plants, which in turn, is lower than plants overexpressing *CKX7* (35S::CKX7) ($n = 20$ siliques; 3 biological replicates). Staging of pistils and fruits is according to Ferrándiz *et al.* (1999) and Roeder and Yanofsky (2006). Adapted from Di Marzo *et al.* (2020).

1.5.2. STK and arabinogalactan proteins

Arabinogalactan proteins are a big family of highly glycosylated hydroxyproline-rich proteins which are divided in four groups: the classical AGPs, the lysine-rich AGPs (Lys-rich AGPs), the arabinogalactan peptides (AG peptides) and the fasciclin-like AGPs (FLAs). The first group is composed by a core domain rich in proline/hydroxyproline and a N-terminal signal peptide which targets the core polypeptide for secretion. The other groups have a structure somewhat similar to the classical AGPs, with some differences; AG peptides contain only 10 to 13 amino-acids, Lys-rich AGPs are rich in lysine and FLAs contain fasciclin-like domains in the polypeptide core. Approximately half of the AGP family members are predicted to be tethered to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor signal in the C-terminal, that is detached from the rest of the protein in order to form the mature protein. Relatively to their location in the cell, AGPs are attached to the outer layer of the plasma membrane oriented extracellularly (Pereira *et al.*, 2015).

There is not a definitive theory about the biological function of AGPs, but there are three models that try to uncover their mode of action: the first one suggests that AGPs' sugars behave as signalling molecules, morphogenes or nutrients (Showalter, 2001; van Hengel *et al.*, 2001); the second model proposes that the entire AGP may be dissociated from the GPI anchor, functioning in signal transduction pathways itself (Schultz *et al.*, 1998); and a last and more recent study presents a model in which AGPs operate as calcium chelators in a pH-dependent manner, through the glucuronidation of arabinogalactan (AG) polysaccharides, enabling the cell to regulate the discharge of calcium (Lamport and Várnai, 2013). Related to this last theory, in *Torenia fournieri*, a bioactive AGP side-chain fragment called AMOR, responsible for the response of the PT to guidance peptides emitted by the synergids (Okuda *et al.*, 2009), was identified as a methyl-Oglucuronosyl-b-D-galactose, suggesting that AMOR may well be an essential component of the AGP Ca²⁺-binding motif (Mizukami *et al.*, 2016). Moreover, it was reported that AG sugar chains specific structures like AMOR have a differential localisation, implying that enzymes responsible for the biosynthesis of AG can locally regulate AGP sugar chains and, most probably, AGP functions (Su and Higashiyama, 2018). Furthermore, Lopez-Hernandez *et al.* (2020) concluded that glucuronidation of AG is essential for several plant developmental mechanisms. In this study, the authors used *Arabidopsis* mutant plants for enzymes responsible for glucuronidation of AG and observed that these mutants not only presented various developmental defects, but also contained AG with a lower calcium binding capacity *in vitro* than AG from WT plants, supporting that AGPs may bind and release cell-surface apoplastic calcium. However,

due to the intricate nature of AGPs and since these models do not conflict with each other, AGPs may well function in different ways following more than one model.

AGPs are frequently associated to several processes of sexual plant reproduction like signalling molecules, cell identity determinants, morphogens, nutrient sources and support for PT growth, among others. These molecules have even been detected in the PT pathway into the embryo sac in a temporal and spatial differential pattern. Figure 7 summarises the AGPs expression patterns described to date, according to Pereira *et al.* (2016a) review article. The information known at this moment relative to the AGPs used in the present research is mentioned below.

Up until recently, AGPs were difficult to study individually, since the Yariv reagent (Yariv *et al.*, 1967) and the antibodies used for their detection, only identify the carbohydrates epitopes, making it impossible to distinguish between different AGPs. The advances in genomics allowed to study more deeply this group of proteins, reaching aspects as the functions and mode of action of the different AGPs (Pereira *et al.*, 2015). AGP4 is one of these cases. During the last stage of the PT growth into the pistil, a successful double fertilisation happens when one PT penetrates one of the synergids and the other synergid degenerates, so that no other PT penetrates it, blocking polyspermy. In the *A. thaliana agp4* knockout mutant, using fluorescence microscopy either with reporter lines or aniline blue staining, Pereira *et al.* (2016b) observed that this mutant did not present polyspermy block, showing more than one PT into one embryo sac, since the persistent synergid cell did not degenerate. Thus, this group concluded that AGP4 – JAGGER – intervenes in the mechanism that regulates the inhibition of polyspermy. Moreover, through *in situ* hybridisation and histochemical detection of the β -glucuronidase activity with *GUS* reporter gene linked to the *JAGGER* promoter, Pereira *et al.* (2016b) detected *JAGGER* expression in the stigma, style, transmitting tract, ovule integuments and inside the embryo sac (Figure 7).

Nevertheless, for most of the AGPs only its gene differential expression pattern is known. Some of these AGPs expressed in *Arabidopsis* female tissues are AGP1, AGP9, AGP12 and AGP15. So far, the findings made concerning these *Arabidopsis* AGPs is their gene expression pattern. Pereira *et al.* (2014) analysed *AGP1*, *AGP9*, *AGP12* and *AGP15* expression by real-time RT-PCR, observing that these genes are upregulated in the pistil tissues relatively to pollen transcript levels. This was also shown by using reporter lines, where the GFP expression was driven by the *AGP1*, *AGP9*, *AGP12* and *AGP15* promoters. With this last technique, Pereira *et al.* (2014) could conclude that *AGP1* gene expression is strongly present in the style tissues, septum, transmitting tract, funiculus and chalazal region of the ovule, being weaker in the stigmatic cells and in the

integuments of the ovule. *AGP1* expression pattern was also analysed through the detection of β -glucuronidase activity, being the *GUS* reporter gene expression also guided by the *AGP1* promoter. Plants expressing this construct presented a low *GUS* activity in the stigmatic cells and a high *GUS* activity in the septum, transmitting tract, funiculus, chalaza and ovule integument cells. As for the *AGP9* gene expression, it is present in the vascular tissues of the pistil transmitting tract, the septum and the funiculus, presenting a weak expression in the chalazal pole of the ovules. *AGP12* expression is located in the stigmatic cells and the chalazal pole of the ovules with a stronger intensity and along the internal tissues of the funiculus and septum with less strength. The authors also confirmed *AGP1* and *AGP12* expression patterns stated above using FISH. Finally, *AGP15* expression was detected in all female reproductive tissues, except in the transmitting tract. This pattern was also verified when the β -glucuronidase activity was observed in plants transformed with a construct containing the *AGP15* promoter fused to the *GUS* reporter gene (Figure 7). According to the authors, these expression patterns indicate that *AGP9* might be involved in the nutrition or signalling between the pistil vasculature and the ovule, especially the endosperm, since the principal nutrient absorption into the endosperm coming from the maternal parent happens through the chalazal pole. This can also be assumed for *AGP1* and *AGP12*, considering their expression at the chalazal tissues of the ovules and at the funiculus. As for the expression of *AGP1*, *AGP12* and *AGP15* present in the female tissues through which the PT grows, it indicates a possible redundant role of these AGPs in PT growth and in fertilisation. *AGP1* and *AGP12* can even be related to pollen-stigmatic cells interactions, due to its expression in the stigmatic cells (Pereira *et al.*, 2014).

Another AGP whose gene differential expression pattern was described was *AGP24*. Tucker *et al.* (2012) and Tucker and Koltunow (2014) used a reporter line in which the *YFP* was regulated by the *AGP24* promoter and found that, during early ovule development, *AGP24* is expressed specifically in the functional megaspore, which indicates that this AGP might be involved with the fate of the functional megaspore. Apart from this, Lalanne *et al.* (2004) also observed *AGP24* expression in roots, stems, leaves, flowers and more intensely in pollen through a RT-PCR analyses. This group also reported a similar expression pattern for *AGP21*, although this gene expression was weaker.

Other AGP genes, like *AGP23* (Pereira *et al.*, 2014), *AGP6*, *AGP11* and *AGP40*, have so far been described as being expressed only in pollen (Pereira *et al.*, 2016a). Pereira *et al.* (2014) also analysed *Arabidopsis* plants transformed with the constructs

pAGP23:GUS and pAGP23:NLS:3GFP and observed reporter genes' expression in the pollen grains and PT, indicating that this AGP is specific to the pollen vegetative cell (Pereira *et al.*, 2014). As for AGP6 and AGP11, through a functional analysis and performing a yeast two-hybrid experiment, Costa *et al.* (2013) proposed that these AGPs are included in the PT endosome machinery (Figure 7). With regards to AGP31, only aspects related to its chemical structure (Hijazi *et al.*, 2012) were found. Despite the difficulties in studying them, there is much yet to discover, AGPs must definitely have an important role in reproduction. This family of proteins may be another member of the "crosstalk" between STK, CKX6 and CKX7.

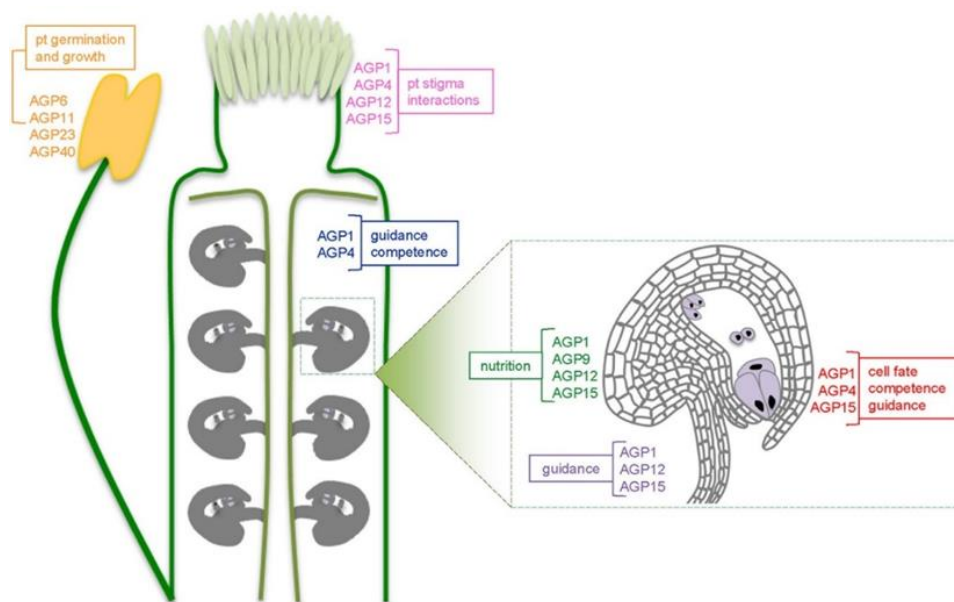


Figure 7. Schematic representation of an *Arabidopsis* stamen and ovule, showing the different AGPs present along the PT path, in the pollen grain and in the endosperm. Orange: pollen-specific AGPs (AGP6, AGP11, AGP23, AGP40); Pink: AGPs present in the stigma (AGP1, AGP4, AGP12, AGP15); Blue: AGPs that are known of being expressed in the transmitting tissue (AGP1, AGP4). In the ovule scheme: Green: AGPs present at the chalazal pole (AGP1, AGP9, AGP12, AGP15); Red: AGPs present along the integuments near the micropyle (AGP1, AGP4, AGP15); Violet: AGPs present in the funiculus (AGP1, AGP12, AGP15). It is also shown the possible functions for each group of AGPs. (Adapted from Pereira *et al.*, 2016a).

There are other molecules that intervene in the gene network responsible for the ovule development and the fertilisation processes. Three of these other molecules are LORELEI (LRE), the LUREs group and MYB98. LORELEI is a GPI-anchor membrane protein expressed in the synergid cell and is one of the molecules responsible for PT growth arrest at the synergid cell (Palanivelu and Tsukamoto, 2012). Relatively to the LUREs, these cysteine-rich peptides that belong to the defensin-like protein subgroup are attractant molecules involved in the PT micropylar guidance. Despite being firstly identified in *Torenia fournieri*, LUREs are also present in *A. thaliana* (AtLUREs), being

produced in the synergids and discharged by the filiform apparatus (Takeuchi and Higashiyama, 2012). There are several LUREs' receptors that were already described in *Arabidopsis* and there are also some LUREs' regulators, such as MYB98. MYB98 expresses a TF whose action leads to the normal formation of the filiform apparatus and acts indirectly in the PT growth micropylar guidance, because it hampers the filiform apparatus formation (Palanivelu and Tsukamoto, 2012).

To sum up, several molecules intervene in the reproductive process in *Arabidopsis*, acting in different or similar ways and regulating all the steps, since pollen and ovule synthesis until seed formation. However, most of the networks in which these molecules interact are not yet fully understood. STK must control at least most of the ovule and seed development mechanisms, but who are its targets and how they interact is something which has not been discovered totally.

1.6. Gene expression networks in plant reproduction

1.6.1. Study TFs direct regulation of target genes expression *in vitro*

When investigating gene regulatory mechanisms, protein-DNA binding is an essential subject that needs to be addressed, since it enables to conclude if a certain TF binds directly to a given gene, more specifically, to a particular promoter, informing about the interference of that TF in the expression of the gene. In the seventies, a method widely used to analyse protein-DNA interactions was the filter binding assay (Riggs *et al.*, 1968). This procedure is based on the capability of proteins to attach DNA to a nitrocellulose filter. Although this technique is simple, sensitive, inexpensive and rapid enough to enable kinetic studies and equilibrium measurements, it does not allow the analysis of the composition of the binding reaction products, nor determining whether a single DNA connects to more than one protein (Fried and Crothers, 1981).

Nevertheless, in the following decade, a new method was developed to analyse protein-DNA interactions without this disadvantage. Electrophoretic Mobility Shift Assay (EMSA) (Fried and Crothers, 1981; Garner and Revzin, 1981) consists in evaluating if a protein causes a retardation in the electrophoretic run of a nucleic acid fragment when bound to it in a complex compared to the run of the same nucleic acid not bound to a protein. This technique not only tests the nucleic acid-protein interaction qualitatively, but also quantitatively, since the mobility of the nucleic acid fragment in the gel decreases as the number of proteins bound to it increases, enabling to detect protein-DNA complexes formed by more than one protein (Figure 8). This way, this method enables to assess the stoichiometric ratio of protein linked to the nucleic acid and the relative binding affinities of a certain protein for two different nucleic acids (Fried and Crothers, 1981). Furthermore, if, in the beginning of the electrophoresis, the nucleic acid fragments that are

not involved in any interaction enter the electrophoretic gel before the dissociation of the complexes formed, it is also possible to quantify the concentration of unbound nucleic acid fragments and, consequently, the concentration of the complexes in the binding reaction (Garner and Revzin, 1981).

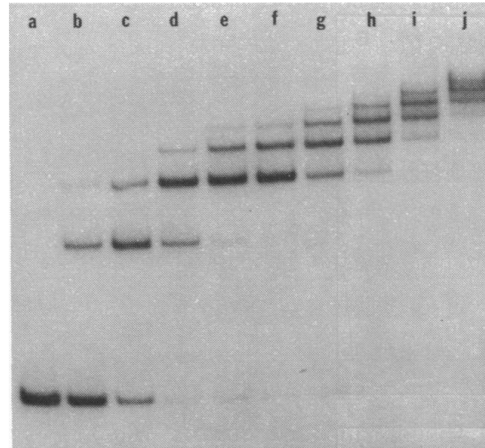


Figure 8. Autoradiography of a polyacrylamide gel that resulted from an EMSA. Crescent amounts of *lac* repressor (a-j) were combined with the *lac* operator fragment, revealing that EMSA also tests the nucleic acid-protein interaction quantitatively, since the mobility of the *lac* operator fragment in the gel decreases as the amount of *lac* repressor bound to it increases (adapted from Fried and Crothers, 1981).

With regards to the EMSA utility, more than being a semi-quantitative technique to follow nucleic acid-binding activity of purified proteins, EMSA can also be used in cell crude extracts, either to verify if one or more proteins bind to a specific nucleic acid, or to examine the distribution of the protein-nucleic acid complex formation across diverse cell-types, tissues, growth states, or even subcellular compartments (Guille and Kneale, 1997). Additionally, there are many methods complement or deriving from EMSA that provide much more information about protein-nucleic acid interactions. Some one-dimensional EMSA variants can measure association and dissociation kinetics, determine binding stoichiometry, evaluate and compare binding affinities involving several proteins or nucleic acids with different protein binding sites, among others. Two-dimensional procedures, in turn, may help in characterising DNA-binding proteins (Hellman and Fried, 2007).

Overall, EMSA is easy to execute, but powerful and sensitive, comprising a broad spectrum of binding conditions and enabling the use of small concentrations of protein and nucleic acid and small sample volumes, especially when the nucleic acid is radio-labelled. This assay is able to test oligonucleotides with a large variety of sizes and structures. Proteins used in EMSA can have very distinct sizes and still efficiently provide mobility shifts. Using EMSA, one can even differentiate how proteins are spread across different interactions with various nucleic acid molecules in one solution, as well as the

existence of complexes with different protein stoichiometry and/or binding site distribution (Hellman and Fried, 2007).

Nevertheless, EMSA also presents some disadvantages. One of them is that, during the electrophoresis run, partial or total dissociation of the nucleic acid-protein complex can happen, or, contrastingly, many complexes get much more stable in the gel than they are in a free solution. Either way, the results can be misleading, since they do not reflect the state of the interaction before the loading of the sample in the electrophoresis gel. Another drawback of this assay is that many factors may be in the origin of a mobility shift, such as the conformation and structure of the complex, and not necessarily the size and number of the proteins present in the complex. Moreover, this assay does not give information about the location of the sequences of the nucleic acid that are linked to the protein. Additionally, the time needed to handle the solutions inhibits the performance of kinetics analyses that require longer periods. Finally, the electrophoretic step of the EMSA technique limits the salt concentrations of the samples to nearly 300 mM or less and the DNAs' size to approximately 5.000 bp or less, depending on the sample and gel compositions, unlike the nitrocellulose filter binding assay (Hellman and Fried, 2007).

Overall, EMSA is an important technique that can help in identifying nucleic acid-protein interactions quite easily and rapidly. This approach proved to be useful when addressing the gene regulatory networks related to sexual plant reproduction (Pajoro *et al.*, 2014). Since the '90s, EMSA is indeed being used to verify the binding between TFs and the respective target genes (Tröbner *et al.*, 1992). For instance, Lucia Colombo's research group has employed this technique for some years now to study regulatory networks, including the ones associated with MADS-box genes and plant reproduction (Kooiker *et al.*, 2005).

1.6.2. Study TFs direct regulation of target genes expression *in planta*

Apart from *in vitro* techniques, approximately three decades ago a different approach was developed to detect protein-DNA interactions: the yeast one-hybrid (Y1H). This method implicates a construct containing the gene of interest, or a region of its promoter, more specifically, its *cis* elements, upstream of a reporter gene and a vector expressing the TF under study fused to a yeast transcription activation domain, usually GAL4. These two elements are inserted into a yeast and, in case the TF does bind to the gene of interest, the activation domain leads to the expression of the reporter gene, being this activation independent of the action of the TF, either it is an activator or a repressor. Unlike *in vitro* techniques, this technique employs a system where a cell is involved, providing much more reliable results since it recreates an environment similar to the one

where the complex under study is formed (Reece-Hoyes and Walhout, 2012). Still, a yeast cell is quite different from a plant cell. Thus, this method does not totally recreate the environment of a plant cell, disregarding some factors that might interfere and determine the protein-DNA interaction being studied. Therefore, an *in planta* technique needed to be developed.

The extraction of protoplasts was first described sixty years ago (Cocking, 1960) and has been used to study many different processes related to plant physiology, immunity, growth and development. This has been done essentially through the identification of important players and mechanisms behind intracellular signal transduction pathways. Specific responses of various types of cells can also be studied through the isolation of protoplasts from those cells. Furthermore, protoplasts maintain many of the same physiological and cellular mechanisms as a whole plant. The creation of different DNA transfection techniques allowed the use of protoplasts to investigate the functions of genes in different plant species (Yoo *et al.*, 2007).

The creation of a Transient Expression in Arabidopsis Mesophyll Protoplast (TEAMP) method by Yoo *et al.* (2007), a protoplast transactivation assay (PTA), enabled to perform an experiment similar to Y1H, but using plant cells, more specifically protoplasts, eliminating the risk of drawing false conclusions from techniques that use different organisms, like mentioned above. Ueda *et al.* (2011) used this protocol to study a specific protein-DNA interaction and complemented it with a procedure that allows the identification of the target gene's *cis* elements. In order to verify if the WRKY2 TF binds to the *WUSCHEL RELATED HOMEBOX8* (*WOX8*) gene, this research group firstly analysed the *WOX8* gene to identify its *cis* elements. They started by producing several constructs, each of which formed by several copies of an isolated fragment of the *WOX8* gene linked to a reporter gene, doing the same with the complete *WOX8* genomic sequence, but using only one copy. Each construct was inserted in a vector for plant transformation. After analysing the phenotypes of the transformed plants, Ueda *et al.* (2011) concluded that the gene of interest contained 4 *cis* elements, because the plants transformed with the constructs containing these fragments presented a similar expression pattern to the one obtained with the complete *WOX8* genomic sequence. After this step, the group followed the TEAMP protocol and transformed Arabidopsis mesophyll protoplasts, previously transformed with another vector containing the *WRKY2* gene, with the constructs containing the *cis* elements. As these protoplasts exhibited reporter gene expression, the group suggested that WRKY2 does bind to the *WOX8* gene, which led to the expression of the reporter gene (Figure 9).

Despite being a very complex procedure that involves numerous steps, PTA is a reliable *in planta* technique when trying to decipher gene expression networks and study gene expression direct regulation.

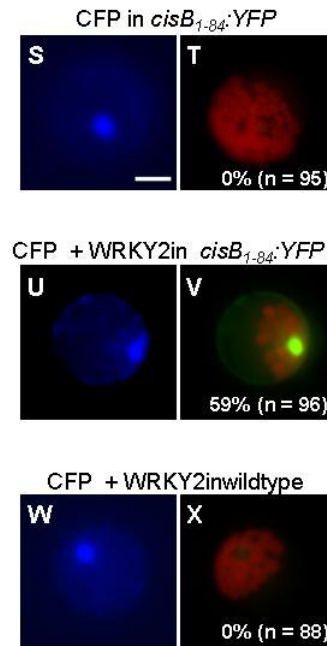


Figure 9. Results from *Arabidopsis* protoplast transactivation. In this work, protoplasts, derived from plants previously transformed with the yellow fluorescent protein (YFP) gene linked to one of the *WOX8 cis* elements (*cisB1-84*) (**S** to **V**) or WT plants (**W** and **X**), were transfected with the cyan fluorescent protein (CFP) gene linked to the nuclear localisation sequence (NLS) (**S** and **T**), or the CFP gene, also linked to the NLS, and the WRKY gene (**U** to **X**). The percentages represent the number of cells displaying YFP and CFP signal in the total number of cells presenting only CFP. Scale bar, 10 μ m (adapted from Ueda *et al.*, 2011).

1.7. Study gene function: genome editing through CRISPR-CAS9

Functional genomic analysis of plants needs the creation of mutant lines. These mutants can be obtained through gene targeting and genome editing. Generally, researchers use T-DNA and transposon insertions to obtain mutants in plants like *Arabidopsis*. The function of a gene can be studied through the analysis of the phenotypes of mutants for that gene. High order mutants are recently being used to address the functions of genes with redundant functions and to investigate epistasis. Nonetheless, using this method for multiple mutagenesis requires a lot of time and labour for performing the crossing of mutant plants and sometimes it cannot even be applied due to the lack of mutant lines available for all the genes of interest. Recently, cheaper, more efficient and easier techniques were created. New genome editing procedures were discovered, making use of sequence-specific nucleases that create double-strand breaks (DSBs) at specific genomic sites (Xing *et al.*, 2014).

Even more recently, a new revolutionary system was created, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) – associated protein 9 (CRISPR-Cas9), which is based in a bacterial immune system that makes use of RNA-guided nucleases (Tsai and Joung, 2016). The CRISPR are repeats present in the bacterial DNA interspaced with DNA specific sequences homologous to strange target DNA (protospacer), forming the CRISPR RNA (crRNA) array. Each target DNA is always accompanied by a protospacer adjacent motif (PAM) sequence that enables the Cas9 enzyme to distinguish the strange DNA from the respective protospacer present in the CRISPR. Once the foreign DNA goes into the bacteria cell, the respective crRNA is transcribed and links to an auxiliary trans-activating crRNA (tracrRNA), forming together the single guide RNA (sgRNA), and then to Cas9, creating a complex that searches the target DNA with the PAM sequence. Then, the crRNA anneals with the strange DNA and Cas9 separates them and creates a DSB in the target gene that leads to a knockout mutation (Ran *et al.*, 2013).

This system can be adapted to create a genome editing technology, simply by changing the target sequence used (Doudna and Charpentier, 2014). The method developed uses a Cas9 enzyme, whose expression is regulated by the Pol II promoter (Lowder *et al.*, 2015), and sgRNAs, whose expression is controlled by Pol III promoters (Tang *et al.*, 2016). These sgRNAs are composed by the crRNA, which, in turn, is formed by the protospacer (usually a specific sequence cassette containing 20 nucleotides homologous to the target DNA) and a partial direct repeat, in the 5' end and by the tracrRNA (or scaffold sequence) in the 3' end. This sgRNA binds to the Cas9 enzyme which cuts the target DNA, forming a DSB. This break can be repaired by non-homologous end-joining (NHEJ), or by homology directed repair (HDR), being the first one more common. Normally, NHEJ leads to insertions or deletions in the target gene, which results in the knockout of that gene, and HDR originates gene editing by introducing point mutations (Komor *et al.*, 2017) (Figure 10).

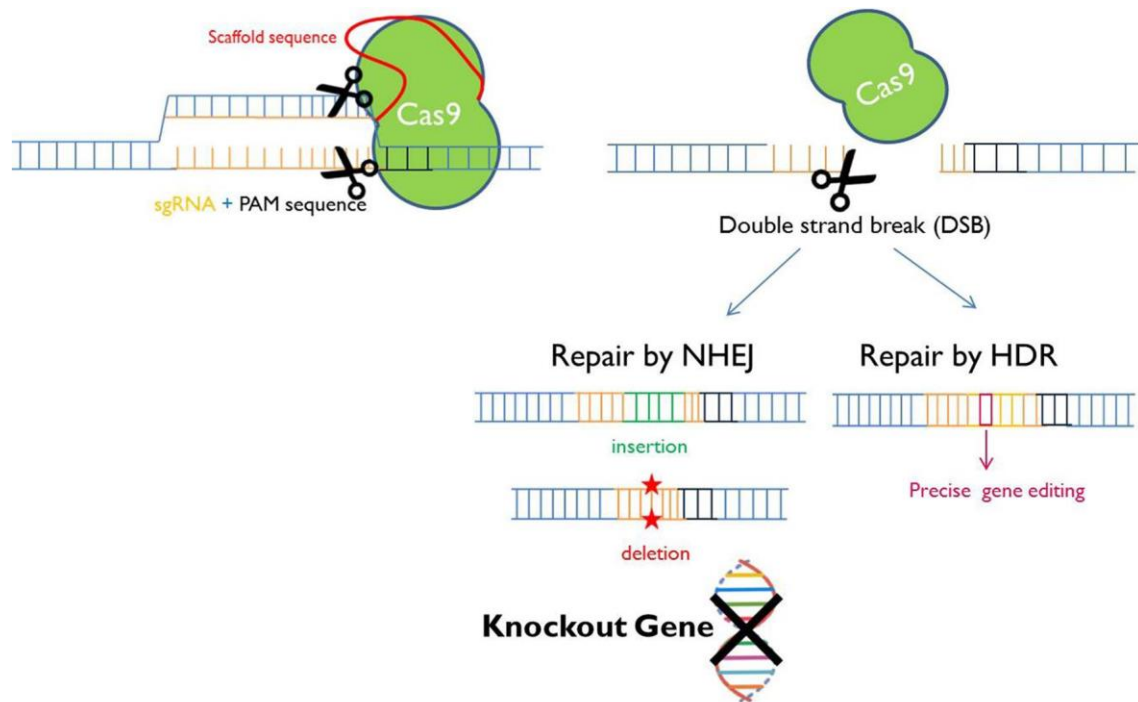


Figure 10. Mechanism of action of the CRISPR-Cas9 technique. The complex formed by the sgRNA specific sequence (orange), the scaffold sequence (red) and the Cas9 enzyme can detect target DNA sequences homologous to the sgRNA specific sequence. If this sequence is upstream of a PAM sequence (black), Cas9 creates a DSB in the target DNA 3 bp upstream of the PAM sequence. This action leads to repair by NHEJ, that creates insertions/deletions in the gene, which causes its knockout, or by HDR, originating gene editing (adapted from Moreira *et al.*, 2020).

2. Objectives

This study was integrated in the SEXSEED project, funded by the EU Horizon 2020 Research and Innovation Staff Exchange programme. It intended to unravel the mechanisms and gene interactions controlling sexual plant reproduction and seed development, more specifically the ones related to STK. STK regulates ovule and seed development, as well as seed dehiscence and its target genes have been revealed by RNA- and ChIP-sequencing assays previously performed. Some of those genes, especially arabinogalactan proteins (AGP) and cytokinin oxidase/dehydrogenase (CKX) genes, chosen for the present study, were also found to be related to plant reproduction processes. However, the direct regulation of these genes by STK has not been confirmed, neither their description nor interactions have been unravelled, in what concerns ovule and seed development.

Several techniques from different areas, such as molecular biology, functional genomics and transcriptomics were used in order to build a more precise and broader model map of the interactions between STK and those targets, as well as to describe these genes' functions during ovule and seed development in *Arabidopsis thaliana*. The expression levels of these genes were accessed and compared between WT and *stk* mutant *Arabidopsis* plants using Real-Time PCR. The expression patterns of these genes were determined and compared between WT and *stk* mutant *Arabidopsis* plants that contained the promoters of these genes linked to reporter genes. An EMSA and a PTA were projected, in order to check direct regulation of these targets' gene expression by STK. The CRISPR-Cas9 technique was also performed aiming to create double mutants for the *STK* gene and each target gene, intending to characterise these mutants and compare their phenotypes with the *stk* mutant. Overall, this project intended to confirm some of *STK* targets identified in the ChIP-sequencing assays and how they interact between each other and with STK.

3. Materials and methods

3.1. Plant material and growth conditions

The WT plants used in all experiments of this work belong to the ecotype Columbia (Columbia 0, [Col-0]) and were obtained from the Nottingham Arabidopsis Stock Center, United Kingdom. As for the *stk-2* (Pinyopich *et al.*, 2003) mutant line, it was gently supplied by Lucia Colombo Lab, Università degli Studi di Milano. The *stk* mutant line and the pAGP21:GUS, pAGP24:GUS, pAGP31:GUS and pAGP9:NLS:3GFP transgenic lines were already available in the laboratory.

Seeds' surface was sterilised in a laminar flow chamber. Firstly, around 10 µL of seeds of each line were submersed in ethanol 70% (v/v) in 1.5 mL eppendorf tubes, which were then shaken. After this step, the seeds were placed in filter paper to dry for half an hour. Then, the seeds were transferred to plates containing solid MS medium (Murashige and Skoog, 1962) supplied with 15 µg/mL of hygromycin when sowing potentially transformed seeds collected from plants used in the CRISPR-Cas9 technique (section 3.7.3.). These plates were firstly placed in the dark for 2 days at 4 °C in order to impel the vernalisation process. Only after this step, the plates were positioned in a growth chamber under 16 hours of light and 8 hours of dark (long-day photoperiod), at a relative humidity of approximately 60% and a temperature around 20 °C. As soon as the seedlings were two weeks old, they were passed to pots containing soil mixed with sand (4:1) and were exposed to the same conditions in the growth chamber.

3.2. Genotyping

The genotypes of WT plants, *stk* mutant plants and *stk* mutant plants transgenic for the pAGP21:GUS, pAGP24:GUS and pAGP31:GUS constructs were confirmed by Polymerase Chain Reaction (PCR) in order to check the presence of the T-DNA insertion in the *STK* gene. The same was performed for WT and *stk* mutant plants transgenic for the pAGP21:GUS, pAGP24:GUS and pAGP31:GUS constructs to verify the presence of the reporter gene in the genomic DNA. Firstly, genomic DNA was extracted from a leaf disk of each plant according to Edwards *et al.* (1991). Then, the DNA samples were used in a PCR in which NZYtaq II 2× Green Master Mix (NZYTech™) was used. The sequences of the specific primers used were designed with Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) and are enumerated in Table 2 as well as the expected band sizes. The PCR mixture composition is present in Table 3 and the PCR conditions are represented in Table 4. A Biometra T3000

Thermocycler (Biometra®) was used in this step. After amplification, in order to observe the PCR results in an electrophoresis, the DNA fragments ran in an 0.8% (w/v) agarose gel in 1× Sodium Borate (SB) buffer (47 g/L boric acid, 200 mM NaOH pH 8.2) with 0.03 µL/mL of GreenSafe Premium (NZYTech™) added before polymerisation. The GeneRuler DNA Ladder Mix (Thermo Scientific™) was used as molecular-weight size marker and 1× SB buffer was used as running buffer. The electrophoretic run was performed at 180 V for 15 minutes. The DNA was observed with a UV transilluminator (302-365 nm) and images were acquired with a KODAK camera and analysed with the KODAK 1D Image Analysis, version 3.5 (KODAK Scientific Imaging Systems).

Table 2. List of the sequences of the primers used in the genotyping PCR.

Name	Primer orientation	Sequence (5'-3')	Expected band size
STK-FW	Forward	TTCTGATAGCACCAACACTAGCA	WT: 364 bp <i>stk</i> : 438 bp
STK-RV	Reverse	GCTTTCTCAAGGCGATTCTC	
GUS-FW	Forward	CCACATGTTACGTCCTGTAGAAACCC	620 bp
GUS-RV	Reverse	CACCACCTGCCAGTCAACAGACG	

Table 3. Genotyping PCR mixture composition.

Reagents	Quantity (µL)
NZYTaq II 2× Green Master Mix	5
Forward Primer (20 µM)	0.5
Reverse Primer (20 µM)	0.5
Genomic DNA sample	1
Nuclease-free water	4

Table 4. Conditions applied in the genotyping PCR with WT and *stk* mutant lines and GUS reporter lines.

Step	Temperature (°C)	Time
Initial denaturation	95	2 minutes
35 cycles	Denaturation	95
	Annealing	30 seconds
	Extension	30 seconds
		30 sec <i>STK</i> 50 sec GUS
Final extension	72	5 minutes
Pause/cooling	8	+∞

3.3. Gene expression analysis through RT-qPCR

3.3.1. RNA extraction

The RNA extraction from *Arabidopsis* WT and *stk* seeds was executed in a fume hood and according to Oñate-Sánchez and Vicente-Carbajosa (2008) with some small changes. Firstly, 10 mg of seeds were collected from siliques at stage 17 of flower development according to Smyth *et al.* (1990), with the help of hypodermic needles (0.4 × 20 mm; B. Braun™), a tweezer (Asa Dental 0645-2) and a Stemi DV4 stereomicroscope (Zeiss™), and were immersed in 550 µL of cool extraction buffer (0.4

M LiCl, 0.2 M Tris pH 8, 25 mM EDTA, 1% (w/v) SDS) in 1.5 mL RNase free microtubes where they were crushed with micro pestles. Then, 550 µL of chloroform was added and the samples were vortexed for 10 seconds and centrifuged during 3 minutes at 4 °C and 16100 g in a Centrifuge 5415 R (Eppendorf™). After this step, the supernatants were transferred to new microtubes where 500 µL of water-saturated acidic phenol was added and, after vortexing, 200 µL of chloroform was added and mixed. Then, a second centrifugation was performed with the same conditions as the previous one and the supernatants were transferred to new microtubes and mixed with 1/3 volume of 8 M LiCl. After precipitating for 1 hour at -20 °C and centrifuging at 4 °C and 16100 g for 30 minutes, the pellets formed were resuspended in 470 µL of nuclease-free water, 7 µL of CH₃COONa (3 M, pH 5.2) and 250 µL of 96% (v/v) ethanol. A centrifugation of 10 minutes at 4 °C and 16100 g followed in order to precipitate the carbohydrates and the supernatants were passed to new microtubes and mixed with 43 µL of CH₃COONa (3 M, pH 5.2) and 750 µL of 96% (v/v) ethanol. The tubes were then placed at -20 °C for 1 hour. Then, a new centrifugation was performed for 20 minutes at -20 °C and 16100 g and the pellets obtained were washed with 250 µL of 70% (v/v) ethanol. Finally, a last centrifugation of 3 minutes at 4 °C and 16100 g created a final RNA pellet which was dissolved in 22 µL of nuclease-free water.

In a second and different approach, this protocol was repeated, but using 30 mg of siliques at stage 17 of flower development according to Smyth *et al.* (1990), excluding the valves, and dissolving the final RNA pellet in 15 µL of nuclease-free water.

3.3.2. RNA quantity and integrity analysis

In what concerns the amount of the RNA samples, they were analysed using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix), according to the manufacturer's instructions, in order to obtain their RNA concentration. Furthermore, to validate the integrity of these samples, 2 µL of each sample was mixed with 3 µL of 1× Orange G loading dye (10 mM Tris-HCl pH 7.6, 0.15% (w/v) Orange G dye, 60% (v/v) glycerol, 60 mM EDTA) and run in an 0.8% (w/v) agarose gel in 1× SB buffer, as described before. The RNA samples were stored at -80 °C.

3.3.3. cDNA synthesis

Right after the RNA quantity and integrity analysis, the RNA samples were initially processed to eliminate genomic DNA contamination using DNase I, RNase free (Thermo Scientific™), according to the manufacturer's instructions, only adding 1 µL of Ribolock RNase Inhibitor (Thermo Scientific™) in the first step of this treatment. Immediately after

this process, the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific™) was used to synthesise first strand cDNA, following the instructions manual. These reactions are summarised in Table 5. The cDNA synthesis was performed using a T100™ Thermal Cycler (Bio-Rad).

Table 5. Genomic DNA degradation and reverse transcription reactions mixture composition.

Genomic DNA degradation reaction		Reverse transcription reaction	
Reagents	Quantity	Reagents	Quantity
RNA sample	1 µg	Oligo (dT) ₁₈ primer	1 µL
10× reaction buffer with MgCl ₂	1 µL	5× Reaction Buffer	4 µL
DNase I, RNase-free	1 µL (1U)	10 mM dNTP Mix	2 µL
RiboLock RNase Inhibitor (20 U/µL)	1 µL	RevertAid M-MuLV RT (200 U/µL)	1 µL
50 mM EDTA	1 µL		

3.3.4. Semi-quantitative and quantitative RT-PCR

The integrity and relative quantity of the cDNA samples obtained in the previous step were tested performing a semi-quantitative RT-PCR reaction that amplifies a region of the *ACTIN8* (*ACT8*) gene using a T100™ Thermal Cycler (Bio-Rad). The primers used in this PCR, the PCR mixture composition and the conditions applied are listed in Tables 6, 7 and 8, respectively. After the PCR, 2 µL of the product of each reaction was then mixed with 3 µL of 1× Orange G loading dye and submitted to an electrophoresis as described in section 3.2., but using Lithium Acetate Borate (LAB) buffer (10mM lithium acetate, 10mM boric acid) as the electrophoresis buffer.

Table 6. List of the sequences of the primers used in the semi-quantitative RT-PCR.

Name	Primer orientation	Sequence (5'-3')	Expected band size
ACT8-FW	Forward	CTCAGGTATTGCAGACCGTATGAG	185 bp
ACT8-RV	Reverse	CTGGACCTGCTTCATCATACTCTG	

Table 7. Semi-quantitative RT-PCR mixture composition.

Reagents	Quantity (µL)
Dream Taq 10× buffer	2.5
Nucleotide mixture (dNTPs) (2 mM)	0.5
Taq DNA polymerase	0.13
Forward Primer (20 µM)	1
Reverse Primer (20 µM)	1
cDNA sample	1
Nuclease-free water	18.87

Table 8. Conditions applied in the semi-quantitative RT-PCR.

Step		Temperature (°C)	Time
Initial denaturation		95	2 minutes
35 cycles	Denaturation	95	30 seconds
	Annealing	57	30 seconds
	Extension	72	40 seconds
Final extension		72	5 minutes
Pause/cooling		8	+∞

As for the quantitative RT-PCR (RT-qPCR), the cDNA amplification used SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), CFX96™ Real-Time PCR Detection System C1000 Touch™ (Bio-Rad) and the primers listed in Table 9. *ACT8* and *RUB1* were defined as reference genes. Each reaction contained a volume of 20 µL and each condition (i.e. same target gene and same cDNA sample) was replicated 3 times. The conditions applied in this RT-qPCR are present in Table 10. The data collected was analysed using the CFX Maestro™ Software (Bio-Rad).

Table 9. List of the sequences of the primers used in the RT-qPCR.

Name	Primer orientation	Sequence (5'-3')
AGP9-RT-qPCR-Fw	Forward	ATCTGTATCGTCCTCATCG
AGP9-RT-qPCR-Rv	Reverse	ATGTTGTGACTGGTGGTG
AGP21-RT-qPCR-Fw	Forward	GCAATGAAGATGAAGATGATG
AGP21-RT-qPCR-Rv	Reverse	GAAACCAGATGCCAAAGC
AGP24-RT-qPCR-Fw	Forward	TTTGTCTATTGGCGACGATG
AGP24-RT-qPCR-Rv	Reverse	ATAGCCAAGACGGTGAAC
AGP31-RT-qPCR-Fw	Forward	CCCCTTCTCTAGCTCCTGCT
AGP31-RT-qPCR-Rv	Reverse	GTGGAGACACTGGTGCCTTA
ACT8-RT-qPCR-Fw	Forward	CTCAGGTATTGCAGACCGTATGAG
ACT8-RT-qPCR-Rv	Reverse	CTGGACCTGCTTCATCATACTCTG
RUB1-RT-qPCR-Fw	Forward	CTGTTACGGAACCCAATTC
RUB1-RT-qPCR-Rv	Reverse	GGAAAAAGGTCTGACCGACA

Table 10. Conditions applied in the RT-qPCR.

Step		Temperature (°C)	Time
Initial denaturation		95	3 minutes
45 cycles	Denaturation	95	10 seconds
	Annealing and extension	60	30 seconds
	Melt Curve	55 – 95	0.5 °C increments each second
Pause/cooling		12	+∞

3.4. *In silico* analyses

3.4.1. CArG boxes and *cis* elements prediction

The sequences of the genes studied in this work were all obtained from the SeqViewer tool (<https://seqviewer.arabidopsis.org/>), available on TAIR (<https://www.arabidopsis.org/>). With regards to the CArG boxes and *cis* elements prediction, the genomic region considered for each gene goes from the first base pair

after the 3' UTR of the upstream gene until the last base pair of the 3' UTR of the gene being analysed.

Based in what was reported by Matias-Hernandez *et al.* (2010) and concerning the CArG boxes identification, potential sequences of these motifs were identified in *AGP9*, *AGP24* and *AGP31* genes, establishing a criterion according to which the sequences should contain a maximum of three differences between themselves and the CArG box exact sequence [CC(A/T)₆GG], either it is a deletion, an insertion, or a replacement. The last twelve identified CArG boxes upstream of the ATG were selected, although potential CArG boxes were identified throughout all the genomic region considered, including introns and exons. These criteria allowed to define, in an acceptably rigorous manner, enough putative CArG boxes to develop this work, being all of them present in the gene promoter or 5' UTR. The twelve CArG boxes selected for each gene are present in Figure 11 and Table 11. The sequences of the oligonucleotides designed for the EMSA technique that contained the CArG boxes are listed in Annex I, Supplemental table 1.

As for the selection of the fragments that would be designated as the *cis* elements for the PTA technique, the methodology used followed what was previously described by Ueda *et al.* (2011). Within the genomic region examined, three sequences were initially tested and isolated: 5' UTR, 3' UTR and a promoter sequence that begins 1000 bp upstream and ends immediately before the 5' UTR.

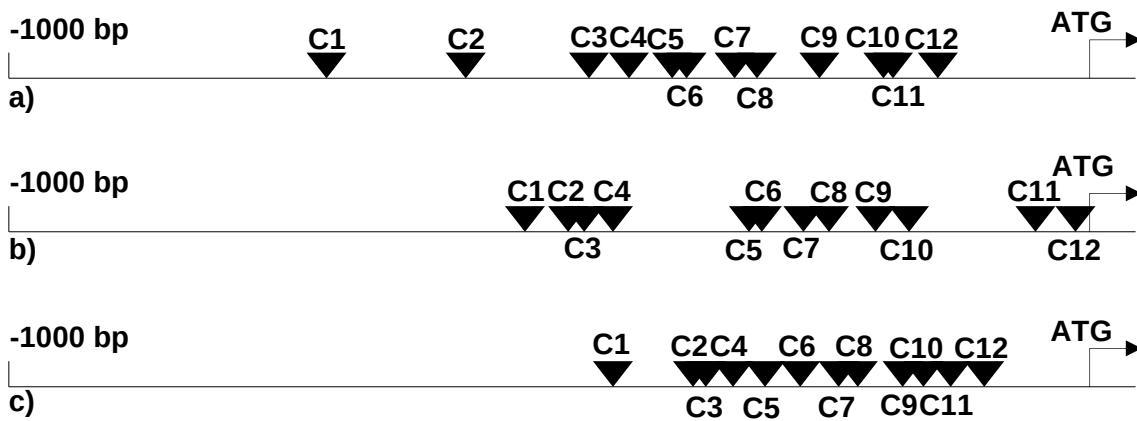


Figure 11. Schematic representation of the position of the twelve selected CArG boxes (C1 to C12) in a section of the promoter region near the ATG (1000 bp upstream of the ATG) of the *AGP9* (a), *AGP24* (b) and *AGP31* (c) genes.

Table 11. Sequences of the twelve selected CArG boxes (C1 to C12 in Annex I, Supplemental table 1) present in the *AGP9*, *AGP24* and *AGP31* genes.

Gene	CArG box	Sequence (5'-3')	Gene	CArG box	Sequence (5'-3')	Gene	CArG box	Sequence (5'-3')
AGP9	C1	CAAATAAGG	AGP24	C1	CCATAATTAAAGG	AGP31	C1	CTAGATGGG
	C2	CTTTGAAGG		C2	CCTTCATTG		C2	CTTTTGTTTGG
	C3	CTAGATATTTG		C3	CCATAAG		C3	CGTAATAG
	C4	CTTATTTG		C4	CTTTATG		C4	CATTTTTTCAG
	C5	CAAATAAAGG		C5	CCATATAATTGG		C5	CCATTTTAATG
	C6	CAAAAAG		C6	CCAATTATAG		C6	CCATCTTG
	C7	CATGTTTGG		C7	CTTTAACAG		C7	CAAAAAAAG
	C8	CTTAAAAG		C8	CATCAAAAAG		C8	CAAAATGG
	C9	CATAATTTTG		C9	CTTATTTTATG		C9	CACAAAATGG
	C10	CTCACATGG		C10	CTTATTTTAAG		C10	CTAATAATGGGG
	C11	CATCATGGG		C11	CATCAAAAAG		C11	CGATTTTG
	C12	CTTCAATACG		C12	CGTTTTTG		C12	CAAATAAAAGG

3.4.2.sgRNAs design for the CRISPR-Cas9 technique

With regards to the CRISPR-Cas9 technique, the sequence of each target gene obtained in the SeqViewer tool was analysed in the CRISPR-P 2.0 platform (<http://crispr.hzau.edu.cn/CRISPR2/>; Liu *et al.*, 2017). This platform helped in choosing the appropriate and specific sequence of the sgRNAs for each target gene (*AGP1*, *AGP4*, *AGP9*, *AGP15*, *AGP21*, *AGP24* and *AGP31*). Although only the *AGP9*, *AGP21*, *AGP24* and *AGP31* genes were studied in the present work, sgRNAs were designed and constructs were assembled for the knockout of the *AGP1*, *AGP4* and *AGP15* genes, as well as of the *CKX5* and *CKX7* genes (data not shown). Two sgRNAs specific cassettes were designed for each construct with the purpose of creating a deletion in the target gene after both cuts effected by Cas9, one per each sgRNA. Furthermore, for each gene two constructs were designed and used in the same plant transformation in order to increase the probability of the knockout. So, four sgRNAs specific cassettes were developed for each target gene. In order to improve the efficiency of this technique, the sgRNAs specific sequences selected (Table 12) respect the following rules:

- exhibit a score higher than 0.3 (Doench *et al.*, 2014);
- are designed for the target genes' exons in order to ensure the knockout;
- present a minimum number of off-targets;
- present a minimum number of tandem repeats;
- start with a guanine (G), since the sgRNA expression will be guided by a U6 promoter, which requires a G in order to start the transcription (Mali *et al.*, 2013);
- are designed for preferably the same exon, but for a region the furthest possible, generally 200 bp, from the region correspondent to the other sgRNA of the same construct, with the objective of creating the bigger deletion possible.

Table 12. List of the sgRNAs specific sequences selected for the *AGP1*, *AGP4*, *AGP9*, *AGP15*, *AGP21*, *AGP24* and *AGP31* genes. Two pairs of sgRNAs cassettes were designed for each gene.

Gene		sgRNAs specific sequences (5'-3')	
AGP1	1 st pair	1 st sgRNA	GAAGGAGAACTTCAGGTGC
		2 nd sgRNA	GACAACCTAGATAACCAAAA
	2 nd pair	1 st sgRNA	GCGGCTGCCGGAGAAGGAGA
		2 nd sgRNA	GACCGCCACAGATCCGAAAC
AGP4	1 st pair	1 st sgRNA	GCAGGAGAAGGAGCAACTGA
		2 nd sgRNA	GCCTCTCCACCAGCACCGGA
	2 nd pair	1 st sgRNA	GGAGAAGGAGCAACTGACGG
		2 nd sgRNA	GTACATAATAGCGGCGAAGG
AGP9	1 st pair	1 st sgRNA	GGTGGTGCTGAGACAGGAGG
		2 nd sgRNA	GGAGAGATAGAATCGGTGCT
	2 nd pair	1 st sgRNA	GGAGGAGGAGAAGAGACTGG
		2 nd sgRNA	GGAAGAGGTGGGCTAGATGA
AGP15	1 st pair	1 st sgRNA	GACCATCATGAGAACAACAA
		2 nd sgRNA	GCAGCGTCTTTCGTGTCAGC
	2 nd pair	1 st sgRNA	GATTATCTCCGTCTGTCAT
		2 nd sgRNA	GTAGATCCAGAAGTAGGACT
AGP21	1 st pair	1 st sgRNA	GCAGCATCAGAAGTTGGGCT
		2 nd sgRNA	GGTGGCTTTCTCAGCTGCCA
	2 nd pair	1 st sgRNA	GGCTTCAACGGTAGCAGCCG
		2 nd sgRNA	GAACAGTGCTGGTACAAACA
AGP24	1 st pair	1 st sgRNA	GATGTTTGTTTCAGATCGCTG
		2 nd sgRNA	GGAACCAACGACGAGAGCCA
	2 nd pair	1 st sgRNA	GCCGTTGTGTCCGCCACGA
		2 nd sgRNA	GTAATAACCAACTTAGTGGT
AGP31	1 st pair	1 st sgRNA	GCTCCTGCTCCTTACCACCA
		2 nd sgRNA	GGAAACAGGTGGCTTGACCG
	2 nd pair	1 st sgRNA	GAGCAAGAAGAACATAACAG
		2 nd sgRNA	GTATACGGTCTCTTTAACGT

3.5. Electrophoretic Mobility Shift Assay

3.5.1. Introduction to the technique

To perform an EMSA, firstly the coding sequence of the gene that codifies the protein of interest needs to be inserted in an expression vector. Then, the transcription and translation of this gene is induced and the protein is extracted and purified. Meanwhile, in what concerns the DNA fragment under study, initially, the fragments were usually radiolabelled. Nowadays, the researchers commonly use digoxigenin labelled probes, due to safety and durability issues. It starts with the annealing and cooling of the oligonucleotides of interest which are then tagged using a digoxigenin labelling kit. After testing the efficiency of this last reaction with a dot blot, the DNA probe and the protein of interest are mixed with a buffer and this mixture is electrophoresed in a polyacrylamide gel. The gel is then transferred to a nylon membrane which is blocked and subsequently incubated with anti-digoxigenin antibody. After going through a washing and equilibration process, the membrane is covered with a detection substrate and blotted on a paper for further analysis of the results (Kass *et al.*, 2000).

3.5.2. STK coding sequence amplification

In order to obtain the STK purified protein, firstly the STK coding sequence was amplified in order to insert it in an expression vector. To achieve this, WT RNA was extracted according to the second RNA extraction approach reported in sections 3.3.1. and 3.3.2.. Right after RNA quantification and integrity analysis, cDNA was synthesised as it is mentioned in section 3.3.3., using 2660 ng of RNA and a specific reverse primer, instead of using the Oligo (dT)₁₈ primer, and altering the reverse transcription reaction temperature from 42 °C to 50 °C. Subsequently, the STK coding sequence was amplified using the cDNA synthesised, the same reverse primer and a specific forward primer (primers' sequences in Table 13) through a PCR, whose conditions are listed in Table 14. The PCR was performed in a T100™ Thermal Cycler (Bio-Rad) and the reaction mixture composition included DreamTaq RNA Polymerase (Thermo Scientific™) and was prepared as mentioned in section 3.3.4.. After the amplification, 2 µL of the reaction product was mixed with 3 µL of 1× Orange G loading dye and submitted to an electrophoresis as described in section 3.2., using LAB buffer as the electrophoresis buffer.

The specific primers were designed with the purpose of cloning the complete STK coding sequence into the pET28a (+) expression vector (Studier and Moffatt, 1986; Rosenberg *et al.*, 1987; Studier *et al.*, 1990; Annex II, Supplemental figure 1). Each primer contained a specific region of the STK sequence, and an upstream non-specific region, that included the recognition sequence for each restriction enzyme (EcoRI in the forward primer and Sall in the reverse primer) and an adapter in the 5' end. The specific region of the forward primer contained the first 20 nucleotides of the first exon of the STK gene and the specific region of the reverse primer was the reverse-complement of the last 24 nucleotides of the last exon of the same gene, excluding the last three nucleotides that codify for the stop codon (Table 13). After the PCR, the amplified fragment was purified using the GeneJET PCR Purification Kit (Thermo Scientific™), according to the instructions manual.

Table 13. Sequences of the primers designed for the STK coding sequence amplification and cloning into the pET28a (+) expression vector. EcoRI (forward primer) and Sall (reverse primer) recognition sequences are underlined. These primers also included an adapter upstream of each recognition sequence and a downstream specific sequence that corresponds to the 5' end of the first STK exon in the forward primer and to the 3' end of the last STK exon, excluding the last three nucleotides that codify for the stop codon, in the reverse primer.

Name	Primer orientation	Sequence (5'-3')	Expected band size
STK-pET28a-FW	Forward	TATGGAATTCATGCTCTTTCCCATGAAAG	788 bp
STK-pET28a-RV	Reverse	TATGGTCTGACTCCGAGATGAAGAATTTTCTTGTC	

Table 14. Conditions applied in the PCR for the STK coding sequence amplification.

Step		Temperature (°C)	Time
Initial denaturation		95	2 minutes
35 cycles	Denaturation	95	30 seconds
	Annealing	53	30 seconds
	Extension	72	1 minute
	Final extension	72	5 minutes
Pause/cooling		8	+∞

3.5.3. Plasmid and amplified fragment double digestion and plasmid dephosphorylation

E. coli colonies transformed with the plasmid pET28a and incubated in Luria-Bertani (LB) solid medium supplied with 1.5% (w/v) agar and kanamycin (50 µg/mL) were kindly provided by Professor Luís G. Pereira. Three colonies were collected with toothpicks from the Petri dishes with the solid medium and were separately inoculated in 15 mL tubes with 5 mL LB liquid medium provided with kanamycin in the same concentration. These cultures were incubated overnight (ON) at 37 °C with agitation (3 g) in an orbital shaker-incubator (CFL 3031). Afterwards, 2 mL of each liquid culture were centrifuged at 12100 g for 2 minutes using miniSpin (Eppendorf™), being the resulting supernatant discarded. Subsequently, the plasmid DNA from each sample was extracted as reported by the manufacturer's manual of the GeneJET Plasmid Miniprep Kit (Thermo Scientific™).

Prior to the ligation of the fragment amplified in the previous step with the expression vector, both parts were digested with EcoRI and Sall restriction enzymes, being the reaction mixture composition described in Table 15. The mixture was incubated for 2 hours at 37 °C, being then incubated at 65 °C for 20 minutes in order to inactivate the restriction enzymes. A volume of 2 µL of the product of each reaction was then mixed with 3 µL of 1× Orange G loading dye and submitted to an electrophoresis as described in section 3.2., using LAB buffer as the electrophoresis buffer.

After this step, the plasmid was dephosphorylated by adding 2.5 µL of alkaline phosphatase (FastAP™ Thermosensitive Alkaline Phosphatase, Thermo Scientific™) and 5 µL of the respective buffer (10× FastAP™ Buffer, Thermo Scientific™). The mixture was incubated at 37 °C for 20 minutes and then at 65 °C for 15 minutes, in order to block the alkaline phosphatase activity. Finally, the products of the amplified fragment digestion and of the linearised plasmid dephosphorylation were purified using the GeneJET PCR Purification Kit (Thermo Scientific™), according to the instructions manual, and a volume of 2 µL of each purification was then submitted to an electrophoresis as described above in this section.

Table 15. Restriction reaction mixture composition.

Reagents	Quantity (μL)	
	Fragment	Plasmid
Buffer O (10X) (Thermo Scientific™)	5	5
EcoRI (Thermo Scientific™)	1.5	1.5
Sall (Thermo Scientific™)	1.5	1.5
DNA	42 (all the PCR purified product)	10
Nuclease-free water	0	32

3.5.4. Ligation and restriction analysis

The linearised plasmid and the amplified fragment were mixed in three different volume proportions (1:1, 1:2 and 1:3). The ligation reactions mixtures composition are described in Table 16 and were held at 4 °C ON. After these reactions and intending to precipitate the respective products, nuclease-free water was added to each reaction to a volume of 200 μL, followed by 2 μL of glycogen (Thermo Scientific™), 10 μL of 5 M NaCl and 500 μL of 96% (v/v) ethanol. This mixture was kept at -20 °C for 30 minutes and then it was centrifuged during 20 minutes at 16100 g. Finally, the dry pellet was dissolved in 10 μL of nuclease-free water.

In order to verify the ligation of the expression vector with the amplified fragment, the final products of the ligation procedure were subjected to a restriction analysis. The restriction reactions were maintained at 37 °C for 15 minutes and then at 80 °C for 5 minutes. The mixture composition of these reactions is described in Table 17. The products of each reaction were electrophoresed as stated in section 3.2., using LAB buffer as the electrophoresis buffer and without adding loading dye, since the used buffer already includes it.

Table 16. Ligation reactions mixtures composition.

Reagents	Quantity (μL)		
	1:1	1:2	1:3
10× T4 DNA Ligase buffer (Thermo Scientific™)	1	2	3
T4 DNA Ligase (Thermo Scientific™)	0.3	0.3	0.3
Amplified fragment	2	2	2
Linearised plasmid	6	12	18
Nuclease-free water	0.7	3.7	6.7

Table 17. Restriction analysis mixture composition.

Reagents	Quantity (μL)
10× FastDigest Green Buffer (Thermo Scientific™)	2
EcoRI (FastDigest, Thermo Scientific™)	1
MluI (FastDigest, Thermo Scientific™)	1
Recombined plasmid	2
Nuclease-free water	14

3.6. Protoplast Transactivation Assay

3.6.1. STK gene cloning

The gene that expresses the TF being studied in a PTA needs to be inserted in protoplasts and, to do so, a miniprep sample of the vector used for this purpose from Ueda *et al.* (2011), MU1110 (Annex III, Supplemental figure 2), was kindly provided by Professor Minako Ueda and was quantified using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix), according to the manufacturer's instructions, and digested with Sall and NotI restriction enzymes, following the procedures described for the restriction reaction referred in section 3.5.3. and preparing a reaction mixture as described in Table 18. The product of this reaction was mixed with 3.2 µL of 1× Orange G loading dye and submitted to an electrophoresis as described in section 3.2.. The band representative of the linearised plasmid was extracted and purified using the GeneJET Gel Extraction Kit (Thermo Scientific™), according to the instructions manual. The final purified and linearised plasmid was quantified again using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix), according to the manufacturer's instructions.

Table 18. Restriction reaction mixture composition.

Reagents	Quantity (µL)
Buffer O (10X) (Thermo Scientific™)	2
Sall (Thermo Scientific™)	0.5
NotI (Thermo Scientific™)	0.5
Plasmid DNA	1 µg
Nuclease-free water	Up to 20 µL

As for the *STK* gene amplification and cloning into this vector, specific primers were designed to amplify the *STK* gene sequence and clone it into the MU1110 vector. These primers were designed for the Gibson Assembly cloning technique (Gibson *et al.*, 2009; Table 19), enabling to perform conventional directional cloning as well. The *STK* genomic sequence was amplified from the 5' end of the first exon until the 3' end of the last one, including introns, since it was going to be inserted in protoplasts, not in prokaryotic cells, not being necessary the use of the *STK* coding sequence.

This PCR was performed using genomic DNA and the high-fidelity PrimeSTAR® HS DNA Polymerase (TAKARA) and a TaKaRa PCR Thermal Cycler Dice® Gradient. The composition of the reaction mixture and the conditions applied in the reaction are present in Tables 20 and 21. The result of this PCR was analysed through electrophoresis in an 1.5% (w/v) agarose gel on 1× TAE buffer (40 mM Tris-acetate and 1 mM EDTA pH 8.3) with 0.5 mg/mL ethidium bromide. A small volume of the final product was mixed with 6× purple dye (2.5% (w/v) Ficoll®-400, 10 mM EDTA, 3.3 mM Tris-HCl, 0.08% (w/v) SDS

and 0.02% (w/v) Dye 1 (pink/red) and 0.0008% (w/v) Dye 2 (blue)) (New England Biolabs®) before the electrophoretic run. 1× TAE buffer was also used as running buffer and Gene Ladder 100 (Nippon Gene) as molecular-weight size marker. The electrophoresis was performed at 135 V and 20 minutes. DNA viewing was done with a UV transilluminator (302-365 nm). The images were obtained with a NIKON camera and a CCD camera control software. The band corresponding to the fragment amplified was immediately excised from the agarose gel and the DNA was recovered with the NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) kit, according to the manufacturer's instructions.

Table 19. Sequences of the primers designed for the *STK* genomic sequence amplification and cloning into the MU1110 vector. Sall (forward primer) and NotI (reverse primer) recognition sequences are underlined. In the 5' end of each primer, there is a specific sequence that corresponds to the plasmid sequence located upstream (forward primer) and downstream (reverse primer) of the restriction enzymes' recognition sequences present in the plasmid. In the 3' end of each primer, there is a specific sequence that corresponds to the 5' end of the first *STK* exon (forward primer) and to the 3' end of the last *STK* exon (reverse primer). In the reverse primer, between the NotI recognition sequence and the *STK* specific sequence, there are six extra nucleotides (in bold), that codify for additional stop codons.

Name	Primer orientation	Sequence (5'-3')	Expected band size
STK-MU-FW	Forward	TTCGCCCTTGGAGTCGACATGCTCTTCCCCATGAA	3825 bp
STK-MU-RV	Reverse	TCTGCAGCCGGCGGCCGCTATCATTATCCGAGATGAAGAATTTCT	

Table 20. PCR mixture composition for the amplification of the *STK* genomic sequence.

Reagents	Quantity (μL)
5× PrimeSTAR GXL Buffer	3
2.5 mM dNTP mix	1.20
50 μM Forward Primer	0.15
50 μM Reverse Primer	0.15
Genomic DNA	1
PrimeSTAR GXL DNA polymerase	0.15
Nuclease-free H ₂ O	9.35

Table 21. Conditions applied in the PCR for the *STK* genomic sequence amplification.

Step	Temperature (°C)	Time
Initial denaturation	95	2 minutes
35 cycles	Denaturation	95
	Annealing	55.1
	Extension	72
Final extension	72	5 minutes
Pause/cooling	8	+∞

The cloning method used was the conventional directional cloning. The amplified sequence was digested following the same procedure of the MU1110 restriction reaction. The respective product was also submitted to an electrophoresis and the band corresponding to the digested fragment was also extracted and purified using the GeneJET Gel Extraction Kit (Thermo Scientific™), according to the instructions manual.

The final purified and amplified fragment was quantified using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix), according to the manufacturer's instructions.

Finally, relatively to the ligation reaction, the linearised plasmid and the amplified *STK* genomic sequence were mixed with T4 DNA ligase and the respective buffer, following the mixture composition referred in Table 22. In order to verify that the correct recombined vector was formed, the product of this reaction was subjected to a restriction analysis. The restriction reaction were maintained at 37 °C for 2 hours and then at 65 °C for 5 minutes. The mixture composition of this reaction is described in Table 23. A volume of 2 µL of the product was then mixed with 3 µL of 1× Orange G loading dye and submitted to an electrophoresis as described in section 3.2., using LAB buffer as the electrophoresis buffer.

Table 22. Ligation reaction mixture composition. The linearised plasmid and amplified fragment concentrations were measured using a DS-11 FX Spectrophotometer/Fluorometers (DeNovix), according to the manufacturer's instructions. *The quantity of amplified fragment used in this reaction was calculated with NEBioCalculator™ v1.10.1 (New England Biolabs®), which advises to use a 2:1 insert:vector molar ratio.

Reagents	Quantity (µL)
10× T4 DNA Ligase buffer (Thermo Scientific™)	2
T4 DNA Ligase (Thermo Scientific™)	1
Amplified fragment	20 ng 2 DNA template:1 backbone vector
Linearised plasmid	
Nuclease-free water	
	Up to 20 µL

Table 23. Restriction analysis mixture composition.

Reagents	Quantity (µL)
Buffer O (10X) (Thermo Scientific™)	2
Sall (Thermo Scientific™)	1
Recombined plasmid	2
Nuclease-free water	15

3.6.2. *cis* elements identification

3.6.2.1. Amplify putative *cis* elements

As described by Ueda *et al.* (2011), after selecting *AGP24* and *AGP31* putative *cis* elements *in silico*, primers were designed for amplification and cloning of 4 copies of each element into the pBluescript II SK (+) vector (resistant to ampicillin; Alting-Mees and Short, 1989; Annex IV, Supplemental figure 3) for further recombination with a reporter gene.

The primers used in the amplification of *AGP24* putative *cis* elements were designed for Gibson Assembly cloning of the first copy of each element, enabling to perform conventional directional cloning of the other copies as well. The ones used in the

amplification of *AGP31* putative *cis* elements were designed only for conventional directional cloning, using XhoI and SalI restriction enzymes and ligase. The primers designed are present in Table 24.

The PCR was performed in a T100™ Thermal Cycler (Bio-Rad). The reaction mixture composition for the amplification of *AGP24* (promoter and 5' UTR) and *AGP31* (promoter, 5' UTR and 3' UTR) putative *cis* elements included genomic DNA and DreamTaq RNA Polymerase (Thermo Scientific™) and was prepared as mentioned in section 3.3.4.. The reaction conditions used in these amplifications are present in Annex V, Supplemental table 2. 1× Orange G loading dye was added to each reaction product, which were then submitted to an electrophoresis under the conditions described before in section 3.2., using LAB buffer as the electrophoresis buffer.

Table 24. Sequences of the primers designed for the *AGP24* and *AGP31* genes predicted *cis* elements amplification and cloning into pBluescript II SK (+). XhoI (forward primer) and SalI (reverse primer) recognition sequences are underlined. These primers included an adapter upstream of each recognition sequence and a downstream sequence specific of the potential *cis* element being amplified. Relatively to the primers designed for the amplification of *AGP24* putative *cis* elements, the region upstream of the restriction enzymes recognition sequences are specific of the pBluescript II SK (+) vector, as it is required by the Gibson Assembly cloning system (Gibson *et al.*, 2009).

Gene	<i>cis</i> element	Primer orientation	Primer (5'-3')
<i>AGP24</i>	Promoter	Forward	TACCGGGCCCCCCTCGAGTTTATAATTGTAGTTTTATCAAGAAGA
		Reverse	TCGATACCGTCGACGTCGACGTAATAGGTTTGTGAGATTTGT
	5' UTR	Forward	TACCGGGCCCCCCTCGAGACAACAACACACAAATCT
		Reverse	TCGATACCGTCGACGTCGACTATCTTCAAAAACGTTTGA
	3' UTR	Forward	TACCGGGCCCCCCTCGAGGTTGGTTATTACATTGCTCGT
		Reverse	TCGATACCGTCGACGTCGACATATACTGCTGCTTCTACTAATAAAAATT
<i>AGP31</i>	Promoter	Forward	TATACTCGAGTGTTATACATACTACTTTCAAAAC
		Reverse	TATAGTCGACTATTAGTTACTAGTACCATTTTGTG
	5' UTR	Forward	TATACTCGAGATGGGGTAATTCATGTC
		Reverse	TATAGTCGACCTCTTACCAATGAAACCC
	3' UTR	Forward	TATACTCGAGAGATGATTTGATGCTCTGG
		Reverse	TATAGTCGACGCTCATTGTTTGTAAATG

3.6.2.2. Plasmid miniprep and restriction analysis

As for the vector used, pBluescript II SK (+), *E. coli* colonies transformed with this plasmid and incubated in Luria-Bertani (LB) solid medium supplied with 1.5% (w/v) agar and ampicillin (50 µg/mL) were kindly provided by Professor Luís G. Pereira. The miniprep of this plasmid was executed as mention in section 3.5.3.. In order to confirm the extraction of pBluescript II SK (+), the samples of plasmid DNA collected were submitted to a restriction analysis using KpnI restriction enzyme. This step was only accomplished after testing different periods for the reaction incubation at 37 °C. The reaction conditions are present in Table 25. The product of each reaction was then submitted to an electrophoresis under the same conditions described in section 3.2., using LAB buffer

as the electrophoresis buffer and without adding a loading dye, since the reaction buffer already includes one.

Table 25. Reaction mixture composition and conditions of the restriction analysis of the plasmid DNA using KpnI restriction enzyme (Thermo Scientific™).

Reagents	Quantity (μL)	Temperature (°C)	Time (minutes)
Nuclease-free water	Up to 20 μL	37	15
10× FastDigest Green Buffer	2		
Plasmid DNA	1 μg		
FastDigest KpnI	1		

3.7. Creation of knockout mutant plants using the CRISPR-Cas9 technique

3.7.1. Assembly of constructs with two sgRNA cassettes for plant transformation

3.7.1.1. Bacterial strains and vectors used

In this section of the study, the bacterial strains used were: *Escherichia coli* Mach1 strain (Thermo Scientific™), since it presents a very fast growth, and *Agrobacterium tumefaciens* GV310:pMP90, which is resistant to rifampicin and gentamicin. The vectors used were the following: pCBC-DT1T2, as a template vector (Xing *et al.*, 2014; Annex V, Supplemental figure 4), and pHEE401E+BliI (pHEEBliI), as a backbone vector (Wang *et al.*, 2015; Annex V, Supplemental figure 5). Both the bacterial strains and the vectors used were gently provided by Dr. Higashiyama's lab. The vectors came from miniprep protocols.

3.7.1.2. Growth media and conditions

The medium used to grow *E. coli* and *Agrobacterium tumefaciens* cells was the LB medium supplied with the adequate antibiotic after sterilisation. The solid media were also supplied with agar and both solid and liquid media were stored at 4 °C. While the *E. coli* cultures were incubated at 37 °C for 12 to 16 hours, using an orbital shaker-incubator (CFL 3031) for the liquid cultures, the *A. tumefaciens* cultures were incubated at 28 °C for 2 days, using a Shel Lab's Shaking Incubator for the liquid cultures.

3.7.1.3. Acquisition of sgRNA cassettes within an amplified template

An initial PCR was performed using primers containing the sgRNAs specific cassettes designed previously (Table 12, section 3.4.2.) and the pCBC-DT1T2 vector as template. Each forward primer included, from the 5' end to the 3' end, an adapter, the BsaI recognition and cutting sequences, being this last sequence equivalent to the U6-26

promoter sequence (also present in the backbone plasmid), the first sgRNA specific cassette and a sequence from the template, which, in this case, corresponds to the first sgRNA scaffold. With regards to the reverse primers, each one was composed, in the same orientation, by an adapter, the BsaI recognition and cutting sequences, matching this last part to the reverse-complement sequence of the second sgRNA scaffold sequence (also present in the backbone plasmid), the reverse-complement sequence of the second sgRNA specific cassette and a sequence from the template, corresponding here to the reverse-complement sequence of the U6-29 promoter sequence. Bearing in mind that the region of the amplified template included the sequences of the first sgRNA scaffold, the U6-26 terminator and the U6-29 promoter, the whole resulting amplified fragment contained an adapter and BsaI recognition and cutting sequences in each end and, in between, it was formed, from the 5' end to the 3' end, by the first sgRNA specific cassette followed by the sequence of the first sgRNA scaffold, succeeded by the sequence of the second sgRNA promoter (U6-29 promoter) and, lastly, the second sgRNA specific cassette (Wang *et al.*, 2015). The primers designed are presented in Table 26. Figure 12 exhibits schematic representations of the primers (a)), the region of the template that was amplified (b)) and the whole amplified fragment (c)).

Table 26. Sequences of the primers designed for the amplification of the DNA template containing sgRNAs. Yellow, sgRNAs specific cassettes; Grey, adapters; Blue, BsaI recognition sequence; Dark purple, BsaI cutting sequence (U6-26 promoter); Light purple, template specific sequence (U6-29 promoter); Dark green, BsaI recognition sequence (second sgRNA scaffold); Light green, template specific sequence (first sgRNA scaffold) (Wang *et al.*, 2015).

GENE	Primers (5'-3')		
AGP1	1 st pair	Forward	atatatggtctcgattGAAGGAGAACTTCAGGTGCgttttagagctagaaatagc
		Reverse	attattggtctcgaagctTTTGGTTATCTAGGTTGTCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGCGGCTGCCGAGAGGAGAGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctGTTTCGGATCTGTGGCGGTCaatctcttagtcgactctac
AGP4	1 st pair	Forward	atatatggtctcgattGCAGGAGAAAGGAGCAACTGAgttttagagctagaaatagc
		Reverse	attattggtctcgaagctTCCGGTGCTGGTGAGAGGCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGGAGAAGGAGCAACTGACGGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctCCTTCGCCGCTATTATGTACaatctcttagtcgactctac
AGP9	1 st pair	Forward	atatatggtctcgattGGTGGTGCTGAGACAGGAGGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctAGCACCGATTCTATCTCTCCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGGAGGAGGAGAAGAGACTGGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctTCATCTAGCCCACCTCTTCCaatctcttagtcgactctac
AGP15	1 st pair	Forward	atatatggtctcgattGACCATCATGAGAACAAAGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctGCTGACACGAAAGACGCTGCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGATTATCTCCGTCGTTGCATgttttagagctagaaatagc
		Reverse	attattggtctcgaagctAGTCCTACTTCTGGATCTACaatctcttagtcgactctac
AGP21	1 st pair	Forward	atatatggtctcgattGCAGCATCAGAAGTTGGGGTgttttagagctagaaatagc
		Reverse	attattggtctcgaagctTGGCAGCTGAGAAAGCCACCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGGCTTCAACGGTAGCAGCCGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctGTGTTGTACCAGCACTGTTCaatctcttagtcgactctac
AGP24	1 st pair	Forward	atatatggtctcgattGATGTTTGTTCAGATCGCTGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctTGGCTCTCGTCTGTTGCCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGCCGTTGTGTCCGCCACAGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctACCACTAAGTTGGTTATTACaatctcttagtcgactctac
AGP31	1 st pair	Forward	atatatggtctcgattGCTCCTGCTCCTTACCACCAgttttagagctagaaatagc
		Reverse	attattggtctcgaagctCGGTCAAGCCACCTGTTTCCaatctcttagtcgactctac
	2 nd pair	Forward	atatatggtctcgattGAGCAAGAAGAACATAACAGgttttagagctagaaatagc
		Reverse	attattggtctcgaagctACGTTAAGAGACCGTATACaatctcttagtcgactctac

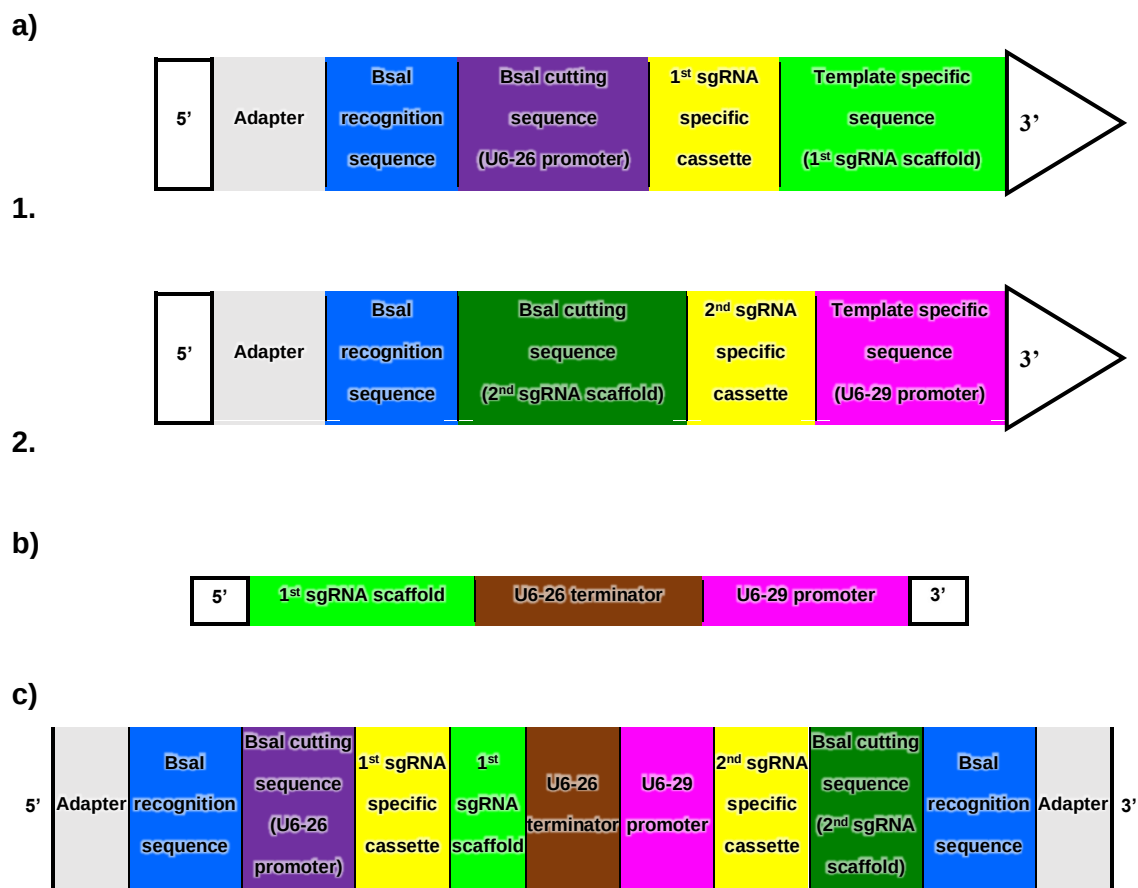


Figure 12. Schematic representation of: **a)** the forward primer (1.) and the reverse primer (2.); **b)** the region of the template that was amplified; **c)** the whole amplified fragment (Wang *et al.*, 2015).

These PCR were performed using the high-fidelity PrimeSTAR® HS DNA Polymerase (TAKARA) and a TaKaRa PCR Thermal Cycler Dice® Gradient. The composition of the reaction mixture, the conditions applied in the reaction and expected band size are present in Tables 27 and 28. The results of these PCR were mixed with 6× purple dye (New England Biolabs®) and analysed through electrophoresis as described in section 3.6.1..

Table 27. PCR mixture composition for the amplification of the DNA template containing sgRNAs.

Reagents	Quantity (µL)
5× PrimeSTAR GXL Buffer	3
2.5 mM dNTP mix	1.20
50 µM Forward Primer	0.15
50 µM Reverse Primer	0.15
DNA Template (pCBC-DT1T2) (1/200)	1
PrimeSTAR GXL DNA polymerase	0.15
Nuclease-free H ₂ O	9.35

Table 28. PCR conditions for the amplification of the DNA template containing sgRNAs and expected band size.

Step		Temperature (°C)	Time	Expected band size (bp)
35 cycles	Initial denaturation	98	2 minutes	626
	Denaturation	98	10 seconds	
	Annealing	55	5 seconds	
	Extension	68	30 seconds	
Pause/cooling		4	+∞	

3.7.1.4. DNA fragments extraction from gel and purification

DNA bands with the fragment of interest from the PCR described in the previous section were immediately excised from the agarose gel and the DNA was recovered with the NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel) kit, according to the manufacturer's instructions.

3.7.1.5. Golden Gate Assembly

Described for the first time in 1996, Golden Gate assembly (Engler *et al.*, 2008; Engler *et al.*, 2009) is a technique that enables the introduction of an insertion into a backbone vector through the sequential (Lee *et al.*, 1996) or simultaneous (Padgett and Sorge *et al.*, 1996) activities of a type IIs restriction enzyme and of the T4 or T7 DNA ligase. Type IIs restriction enzymes act out of the recognition sequence and create 4 bp overhangs, which, consequently, can be customised in order to perform direct cloning of the DNA fragments and leave no marks of the cloning after the process. This is why each primer designed for the amplification of the DNA template with the sgRNAs specific cassettes (section 3.4.2.) needed to include a BsaI (a type IIs restriction enzyme) cutting sequence that would correspond to the sequence from the backbone vector that surround downstream, in the forward primers, or upstream, in the reverse primers, the BsaI recognition sequence, which is also where the DNA template was inserted.

In this work, the Golden Gate assembly protocol used was based on Wang *et al.* (2015) with few changes. The plasmid pHEE401+BliI, used as backbone vector, and the DNA templates with two sgRNAs specific cassettes, amplified previously (section 3.7.1.3.), were cut with the BsaI restriction enzyme and linked with the T4 DNA ligase in the same reaction, in order to create the final constructs with two sgRNAs each. The Golden Gate assembly was performed using a TaKaRa PCR Thermal Cycler Dice® Gradient and the reaction composition and conditions are present in Tables 29 and 30. A schematic representation explaining the Golden Gate assembly is shown in Figure 13.

Table 29. Golden Gate assembly reaction composition. Backbone vector and DNA template concentrations were measured using a DS-11 FX Spectrophotometer/Fluorometers (DeNovix), according to the manufacturer's instructions. *The quantity of DNA template used in this reaction was calculated with NEBioCalculator™ v1.10.1 (New England Biolabs®), which advises to use a 2:1 insert:vector molar ratio.

Reagents	Quantity (μL)
Backbone vector (pHEE401+BliI)	100 ng
DNA template (pCBC-DT1T2 + 2 sgRNAs specific cassettes)	*
10× T4 DNA ligase buffer	0.30
0.1% (w/v) BSA	0.30
T4 DNA ligase	0.20
BsaI	0.20
Nuclease-free water	Up to 3 μL

Table 30. Golden Gate assembly reaction conditions.

Temperature (°C)	Time (minutes)	Nº of cycles
37	5	35
16	5	
55	10	1
80	5	

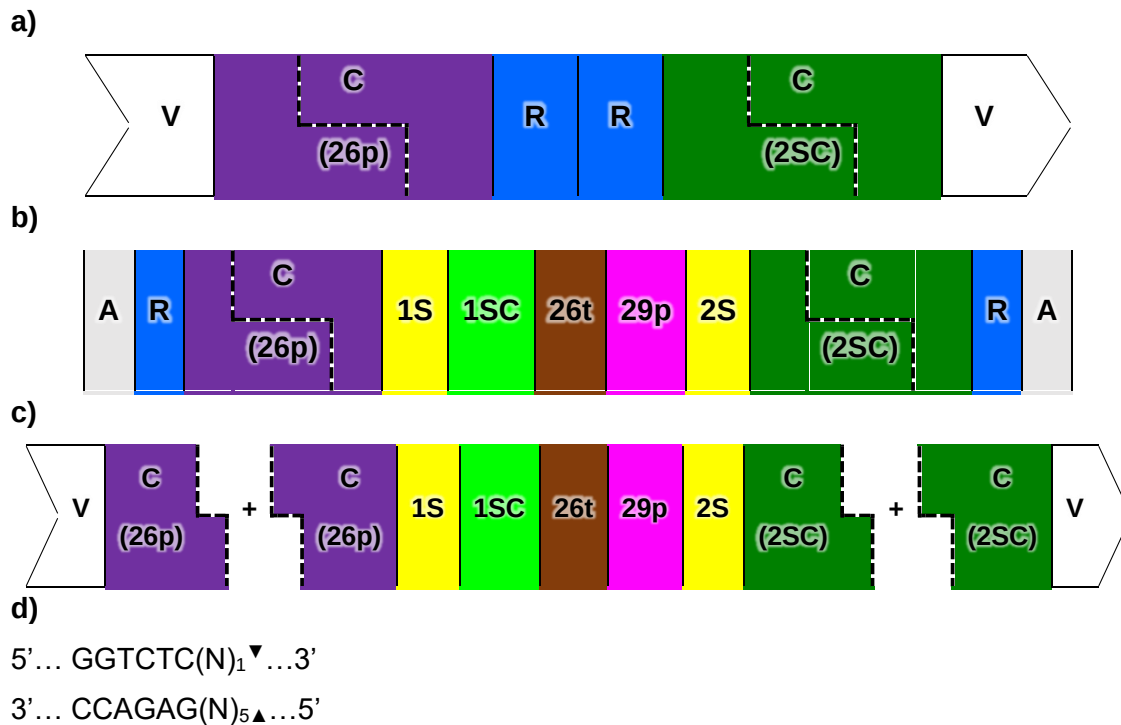


Figure 13. Schematic representation explaining the Golden Gate assembly. From the 5' end to the 3' end: **a)** region of the backbone vector where the DNA template was inserted; **b)** the whole amplified DNA template with both sgRNAs; **c)** assembly of the DNA template with the backbone vector; **d)** BsaI cutting pattern (A, adenine; C, cytosine; G, guanine; T, thymine; N, random nucleotide; ▼▲, cut site). 1S, 1st sgRNA specific cassette; 1SC, 1st sgRNA scaffold; 2S, 2nd sgRNA specific cassette; 26t, U6-26 terminator; 29p, U6-29 promoter; A, Adapter; C (26p), BsaI cutting sequence (U6-26 promoter); C (2SC), BsaI cutting sequence (2nd sgRNA scaffold); R, BsaI recognition sequence; V, surrounding regions of the backbone vector; +, action of the DNA ligase; dashed lines, BsaI cuts.

3.7.1.6. *E. coli* Mach1 and *A. tumefaciens* transformation

E. coli Mach1 were transformed using heat shock chemical transformation. To do so, 1 µL of DNA was added to 50 µL of previously frozen chemical competent cells thawed on ice and then this solution was placed on ice for 30 minutes. Immediately after this process, the solution was transferred to 42 °C for 45 seconds, followed by 2 minutes on ice. Lastly, the cells were incubated in 500 µL of SOC medium (Hanahan, 1983) on ice for 1 hour at 37 °C with shaking (4 g) and then they were precipitated through centrifugation at 8 g for 2 minutes at room temperature and plated on LB solid medium supplemented with 50 µg/mL kanamycin.

As for *A. tumefaciens*, initially, *Agrobacterium* cells from the GV3101:pMP90 strain were transformed using electroporation. Firstly, 50 µL of previously frozen *A. tumefaciens* competent cells were thawed on ice and mixed with the same amount of 15% (v/v) glycerol. Then, this solution was divided in two equal volumes and each one was mixed with 1 µL of one of the constructs assembled. Each mixture was then transferred to a cuvette which was placed on an Eporator™ (Eppendorf™), where the electric pulse (2.50 kV) was performed. Immediately after this step, 700 µL of LB liquid medium was loaded into the cuvette and the mixture was transferred into a microtube and incubated at 28 °C for 2 hours with shaking (4 g). Lastly, the culture was plated on LB solid medium supplemented with 50 µg/mL kanamycin and 10 µg/mL gentamicin and incubated for 2 days at 28 °C.

In a second approach, 50 µL of competent *Agrobacterium* cells from the same strain were mixed with 1 µL of one of the constructs created, being each mixture transferred to a cuvette, which, in turn, was placed in a MicroPulser™ electroporator (Bio-Rad). There, a 2.50 kV electric pulse was applied. The transformed bacteria were plated in LB solid medium supplemented with gentamicin (20 µg/mL), kanamycin (50 µg/mL) and rifampicin (100 µg/mL).

3.7.1.7. Colony-PCR

In order to verify which *E. coli* and *Agrobacterium tumefaciens* colonies were transformed with the correct construct, a colony-PCR was performed using the colonies formed after the transformation of *E. coli* cells and after the second transformation attempt of *Agrobacterium* cells. Previously designed primers that correspond to the sequences of the U6-29 promoter (forward primer) and of the U6-26 terminator (reverse primer) (Table 31) were also used in this PCR. The reaction was performed in a TaKaRa PCR Thermal Cycler Dice® Gradient and using the enzyme GoTaq® DNA Polymerase, the primers mentioned above and a small fraction of the colony being tested. The mixture

composition and PCR conditions are present in Tables 32 and 33. The results of these PCR were analysed through electrophoresis as described previously (section 3.6.1.). Each culture was tested using four colonies (or three in the colony-PCR performed with *Agrobacterium* cells) and the *E. coli* colonies transformed with the correct construct, presenting a 443 bp DNA fragment, were incubated in LB liquid medium supplemented with 50 µg/mL kanamycin for 12 to 16 hours at 37 °C with shaking (4 g).

Table 31. Sequences of the primers for colony-PCR. Size of the amplified fragment: 443 bp.

Name	Sequence (5'-3')
Forward primer (U6-29p)	TTAATCCAACTACTGCAGCCTGAC
Reverse primer (U6-26t)	AGAAATTGAACGCCGAAG

Table 32. Colony-PCR mixture composition.

Reagent	Quantity (µL)
GoTaq® DNA Polymerase	3
Forward primer (50 µM)	0.06
Reverse primer (50 µM)	0.06
Colony	(little fraction collected with a toothpick)
Nuclease-free water	3

Table 33. Colony-PCR conditions.

Step	Temperature (°C)	Time
Initial denaturation	94	1 minute
35 cycles	Denaturation	10 seconds
	Annealing	10 seconds
	Extension	1 minute
Final extension	72	5 minutes
Pause/cooling	15	+∞

3.7.1.8. Plasmid DNA extraction

Each plasmid extraction, performed before the sequencing step, was performed using the Wizard® Plus SV Minipreps DNA Purification System (Promega) kit according to manufacturer's instructions.

3.7.1.9. Sequencing

The sequencing reactions were performed according to the BigDye™ Terminator v3.1 Cycle Sequencing (Applied Biosystems™) kit, with small changes. The sequence of the primers used, the reaction composition and conditions are present in Tables 34, 35 and 36. Two primers were used so that the sequencing could comprise the regions of the plasmid DNA that include both sgRNAs specific cassettes. After this reaction, its product was mixed directly with 67 µL of a solution containing around 70% (v/v) ethanol and 90 mM CH₃COONa. This mix was vortexed for 2 seconds and centrifuged for 20 minutes at 16100 g and 4 °C. Then, the obtained supernatant was discarded and the sediment was

mixed with 100 μ L of 70% (v/v) ethanol and this mixture was centrifuged at 16500 g and 4 °C for another 20 minutes. The pellet formed was then air dried at room temperature for 30 minutes and, finally, it was mixed with 10 μ L of Hi-Di™ Formamide (Applied Biosystems™), incubated at 95 °C for 3 minutes and kept on ice for 5 minutes. The samples were sent for sequencing at the Center for Gene Research (Nagoya University, Japan) and the resulting sequences were analysed using the ApE software (A plasmid Editor).

Table 34. Primers used in the sequencing reaction.

Name	Sequence (5'-3')
U6-26p	TGTCCCAGGATTAGAATGATTAGGC
U6-29p	TTAATCCAACTACTGCAGCCTGAC

Table 35. Sequencing reaction mixture composition.

Reagent	Quantity (μ L)
DNA	200 ng
5× sequencing buffer	1
Primer (U6-26p or U6-29p) (5 μ M)	1
BigDye™ Terminator v3.1 Ready Reaction Mix	0.25
Nuclease-free water	Up to 5 μ L

Table 36. Sequencing reaction conditions.

Step	Temperature (°C)	Time
Initial denaturation	94	1 minute
35 cycles	Denaturation	10 seconds
	Annealing	5 seconds
	Extension	2 minutes
Pause/cooling	15	+∞

3.7.2. Plant transformation by floral dip mediated by *A. tumefaciens*

After transforming *A. tumefaciens* cells with the assembled constructs for the knockout of *AGP9*, *AGP24* and *AGP31* (two constructs each gene) in the first transformation attempt and inducing the formation of colonies in solid medium, one colony from each plate was inoculated in a test tube with 2 mL of LB liquid medium supplemented with 50 μ g/mL kanamycin and 10 μ g/mL gentamicin and incubated at 28 °C for 12 to 16 hours with shaking (4 g). After this incubation, 25 μ L of each liquid culture was mixed with the same amount of the liquid culture that contained bacteria transformed with the other construct assembled for the knockout of the same gene. This mixture was then centrifuged for 5 minutes at 2500 g and room temperature, the resulting supernatant was discarded and the cell pellet was resuspended in a 1.5 mL solution with 10% (w/v) sucrose, 1× MES (2-(N-morpholino)ethanesulfonic acid), 1 μ L/mL KOH (pH: 5.7) and 0.05% (v/v) Silwet L-77. Siliques and opened flowers were removed from WT

Arabidopsis plants used to increase transformation efficiency and all floral meristems remaining were covered with 1 to 3 drops of the prepared *Agrobacterium* solution. One WT plant was used for the knockout of each gene.

Relatively to the *Agrobacterium* second transformation attempt, after incubating cultures transformed with the assembled constructs for the knockout of *AGP9*, *AGP21*, *AGP24* and *AGP31* (two constructs each gene) in solid medium for 2 days at 28 °C and performing a colony-PCR (section 3.7.1.7.), one *Agrobacterium* colony carrying the correct construct was transferred to 5 mL LB liquid medium supplemented with gentamicin (20 µg/mL), kanamycin (50 µg/mL) and rifampicin (100 µg/mL). Each liquid culture was incubated for another 2 days at 28 °C. LB liquid medium with the same antibiotic supplementation was added to each liquid culture to a volume of 250 mL. These last cultures were incubated for one day at 28 °C. After this step, each culture was centrifuged for 5 minutes at 11000 g. Finally, after one day of removing the siliques and opened flowers, the floral meristems of 2 WT and 2 *stk* plants were submerged during 5 seconds in a 250 mL solution containing the resuspended *Agrobacterium* cells from the previous centrifugation, 5% (w/v) sucrose and 0.05% (v/v) Silwet-L77. Each one of these solutions contained *Agrobacterium* cells transformed with constructs assembled for the knockout of the same gene.

In both attempts, after the last step, each floral meristem was covered with plastic wrap and the plants were kept in the dark for 12 to 16 hours.

3.7.3. Transformed plants selection and identification of genome edited plants

After a month of transforming *Arabidopsis* plants, the seeds produced by these plants were collected and conserved in microtubes. After selecting the seeds that presented red fluorescence when observed at the Nikon Eclipse Ti-S inverted fluorescence microscope, all the collected seeds were sown in MS medium supplemented with 15 µg/mL hygromycin, as described in section 3.1., being the selected ones sown separately.

After transferring the seedlings to pots, their genomic DNA was extracted as described in section 3.2.. Then, using this DNA and specific primers for the target gene potentially edited in each plant, a PCR was performed in order to amplify a fragment that embraces the sequences homologous to the four sgRNAs specific cassettes designed for that gene (Table 37). The PCR mixtures included DreamTaq RNA Polymerase (Thermo Scientific™) and were prepared as mentioned in section 3.3.4.. The reaction conditions used are present in Table 38. 1× Orange G loading was added to 2 µL of each reaction

product, which were then submitted to an electrophoresis under the conditions described before in section 3.2., with LAB buffer as the electrophoresis buffer.

Table 37. Sequences of the primers used in the amplification of the target gene fragment that includes the sgRNAs specific cassettes. The primers designed are respective to *AGP9*, *AGP21* and *AGP31* genes, since the only seeds that germinated were the ones produced from plants transformed with the constructs designed for the knockout of these genes.

Name	Primer orientation	Sequence (5'-3')	Expected band size
Gen&Seq AGP9 Fwd	Forward	CCAACAACCTCCTCCACC	414 bp
Gen&Seq AGP9 Rev	Reverse	GTTACTAACCTGGTCGTTTC	
Gen&Seq AGP21 Fwd	Forward	GGAGGCAATGAAGATGAAG	160 bp
Gen&Seq AGP21 Rev	Reverse	GTTGTTGCTTTGGCATCT	
Gen&Seq AGP31 Fwd	Forward	TTCATGACACTCGAAGCG	1694 bp
Gen&Seq AGP31 Rev	Reverse	TCAACCGTACAGTAGACG	

Table 38. PCR conditions used for the amplification of the target gene fragment that includes the sgRNAs specific cassettes. The PCR conditions presented are respective to the amplification of fragments from *AGP9*, *AGP21* and *AGP31* genes, since the only seeds that germinated were the ones produced from plants transformed with the constructs designed for the knockout of these genes.

Step	Temperature (°C)		Time	
Initial denaturation	95		2 minutes	
Denaturation	95		30 seconds	
Annealing	52.7	[AGP9]	30 seconds	
	52.9	[AGP21]		
	53.3	[AGP31]		
Extension	72		30 seconds	[AGP9]
			15 seconds	[AGP21]
			1 minute and 45 seconds	[AGP31]
Final extension	72		5 minutes	
Pause/cooling	8		+∞	

The amplified fragments were purified using the GeneJET PCR Purification Kit (Thermo Scientific™) and quantified using a DS-11 FX Spectrophotometer/Fluorometer (DeNovix), according to the manufacturer's instructions. Finally, using the primers in Table 37, 5 µL of each purified fragment were mixed with 5 µL of the respective forward primer (5 µM), doing the same with the reverse primer. These mixtures were sent for sequencing using the ©Eurofins Scientific LightRun Tube service and the resulting sequences were analysed using the ApE software (A plasmid Editor).

3.8. Controlled crosses

Closed flowers at stages 11 and 12 of flower development according to Smyth *et al.* (1990) belonging to *stk* mutant plants were emasculated 24 hours before the pollination with hypodermic needles (0.4 × 20 mm; B. Braun™), a tweezer (Asa Dental, 0645-2) and with the help of a Stemi DV4 stereomicroscope (Zeiss™) and protected with cellophane paper. The cross fertilisation was performed using pollen from dehiscent anthers of

transgenic plants for the pAGP9:NLS:3GFP, pAGP21:GUS, pAGP24:GUS, or pAGP31:GUS constructs and with the help of the same laboratory instruments. After this step, the pistils were protected once again with the cellophane paper for one day and the plants were placed in the growth chamber.

3.9. Phenotype characterisation

3.9.1. Sample preparation for microscopy

The flowers and siliques that were used in microscopy analyses were collected at stages 13, 15 and 17 of flower development according to Smyth *et al.* (1990). The plant material collected was treated accordingly to the respective treatment (sections 3.9.2. to 3.9.5.) and the pistils were isolated and dissected using hypodermic needles (0.4 × 20 mm; B. Braun™), a tweezer (Asa Dental, 0645-2) and with the help of a Nikon C-DSD230 stereomicroscope.

3.9.2. Clear *A. thaliana* flowers and siliques and ovule microscopic observation

Flowers and siliques from WT and *stk* mutant plants collected as described in section 3.9.1. were cleared and examined as described by Brambilla *et al.* (2007) with a few modifications. The biological material was fixed in 10% (v/v) acetic acid in ethanol ON at 4 °C. Then, it was washed consecutively in 90% (v/v) and 70% (v/v) ethanol, 10 minutes each. Finally, the samples were immersed in a solution previously prepared by mixing 160 g of chloral hydrate, 100 mL of distilled water and 50 mL of glycerol. After dissection, the samples' ovules were observed utilising a Zeiss Axio Imager AZ microscope equipped with Differential Interference Contrast (DIC) optics and the images were captured with a ZeissAxioCam MRc3 camera and using the Zen Imaging software.

3.9.3. Aniline blue staining of *A. thaliana* PTs

The aniline blue staining protocol used followed Pereira *et al.* (2016b). Flowers and siliques collected from WT and *stk* mutant plants previously (section 3.9.1.) were fixed in 10% (v/v) acetic acid in ethanol ON at 4 °C. Then, the samples were washed three times, 5 minutes each, and placed in 8 N NaOH at 4 °C ON. Finally, the samples were washed again three times, 20 minutes each, and stained with 0.1% (w/v) DAB (decolourised aniline blue solution) (in 108 M K₃PO₄) at 4 °C ON. The material was dissected using hypodermic needles (0.4 × 20 mm; B. Braun™), a tweezer (Asa Dental, 0645-2) and with the help of a Nikon, C-DSD230. The ovules were observed using a Leica DMLB upright fluorescence microscope using UV fluorescence and the images were captured with a

Nikon DS-Ri2 camera and the NIS-Elements Basic Research software and processed with ImageJ imaging software.

3.9.4. Histochemical detection of GUS activity

After collecting plant material from WT and *stk* mutant plants transgenic for the pAGP21:GUS, pAGP24:GUS and pAGP31:GUS constructs as described in section 3.9.1., each sample was incubated in 90% (v/v) acetone ON at -20 °C. Then, the samples were then washed twice with sodium buffer (0.2 M NaH₂PO₄, 0.2 M Na₂HPO₄) for 10 minutes. During the following steps, the samples were always protected from the light. This washing buffer was then replaced by X-Gluc solution (50 mM Na₂HPO₄, 50 mM NaH₂PO₄, 0.2% (v/v) Triton X-100, 2 mM potassium hexacyanoferrate(II) trihydrate, 2 mM potassium hexacyanoferrate(III), 1 mg/mL X-gluc (5-bromo-4-chloro-3-indolyl β-D-glucuronic cyclohexylammonium salt)) and the samples were incubated at 37 °C ON. Finally, the samples were washed with 90% (v/v) ethanol and then with 70% (v/v) ethanol, 10 minutes each, and submersed in chloral hydrate ON at 4 °C.

The samples were observed using Zeiss™ Axiolmager AZ microscope with Differential Interference Contrast (DIC) optics and photographed with a Zeiss™ Axiocam MRc3 camera through the Zen Imaging acquisition software (SP1, Zen 2011).

3.9.5. Observation of *A. thaliana* reproductive tissues in a fluorescence microscope

As for the WT transgenic plants for the pAGP9:NLS:3GFP construct, after collecting the plant material as described in section 3.9.1., the samples were observed using a Leica DMLB upright fluorescence microscope and the images were captured with a Nikon DS-Ri2 camera and the NIS-Elements Basic Research software and processed with ImageJ imaging software.

3.9.6. Seed set analysis

The seed set of WT plants was compared to the one of *stk* mutant plants. Firstly, to verify if the results obtained when evaluating individual plants were maintained when analysing the groups of plants under study as a whole, ten siliques were randomly collected from the group of WT plants and another ten were also obtained from the group of mutant plants. Each silique was measured and dissected with hypodermic needles (0.4 × 20 mm; B. Braun™), a tweezer (Asa Dental, 0645-2) and with the help of a Stemi DV4 stereomicroscope (Zeiss™) to analyse and count the seeds produced in each silique, being the seeds discriminated in mature green seeds (seeds with a correct

development), white seeds (seeds whose embryos were arrested at the globular or heart stages), aborted seeds (seeds whose embryos did not develop), aborted ovules (unfertilised ovules) and seeds whose embryos were arrested at the linear-cotyledon stage. In order to verify the homogeneity of variances, a Cochran's C test was performed. To evaluate the statistical significance of the differences found between WT and mutant plants, the ANOVA statistic technique was used. Finally, a Tukey's test was carried out to execute the multiple comparison test (<https://www.fc.up.pt/pessoas/amsantos/>). The ten WT siliques collected and the ten *stk* siliques extracted were considered to be the samples of this analysis.

In a second approach, the process was repeated, but using three WT and three *stk* mutant plants (biological replicates), analysing ten siliques (technical replicates) from each one, enabling to draw correct conclusions when comparing the seed set per silique of a WT plant with the one of a *stk* mutant plant.

4. Results

4.1. Genotyping

In order to confirm the genotype of the plants used and check the presence of the T-DNA insertion in the *STK* gene, *stk-2* (Pinyopich *et al.*, 2003) mutant line plants, kindly provided by Lucia Colombo Lab, Università degli Studi di Milano, and *stk* mutant plants transgenic for the pAGP21:GUS, pAGP24:GUS and pAGP31:GUS constructs were genotyped, performing a PCR that used the primers listed in Table 2 (section 3.2.) and genomic DNA extracted from the plants used. This reaction leads to the amplification of a 364 bp fragment, if the T-DNA insertion was not present in the *STK* gene, representative of the WT plants' genotype, or a 438 bp fragment, in case there was a T-DNA insertion in the *STK* gene, which indicates that the respective plant is a *stk* mutant plant. The result of the genotyping of 7 WT and 7 *stk* mutant plants can be observed in Figure 14.

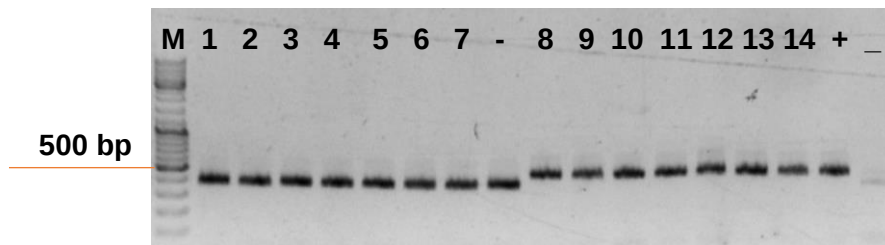


Figure 14. Electrophoresis of the PCR products that correspond to the genotyping of 7 WT plants (1 to 7) and 7 *stk* mutant plants (8 to 14). The PCR performed with genomic DNA from plants 1 to 7 present amplified fragments with a size below 400 bp and near the size of the band corresponding to the negative control for the insertion (-, 364 bp), meaning that these plants were indeed WT plants. As for plants 8 to 14, the amplified fragments present a size above 400 bp and near the size of the positive control for the insertion (+, 438 bp), that resulted from a PCR that used genomic DNA from a previously confirmed *stk* plant, indicating that these plants were indeed *stk* mutant plants. -, negative control for the PCR reaction; M, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

The same was performed for WT and *stk* mutant plants transgenic for the pAGP21:GUS, pAGP24:GUS and pAGP31:GUS constructs to verify the presence of the *GUS* reporter gene in the genomic DNA. In this case, the primers referred in Table 2 lead to the amplification of a fragment with 620 bp in case the *GUS* gene is present in the genome of the plant being tested. Figure 15 demonstrates the result of a PCR that genotyped WT plants transgenic for the pAGP21:GUS construct.

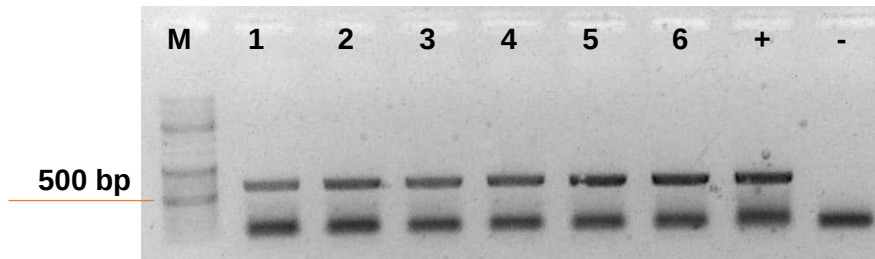


Figure 15. Electrophoresis of the PCR products that correspond to the genotyping of 6 transgenic plants for the pAGP21:GUS construct (1 to 6). The PCR performed with genomic DNA from plants 1 to 6 present amplified fragments with a size above 500 bp and near the size of the positive control (+, 620 bp), that resulted from a PCR that used genomic DNA from a previously confirmed transgenic plant for a construct containing the *GUS* gene. -, negative control reaction; M, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

4.2. Gene expression analysis: from RNA extraction to RT-qPCR

In order to evaluate if the knockout of the *STK* gene interferes in the expression levels of *AGP9*, *AGP21*, *AGP24* and *AGP31*, potential *STK* targets recently identified by ChIP-sequencing analyses (unpublished data), a Real-Time PCR was performed using cDNA from WT and *stk* mutant plants.

Initially, RNA was extracted from 10 mg of seeds of 2 WT (WTI and WTII) and 2 *stk* mutant (*stkl* and *stklI*) plants according to Oñate-Sánchez and Vicente-Carbajosa (2008), with some small changes, since the *STK* gene is expressed in seeds and given that this work was focused in studying ovule and seed development. After checking the quality of the RNA samples (Figure 16) and quantifying their concentration, cDNA was synthesised twice for each RNA sample, always using 1 µg of RNA. In order to validate that these cDNA samples did not contained any contamination from genomic DNA, a PCR was performed using the cDNA samples and specific primers to amplify a fragment of the *ACT8* gene that embraces an intron. In case there is genomic DNA contamination, an additional band (292 bp) is formed, bigger than the only fragment amplified when there is no genomic DNA contamination (185 bp). The *ACT8* gene was used as reference gene, since its expression is ubiquitous and, most probably, does not depend on *STK*'s expression. Figure 17 shows the results from this PCR and it can be observed that only the PCR that used the cDNA sample from the WTI plant originated in the second cDNA synthesis does not present an 185 bp band so intense as the other samples. It can also be noticed that there was no genomic DNA contamination in any of the cDNA samples.

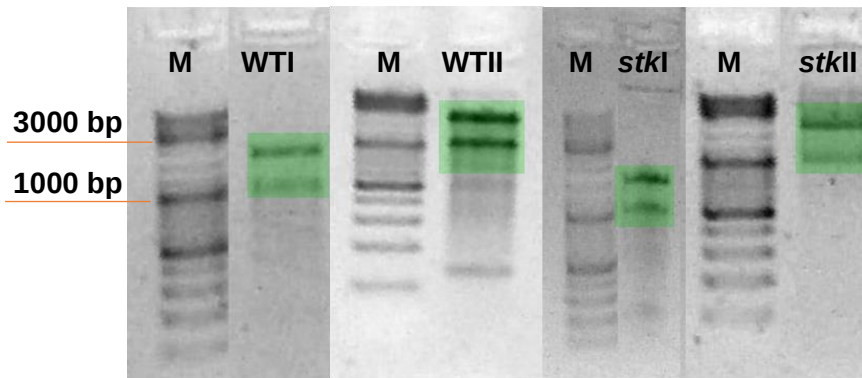


Figure 16. Electrophoresis of the seed RNA samples from 2 WT plants (**WTI** and **WTII**) and 2 *stk* mutant plants (**stkl** and **stklII**). **Green boxes**, 28S (upper bands) and 18S (lower bands) rRNA bands, meaning that the RNA samples present a good quality; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).



Figure 17. Electrophoresis of the semi-quantitative RT-PCR products using cDNA samples from 2 WT (**WI**, **WII**, **wl** and **wll**) plants and 2 *stk* mutant plants (**SI**, **SII**, **sl** and **sll**). The samples **WI**, **WII**, **SI** and **SII** were originated in the first cDNA synthesis and the samples **wl**, **wll**, **sl** and **sll** were created in the second cDNA synthesis. **Green boxes**, fragments amplified (upper fragment) with primers specific of the *ACT8* gene with a size quite similar to the size of the fragment amplified using a positive control cDNA previously obtained from a WT plant (**x**, 185 bp). **+**, positive control that resulted from a PCR that used genomic DNA from a WT plant (292 bp); **-**, negative control reaction; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

The cDNA synthesis was performed twice using the same RNA samples referred above in this section, but the RT-qPCR performed using the cDNA samples originated in the first synthesis did not present coherent results (Figure 18 and 19), which was associated with a possible poor condition of the cDNA samples. Consequently, from these RNA samples, two RT-qPCRs were developed. *ACT8* and *RUB1* were always defined as reference genes and the expression levels were normalised according to these genes, since their expressions are unaffected by differences in samples and experimental treatments and are independently regulated, particularly in *Arabidopsis* female tissues. Here, the expression levels for each gene were compared between *stk* and WT samples, considering each WT sample as the control sample in separate analyses.

Analysing the RT-qPCR performed using the cDNA samples from the first synthesis (Figures 18 and 19), regarding the expression levels of the *AGP21* gene, it can be noticed that, while *stklII* seeds present an expression of this gene significantly lower than both WT seeds, *stkl* seeds show an expression level significantly higher than WTI seeds and practically no difference in the expression of this gene relatively to WTII seeds. As

for the *AGP24* gene, the expression levels quantified in both *stk* plants' seeds are substantially lower than the ones observed in WTI seeds, but considerably higher than the expression levels detected in WTII seeds. Finally, considering the *AGP31* expression, while the *stkII* seeds present an expression level of this gene significantly lower than both WT plants' seeds, *stkI* seeds show an expression significantly lower than WTI seeds, but an expression level significantly higher than WTII seeds. The *AGP9* target gene was included in this RT-qPCR, but none of the melting temperatures reached the threshold of the melt peak graphic, inhibiting to properly analyse the relative normalised expression of this gene.

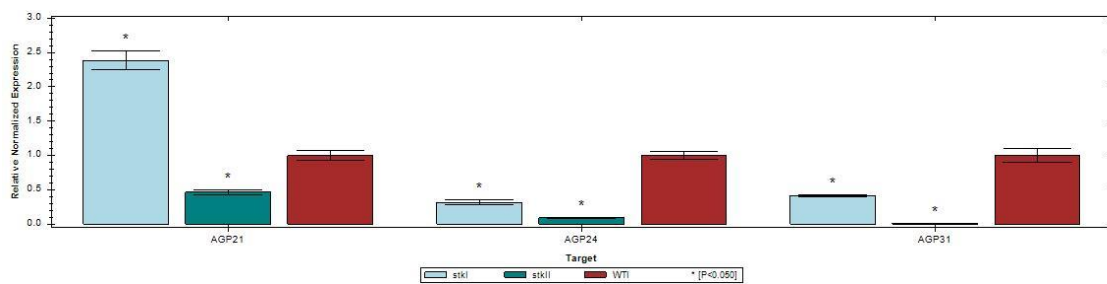


Figure 18. Relative normalised expression of *AGP21*, *AGP24* and *AGP31* in WTI, *stkI* and *stkII* seeds obtained in the first RT-qPCR, defining WTI as the control. *AGP21* expression in *stkI* seeds is significantly higher than in WTI. The other expressions quantified in both *stk* samples are significantly lower than in the control. Error bars, standard error; *, results statistically different from WTI for a $P < 0.05$.

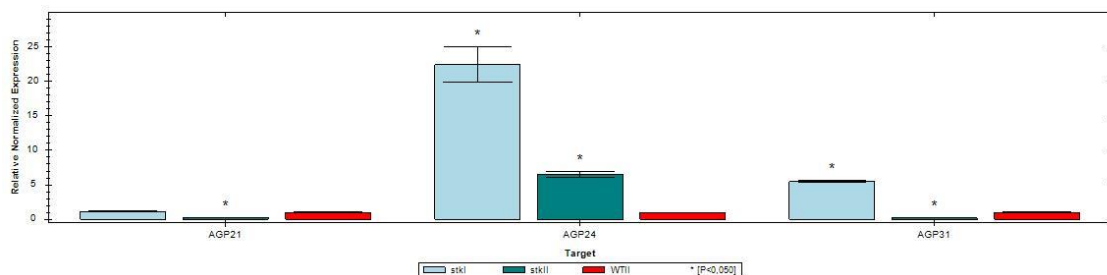


Figure 19. Relative normalised expression of *AGP21*, *AGP24* and *AGP31* in WTII, *stkI* and *stkII* seeds obtained in the first RT-qPCR, defining WTII as the control. In *stkI*, only *AGP24* and *AGP31* expressions are significantly different from WTII, being higher than in the control. In *stkII*, *AGP24* expression is considerably higher than in WTII and the other expressions are significantly lower. Error bars, standard error; *, results statistically different from WTII for a $P < 0.05$.

As for the RT-qPCR performed using the cDNA samples from the second synthesis (Figures 20 and 21), relatively to *AGP21* expression level, while *stkII* seeds present an expression of this gene significantly lower than both WT plants' seeds, *stkI* seeds also show an expression level significantly lower than WTII seeds, but present an expression significantly higher than WTI seeds. In what concerns the *AGP24* gene, similarly to the results obtained in the first RT-qPCR, the expression levels quantified in both *stk* plants' seeds are substantially lower than the ones observed in WTI seeds, but considerably

higher than the expression levels detected in WTII seeds. This scenario was also observed in the *AGP31* gene expression levels. Finally, in this RT-qPCR it was possible to quantify the relative normalised expression of the *AGP9* gene. While the *stkI* seeds present an expression level of this gene significantly lower than both WT plants' seeds, *stkII* seeds also show an *AGP9* expression lower than both WT plants' seeds, but which is only significant when compared to the *AGP9* expression in WTI seeds.

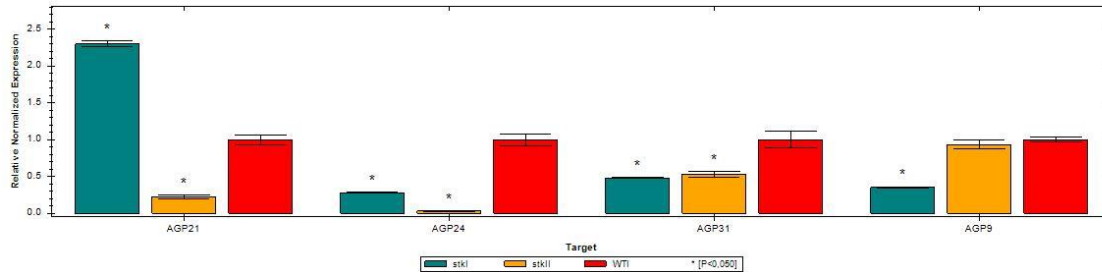


Figure 20. Relative normalised expression of *AGP9*, *AGP21*, *AGP24* and *AGP31* in WTI, *stkI* and *stkII* seeds obtained in the second RT-qPCR, defining WTI as the control. *AGP21* expression in *stkI* is significantly higher than in WTI. *AGP9* expression in *stkII* is not statistically different from WTI. The other expressions quantified in both *stk* samples are significantly lower than in the control. Error bars, standard error; *, results statistically different from WTI for a $P < 0.05$.

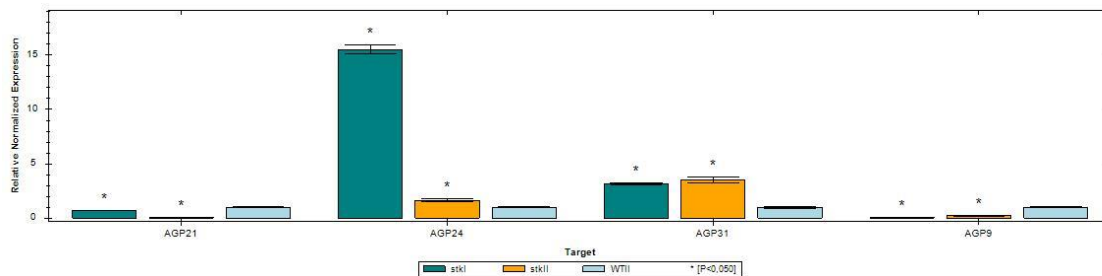


Figure 21. Relative normalised expression of *AGP9*, *AGP21*, *AGP24* and *AGP31* in WTI, *stkI* and *stkII* seeds obtained in the second RT-qPCR, defining WTI as the control. *AGP21* and *AGP9* expressions are significantly lower in both *stk* seeds than in WTI. The other expressions quantified are considerably higher in both mutants than in the control. Error bars, standard error; *, results statistically different from WTI for a $P < 0.05$.

Ultimately, given the persistence of inconsistencies in the second RT-qPCR results, a third attempt was conducted, this time repeating both the RNA extraction and cDNA synthesis, but making some alterations to the method previously performed. The plant material used in this new RNA extraction included not only seeds, but all the pistil tissues, excluding the valves. Furthermore, the amount of plant material used increased to 30 mg. This allowed to obtain higher RNA concentrations for the same final volume, which was needed later for the amplification of the *STK* coding sequence (section 3.5.2.), while still using plant material in which *STK* is expressed. However, the amount of RNA used in the synthesis of cDNA samples for the RT-qPCR remained 1 µg.

In order to avoid incompatible results when comparing gene expression levels between the *stk* samples and more than one WT sample, this third RT-qPCR included only one WT cDNA sample, originated from a new WT RNA sample (WT). This way, only this new WT sample was defined as the control sample. Three *stk* cDNA samples were synthesised from three new *stk* RNA samples (*stkI*, *stkII* and *stkIII*). This time, special care was taken to avoid long temporal distances between RNA extraction and cDNA synthesis, in order to avoid RNA degradation. However, *stkII* and WT cDNA samples used in the third RT-qPCR were synthesised much later relatively to the RNA extraction than the other two *stk* cDNA samples. This happened because, as *stkII* and WT RNA samples, which were the first ones to be extracted, presented high RNA concentrations, much bigger than the other samples, the RNA amount initially used for the *stkII* and WT cDNA syntheses was much higher than 1 µg and what would be possible with the other RNA samples. So, in order to normalise the amount of RNA used in all the cDNA syntheses, *stkII* and WT cDNA syntheses were repeated later, but using only 1 µg of RNA, even considering the possibility that the respective RNA samples could already be slightly degraded, although they were stored at -80 °C, purified with DNase and protected with RiboLock RNase Inhibitor.

As before, both RNA and cDNA samples were tested by electrophoresis and semi-quantitative RT-PCR, respectively. As it can be observed in Figure 22, that shows the electrophoresis of the RNA samples, only *stkI* and *stkIII* RNA samples present a good quality. As for the electrophoresis of the semi-quantitative RT-PCR products (Figure 23), all the cDNA samples newly synthesised present no genomic DNA contamination, being amplified only a 185 bp fragment in each PCR.

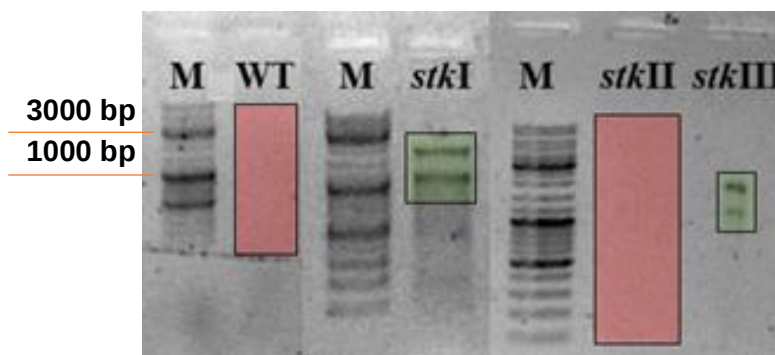


Figure 22. Electrophoresis of the seed RNA samples from 1 WT plant (WT) and 3 *stk* mutant plants (*stkI*, *stkII* and *stkIII*). **Green boxes**, 28S and 18S rRNA bands, meaning that the RNA samples present a good quality; **red boxes**, RNA samples with a less good quality, since there are not any bands; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

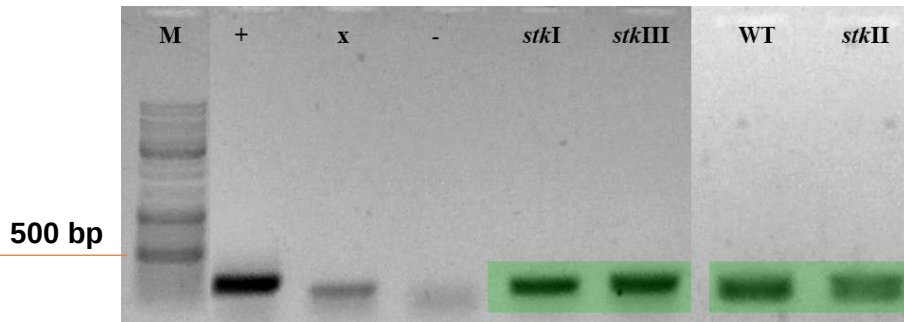


Figure 23. Electrophoresis of the semi-quantitative RT-PCR products using cDNA samples from 1 WT plant (WT) and 3 *stk* mutant plants (*stkI*, *stkII* and *stkIII*). **Green boxes**, fragments amplified (upper fragment in *stkII*) with primers specific of the *ACT8* gene with a size quite similar to the size of the fragment amplified using a positive control cDNA previously obtained from a WT plant (x, 185 bp). +, positive control that resulted from a PCR that used genomic DNA from a WT plant (292 bp); -, negative control reaction; M, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

As for the third RT-qPCR results (Figure 24), the *stkII* sample had to be deleted from this analysis, since the results it presented diverged considerably from the expression levels of the other two *stk* samples. As for these other two samples, the expression levels of the *AGP31* gene are significantly lower in both *stk* samples than in the WT sample, contrary to what happens in the expression levels of the *AGP9* gene that are significantly higher in both *stk* samples than in the WT sample. Although the expression level of the *AGP21* gene was lower in the *stkIII* sample than in the WT sample, the *stkIII* sample did not diverge significantly from the WT sample in what concerns the expression levels of *AGP21* and *AGP24*. The *stkI* sample presents an expression of the *AGP21* gene considerably lower than the expression level of the same gene in the WT sample, opposing to the expression of *AGP24*, which is significantly higher in the *stkI* sample than in the WT sample. *ACT8* and *RUB1* were again defined as reference genes.

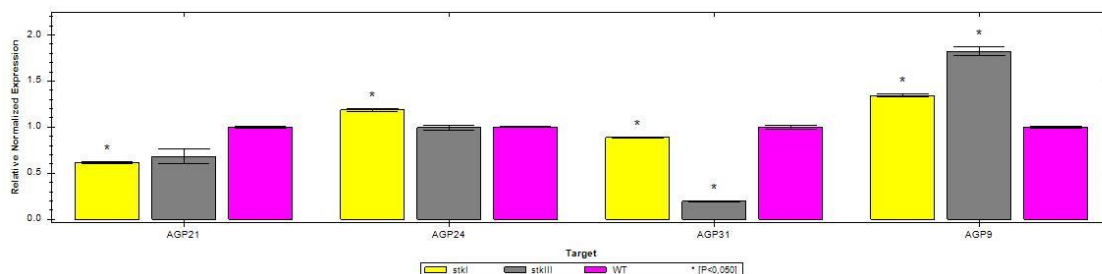


Figure 24. Relative normalised expression of *AGP9*, *AGP21*, *AGP24* and *AGP31* in WT, *stkI* and *stkIII* samples obtained in the third RT-qPCR, defining WT as the control. In *stkI*, *AGP21* expression is considerably lower than in WT and *AGP24* expression is significantly higher. *AGP21* and *AGP24* expressions in *stkIII* do not diverge significantly from the WT. While *AGP31* expression is significantly lower in both *stk* samples than in WT, *AGP9* expression is considerably higher. Error bars, standard error; x, results statistically different from the WT for a $P < 0.05$.

4.3. STK direct interaction with potential target genes

4.3.1. Electrophoretic Mobility Shift Assay

In an attempt to discover if the STK TF binds directly to the putative target genes of this study, an EMSA was projected to check this interaction *in vitro*. This assay verifies if the binding of a TF to sections of the target gene delays the electrophoretic run of these fragments. The sections selected can correspond to the binding site motif of the TF under study. In this case, since STK belongs to the MADSbox family, the last twelve potential CArG boxes upstream of the ATG of *AGP9*, *AGP24* and *AGP31* genes were identified *in silico* (Figure 11 and Table 11, section 3.4.1.). Then, twelve oligonucleotides were designed containing these CArG boxes (Annex I, Supplemental table 1), in order to be used in the EMSA, according to the DIG Gel Shift Kit, 2nd generation (Roche).

As for the cloning and amplification of the *STK* coding sequence, necessary to the expression of the respective protein, firstly, using 2660 ng of the WT RNA sample obtained in the second RNA extraction approach, WT cDNA was synthesised using a specific reverse primer for the *STK* coding sequence. Therefore, only *STK* cDNA was obtained. Then, the *STK* coding sequence was amplified using this cDNA. The electrophoresis of the respective PCR product is present in Figure 25. In addition to non-specific bands, it can be observed a band that has a size quite similar to the one expected (788 bp). The PCR product was then purified in order to insert this fragment in the expression vector. The product of the PCR was purified using a GeneJET PCR Purification Kit (Thermo Scientific™), instead of extracting and purifying the specific band from the agarose electrophoresis gel, to avoid a low final DNA concentration, as it happened in previous attempts.

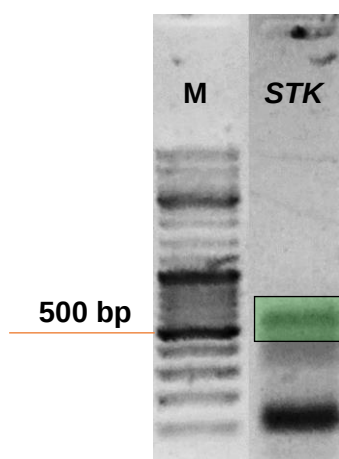


Figure 25. Electrophoresis of the PCR product that amplified the *STK* coding sequence (*STK*). Some non-specific bands are also shown. **Green box**, band with a size close to the predicted (788 bp); **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

After the pET28a plasmid miniprep, this vector and the *STK* coding sequence, amplified in the previous step, were digested with EcoRI and SalI restriction enzymes, being the respective products shown in the electrophoresis present in Figure 26. While the pET28a linearised vector can be detected in the electrophoresis, as it can be seen by the presence of a band with a size close to the expected 5356 bp, the band that would correspond to the digested *STK* coding sequence was not revealed.

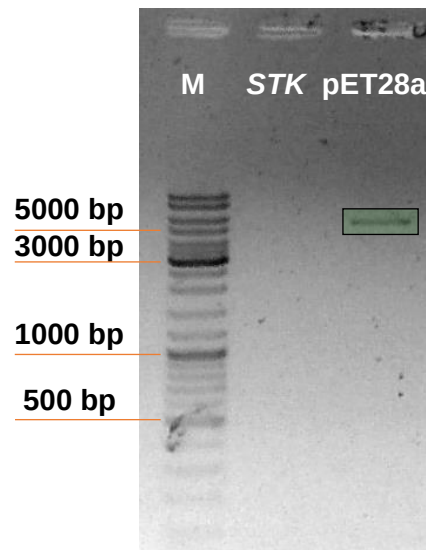


Figure 26. Electrophoresis of the products of the pET28a plasmid and *STK* coding sequence restriction. The lane that corresponds to the restriction of the *STK* coding sequence (**STK**) does not present any band, contrary to the lane that represents the pET28a plasmid restriction (**pET28a**), that shows a band with a size close to the predicted (5356 bp) (**Green box**). **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

Even though there was no band in the electrophoresis of the products of the *STK* coding sequence restriction, the cloning procedure continued, since the absence of this band in the electrophoresis may have resulted from the low concentration of this DNA fragment in the final volume of the restriction product, which does not totally prevent its use in the following steps, despite decreasing their efficiency.

After the dephosphorylation of the linearised pET28a vector, the digested fragment and the linearised vector were purified, in order to remove all the artefacts or molecular machinery used in the previous steps. Again, the products of these purifications were submitted to an electrophoresis. However, no bands were observed (data not shown). Once more, the ligation process was proceeded, given the possibility that the DNA concentrations in both products were just too low to lead to the formation of bands during the electrophoresis.

Subsequently, in the ligation step itself, to facilitate, instead of quantifying the linearised vector and amplified fragment concentrations and combining them according to molar

ratios, both parts were mixed in three different volume proportions (1:1, 1:2 and 1:3). Finally, the products of each ligation reaction were precipitated and the formation of a final recombinant vector was verified by a restriction analysis. The resulting electrophoresis did not show any band: nor the band representative of the amplified fragment, not even the one corresponding to the linearised vector (data not shown).

4.3.2. Protoplast Transactivation Assay

In order to understand if the STK TF does bind directly to the putative target genes of this study, a PTA was planned in order to verify this connection *in planta* using protoplasts and according to the experience performed by Ueda *et al.* (2011).

Firstly, the *STK* genomic sequence needed to be cloned into the MU1110 vector for future protoplast transfection. So, firstly, this sequence was amplified as described in section 3.6.1. The result of the electrophoresis of the PCR product is exhibited in Figure 27. It can be observed that, even though the molecular-weight size marker used was not the appropriate one (the expected size of the amplified fragment (3825 bp) is higher than the size of the biggest fragment of this marker (2000 bp)), the band shown in this electrophoresis is located in an area close to the expected. Therefore, the band was extracted and the DNA fragment was purified.

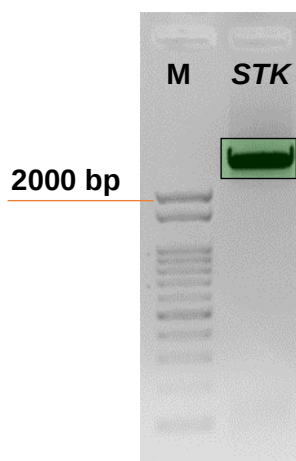


Figure 27. Electrophoresis of the PCR product that amplified the *STK* genomic sequence (*STK*). **Green box**, band that might correspond to the *STK* genomic sequence (located around 4000 bp); **M**, molecular-weight size marker (Gene Ladder 100, Nippon Gene).

The DNA fragment amplified in the previous step and the MU1110 vector were digested using the same restriction enzymes (*Sall* and *NotI*), according to the directional cloning technique. This also enabled to remove the TF gene originally inserted in the plasmid. The electrophoresis of the restriction products are presented in Figure 28. As it can be observed, both restriction reactions succeeded, leading to the formation of two bands

with the expected sizes in the lane corresponding to the restriction of the MU1110 vector and one band in the lane that represents the digestion of the amplified fragment, also with the predicted size. Consequently, it can be assumed that the band shown in the preceding step refers indeed to the *STK* genomic sequence.

The cloning process continued by extracting and purifying the bands that correspond to the linearised MU1110 vector and the digested *STK* genomic sequence. The DNA concentrations of both samples were then quantified in order to calculate the volume to add of each in the ligation reaction following the 2:1 molar ration. Then, the ligation reaction was performed using T4 DNA ligase (Thermo Scientific™). The final product was analysed in a restriction test using the *Sall* restriction enzyme. Unfortunately, the electrophoresis of the product of this restriction reaction did not show any band (data not shown). This can be attributed to the low amounts of DNA obtained after the band extraction and purification processes. Low DNA concentrations, either of the linearised MU1110 vector, either of the *STK* genomic sequence, may have affected the ligation reaction and the amount of recombinant plasmid that might have been formed. Hence, no band was observed in this last electrophoresis.

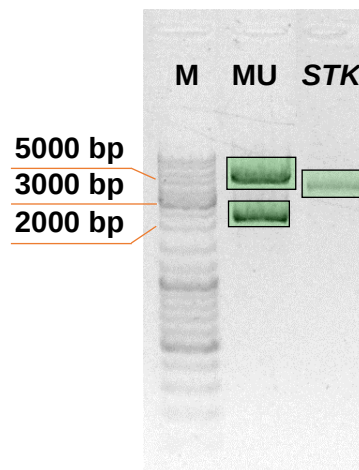


Figure 28. Electrophoresis of the restriction reactions' products that digested the MU1110 vector (**MU**) and the *STK* genomic sequence (**STK**). In the **MU** lane there are two bands, one with a smaller size, close to 2000 bp, and one with a bigger size, approximately 5000 bp, which indicates that these bands correspond to the original TF gene inserted in the plasmid (2074 bp) and to the linearised plasmid (4608 bp), respectively. As for the **STK** lane, it shows one band with a size close to 4000 bp, which suggests that this band represents the digested *STK* genomic sequence (3795 bp). **Green box**, bands with the expected size; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

Along with protoplast transfection using a vector recombined with the TF gene, a PTA also requires the transformation of the same protoplasts with plasmids containing the *cis* elements of the target gene linked to a reporter gene. Thus, these *cis* elements must be identified. The putative *cis* elements of the target gene under study must be selected, *in silico*, and cloned each one in a vector together with a reporter gene, doing the same for

the whole genomic sequence of the target gene. This way, the putative *cis* elements of the *AGP24* (promoter and 5' UTR) and *AGP31* (promoter, 5' UTR and 3' UTR) genes selected in section 3.4.1. were amplified in order to be cloned into the pBluescript II SK (+) vector. The electrophoreses of the PCR products are present in Figure 29 and 30. As it can be observed, concerning the putative *AGP24* gene *cis* elements, only the *AGP24* 5' UTR seems to have been amplified. Although light, a band with a similar size to the one of this fragment (120 bp) can indeed be identified in the respective electrophoresis lane. The negative control in this case, also presents a light band with the same size, which might have resulted from DNA contamination of the reaction mixture. The PCR to amplify the potential *cis* element present in the *AGP24* gene promoter did not succeed, since no band was observed in the respective lane, requiring an improvement of the reaction. As for the amplification reactions of the predicted *AGP31* gene *cis* elements, all of them were successful, since the lanes corresponding to these amplifications revealed bands with approximate sizes of the expected values.

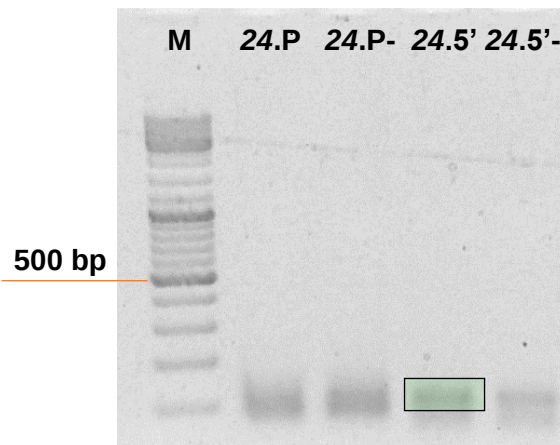


Figure 29. Electrophoresis of the PCR products of the putative *AGP24* gene *cis* elements (promoter and 5' UTR) amplification reactions. As it can be observed, the amplification of the *AGP24* 5' UTR was the only successful reaction, since in this lane (**24.5'**) there is a band with a size above 100 bp and close to the size of the *AGP24* 5' UTR (120 bp). Relatively to the lane corresponding to the negative control for this reaction (**24.5'-**), a tenuous band with the same size can also be observed. With regards to the amplification reaction of the predicted *cis* element present in the *AGP24* gene promoter, no band is observed in the lanes where the products of this reaction (**24.P**) and of the respective negative control (**24.P-**) were electrophoresed. **Green box**, band with the expected size; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

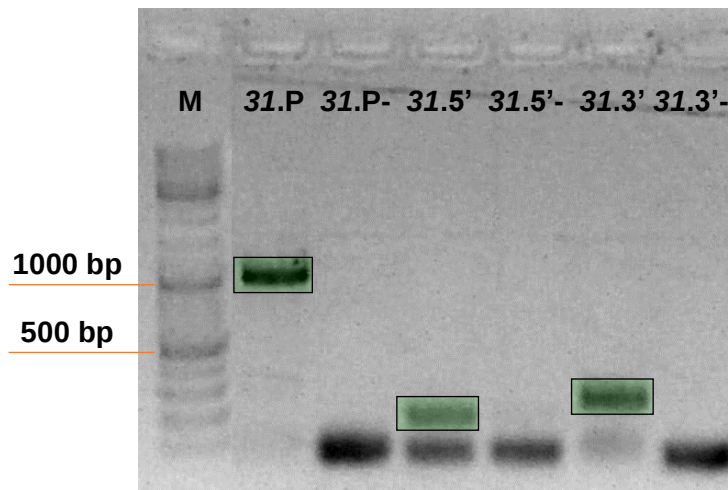


Figure 30. Electrophoresis of the PCR products of the putative *AGP31* gene *cis* elements (promoter, 5' UTR and 3' UTR) amplification reactions. In the second lane, which represents the amplification of the potential *cis* element present in the *AGP31* gene promoter (**31.P**), there is a band with a size above 1000 bp and close to the predicted size of this fragment (1043 bp). The fourth lane, corresponding to the amplification of the *AGP31* 5' UTR (**31.5'**), shows a band with a size close to 200 bp and to the expected size of this fragment (189 bp). Finally, the sixth lane, which is where the products of the *AGP31* 3'UTR amplification were electrophoresed (**31.3'**), a band with a size above 200 bp and close to the presumed size of this sequence (250 bp) is presented. The third (**31.P-**), fifth (**31.5'-**) and seventh (**31.3'-**) lanes correspond to the negative controls of the respective amplification reactions. **Green boxes**, bands with the expected size; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

In what concerns the pBluescript II SK (+) plasmid, after performing a miniprep to obtain this vector from an *E. coli* colony, this plasmid was submitted to a restriction analysis in order to confirm its extraction. Figure 31 shows 3 bands in the lane that represents this restriction reaction. The band with the bigger size might correspond to the non-linearised vector with a similar size to the band present in the lane corresponding to the negative control reaction. The band with an intermediate size might represent the linearised pBluescript II SK (+) plasmid containing the mCherry gene, considering that this band has a size slightly above 3000 bp and close to the size of this recombined vector (3660 bp) and that the *E. coli* cells from which the vector was extracted could also be transformed with pBluescript II SK (+) plasmids containing this gene. Finally, the band with a smaller size can correspond to the linearised non-recombined pBluescript II SK (+) plasmid, because this band has a size just below 3000 bp and close to the size of this plasmid (2961 bp).

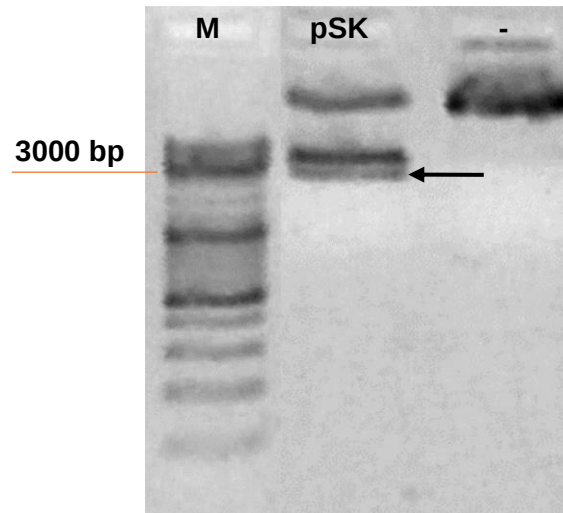


Figure 31. Electrophoresis of the products of the pBluescript II SK (+) plasmid restriction analysis. The arrow points towards the band present in the lane corresponding to the restriction reaction (**pSK**) that has a size close to the expected one of the linearised vector (2961 bp). In the same lane, the band with an intermediate size might represent the linearised pBluescript II SK (+) vector containing the mCherry gene, since the band is similar in size to this recombined vector (3660 bp), and the band with the bigger size might correspond to the non-linearised plasmid, considering it has quite the same size of the band present in the lane that represents the negative control reaction (-). **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

4.4. Creation of knockout mutant plants using the CRISPR-Cas9 technique

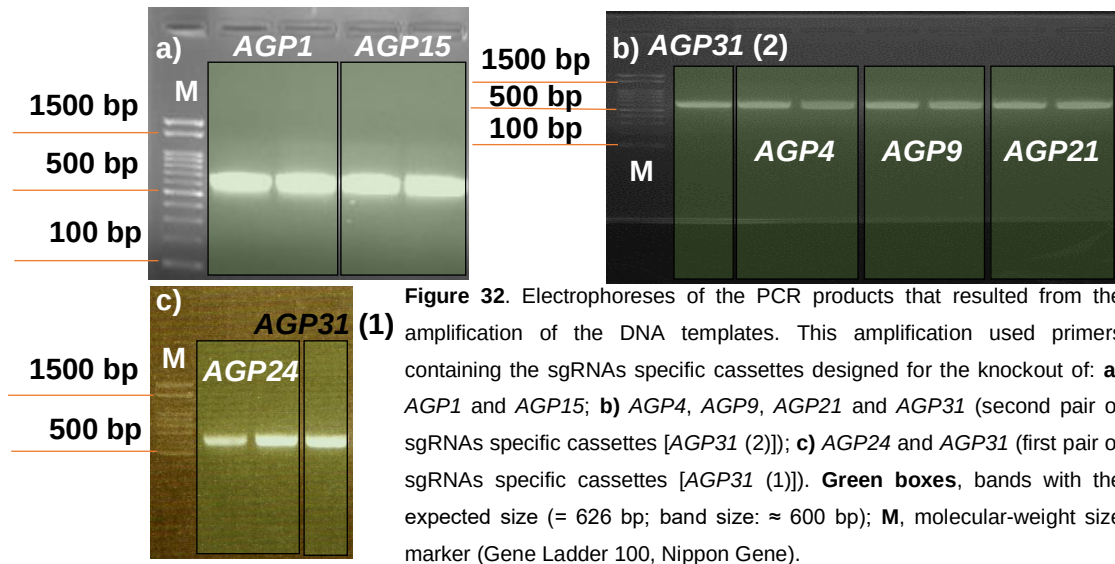
The CRISPR-Cas9 technique was performed in order to create single mutants for *AGP9*, *AGP21*, *AGP24* and *AGP31* and double mutants for each of these gene and for the *STK* gene in order to analyse the respective phenotypes and compare them with the *stk* mutant and try to understand the crosstalk between these genes and its influence in ovule and seed development. Firstly, the constructs used in this technique were assembled, followed by plant transformation by floral dip mediated by *A. tumefaciens*.

4.4.1. Assembly of constructs for plant transformation

4.4.1.1. DNA template amplification

Initially, plasmid constructs were assembled for the CRISPR-Cas9 technique to create knockout mutant plants for all the possible target genes (*AGP1*, *AGP4*, *AGP9*, *AGP15*, *AGP21*, *AGP24* and *AGP31*). Even though only the *AGP9*, *AGP21*, *AGP24* and *AGP31* genes were analysed in this study, constructs were also assembled for the knockout of the *AGP1*, *AGP4* and *AGP15* genes, as well as of the *CKX5* and *CKX7* genes (data not shown) in order to conduct the same experiences with these genes in a near future. To do so, firstly a region of the DNA template pCBC-DT1T2 was amplified in order to include the sgRNAs specific cassettes. The design of each primer and of the DNA template enabled to synthesise PCR products that display the necessary characteristics to be

connected to the backbone vector (pHEE401E+BliI), leading to the incorporation of both sgRNAs specific cassettes into this plasmid between the respective promoter and scaffold. Four sgRNAs specific cassettes were designed, but each template contained only two, so two final plasmids were assembled for each knockout. After the amplification reactions, an electrophoresis was performed to confirm the effective amplification of each template. In Figure 32, it can be observed that all the amplifications performed were successful, since the amplified fragment presents the predicted size.



4.4.1.2. Colony-PCR

After DNA template amplification and purification, each template was inserted in the backbone vector pHEE401E+BliI through Golden Gate assembly. The final assembled vectors were introduced in *E. coli* Mach1 cells. Colony-PCR followed by an electrophoresis was performed to select the colonies transformed with the recombinant plasmid. The recombinant clones present a band with a size of approximately 400 bp, which is more or less the expected band size of 443 bp (Figure 33).

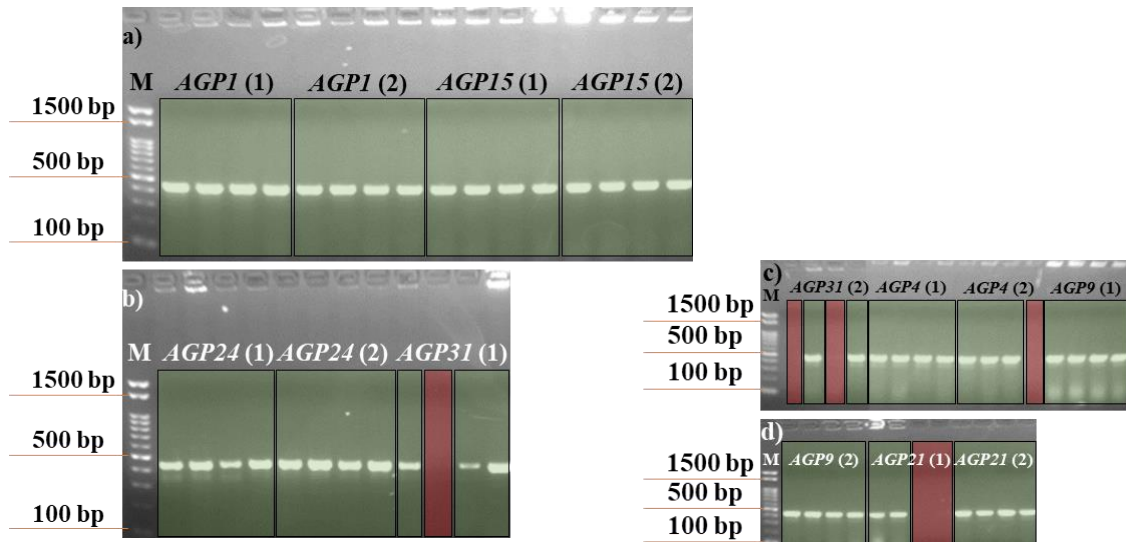


Figure 33. Electrophoreses of the products of the colony-PCR performed to identify the *E. coli* clones transformed with the vectors assembled for the knockout of: **a)** *AGP1* and *AGP15*; **b)** *AGP24* and *AGP31* (1st pair of sgRNAs specific cassettes [*AGP31* (1)]); **c)** *AGP31* (2nd pair of sgRNAs specific cassettes [*AGP31* (2)]); *AGP4* and *AGP9* (1st pair of sgRNAs specific cassettes [*AGP9* (1)]); **d)** *AGP9* (2nd pair of sgRNAs specific cassettes [*AGP9* (2)]) and *AGP21*. **(1)**, 1st pair of sgRNAs specific cassettes; **(2)**, 2nd pair of sgRNAs specific cassettes; **green boxes**, clones transformed with the recombinant plasmid; **red boxes**, clones not transformed with the recombinant plasmid; **M**, molecular-weight size marker (Gene Ladder 100, Nippon Gene). Expected band size: 443 bp; band size: $\approx$ 400 bp.

4.4.1.3. Sequencing

Before *Agrobacterium* and plant transformation and after plasmid DNA extraction from *E. coli* colonies transformed with the recombinant plasmid, the last step in assembling constructs with two sgRNAs specific cassettes is the sequencing of these plasmids and their alignment with the respective predicted sequences, as described in section 3.7.1.9. of Materials and Methods. This step intends to confirm the right insertion of the sgRNAs specific cassettes with no mistakes. The alignment performed proved that the sgRNAs specific cassettes were inserted in the pHEE401E+BllI vector in the right place and orientation (Annex VII, Supplemental figure 6).

4.4.2. *Agrobacterium tumefaciens* and plant transformation

A. tumefaciens cells were transformed with the recombined vectors containing the sgRNAs specific cassettes for the knockout of *AGP9*, *AGP24* and *AGP31* genes in a first attempt, adding the plasmids assembled for the knockout of the *AGP21* gene in a second transformation process, as described in section 3.7.2.. Each *Agrobacterium* cells sample was transformed with one of the plasmids assembled for the knockout of each gene. To confirm the correct transformation of these cells and select the colonies that will be used in the floral dip procedure, a colony-PCR was performed using three of the colonies formed after the second transformation process. The electrophoresis of the products of

this PCR is shown in Figure 34. As it can be observed, all the bands formed in this electrophoresis present a size close to the expected (443 bp), proving that these clones were effectively transformed with the assembled constructs. The lane corresponding to the negative control reaction, also contains a tenuous band, maybe derived from DNA contamination of the reaction mixture.

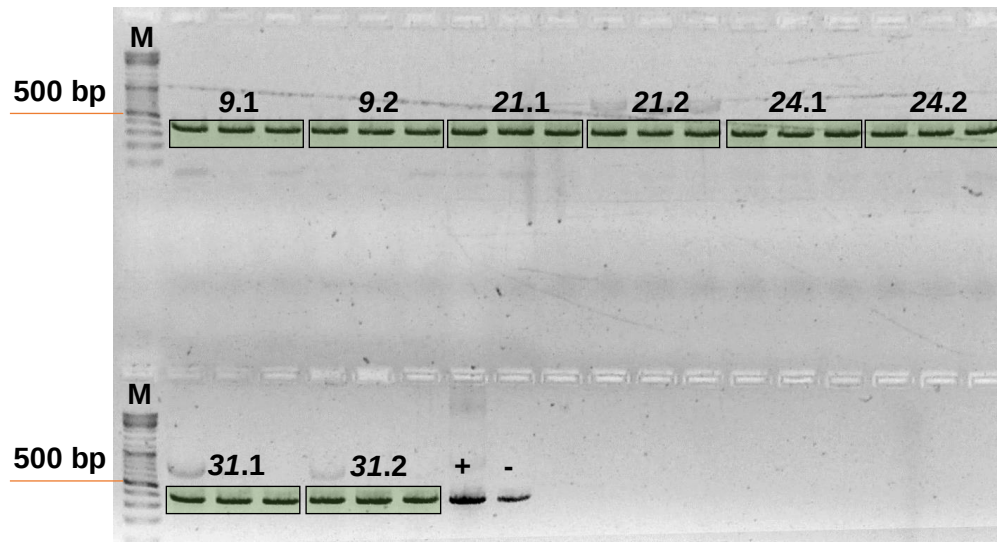


Figure 34. Electrophoresis of the products of the colony-PCR performed to identify the *Agrobacterium* clones transformed with the vectors assembled for the knockout of *AGP9* (**9.1** and **9.2**), *AGP21* (**21.1** and **21.2**), *AGP24* (**24.1** and **24.2**) and *AGP31* (**31.1** and **31.2**). All bands present a size close to 400 bp and to the expected 443 bp. +, positive control reaction (this reaction used one of the plasmids assembled as the DNA template of this reaction); -, negative control reaction; **green boxes**, bands with the expected size; **M**, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

The positive clones were then used in WT and *stk* *Arabidopsis* plants' transformation, also in two different attempts, as described in section 3.7.2.. All seeds collected after the transformation processes were sown, as described in section 3.7.3., since there could be transformed seeds that did not present red fluorescence when observed in a fluorescence microscope, because the mCherry gene was not expressed, or because its expression was not high enough to be detected.

In the first transformation attempt, only one seed collected after the transformation with *AGP31* gene knockout constructs germinated. This seed had presented red fluorescence when observed in the inverted microscope.

As for the second transformation attempt, six plants germinated from the transformation with *AGP9* gene knockout constructs. One of these plants germinated from the only seed that had not presented red fluorescence from all the seeds that germinated. As for the *AGP21* gene knockout, only two germinated, both presenting red fluorescence. Lastly, in a second transformation for the knockout of the *AGP31* gene, one seed germinated,

also with red fluorescence. None of the seeds collected after the transformation for the knockout of the *AGP24* gene germinated, although some of them had presented red fluorescence.

Only the WT plants used in the transformation procedures produced seeds that germinated.

The germinated plants' DNA were analysed to check whether the genome-editing machinery had indeed operated on the respective target gene. To do so, the extracted genomic DNA was used to amplify fragments with homologous sequences to the four sgRNAs specific cassettes designed for the gene potentially edited in each plant (Figure 35), the products of these amplifications were sequenced and the sequences obtained were aligned with the published sequences of the same fragments, in order to identify which plants had indeed the respective genome edited by the CRISPR-Cas9 technique. The results of these alignments are present in Supplemental Figure 7 (Annex VIII). Apparently, the germinated plants were not mutant, since no major differences were detected between the obtained and the original published sequences, mainly in the segments next to the sgRNAs specific cassettes, unlike what was expected.

Only the amplified fragments that comprise the sequences homologous to the four sgRNAs specific cassettes designed for the *AGP21* gene were not properly sequenced (data not shown), which would require improvement of the previous steps in order to conclude whether the respective plants were mutants or not.

As for the plant that germinated from the single seed that had not presented red fluorescence and that had been produced following the transformation performed for the knockout of the *AGP9* gene, the fragment that would be amplified with the DNA of this plant was not even sent for sequencing, which may be explained by the bad condition of its genomic DNA.

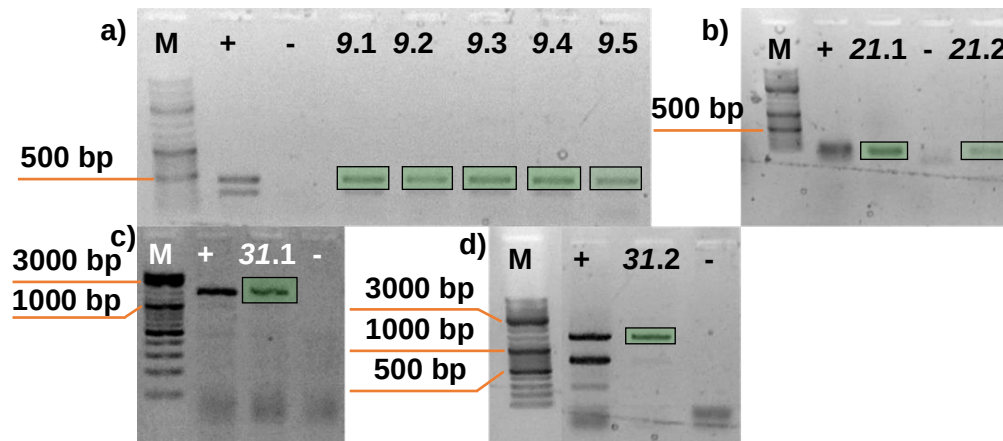


Figure 35. Electrophoreses of the products of the amplification reactions that used genomic DNA from the germinated plants in order to amplify fragments that comprise the sequences homologous to the four sgRNAs specific cassettes designed for the *AGP9* (a) 9.1 to 9.5), *AGP21* (b) 21.1 and 21.2) and *AGP31* (c) 31.1 and d) 31.2) genes. All the fragments amplified present sizes close to the ones of the fragments amplified in the positive control reactions and to the expected sizes (*AGP9*: 414 bp; *AGP21*: 160 bp; *AGP31*: 1694 bp). **Green boxes**, bands formed in the lanes corresponding to the amplification reactions and with a size close to the expected; +, positive control reactions that resulted from PCR that used genomic DNA from a WT plant; -, negative control reactions; M, molecular-weight size marker (GeneRuler DNA Ladder Mix, Thermo Scientific™).

4.5. Phenotype characterisation

4.5.1. WT and *stk* phenotypes

4.5.1.1. *Arabidopsis* flowers and siliques clearing

To address some of the functions of the target genes and their interactions, the phenotypes of siliques, ovules and seeds of the mutant plants created were characterised and compared between them and with *stk* and WT plants. Firstly, ovules and seeds collected from WT and *stk* siliques were cleared and observed with DIC microscopy, so that *stk* plants' phenotypes could be confirmed and then compared to phenotypes of other mutants. With this technique it was possible to observe that ovules collected from *stk* flower buds (Figure 36. b)) are apparently normal and like the ovules extracted from WT flowers at the same stage (Figure 36. a)), exhibiting a nucellus and integuments in the beginning of their development, specific features of stage 2-IV of ovule development. At a further stage, it could be noticed that either WT (Figure 36. c)) and *stk* (Figure 36. d) – e)) newly opened flowers present ovules containing an embryo sac with at least a central cell, two synergids and an egg cell, typical constituents found in an ovule at stage 3-VI of development. Furthermore, it was also perceived that some fertilised ovules of *stk* siliques extracted at stage 15 of flower development present four nuclei which most probably resulted from the second nuclear division in the endosperm taking place at stage 4-IV of ovule development (Figure 36. f)). The flower developmental

stages mentioned are according to Smyth *et al.* (1990) and the ovule developmental stages are according to Schneitz *et al.* (1995).

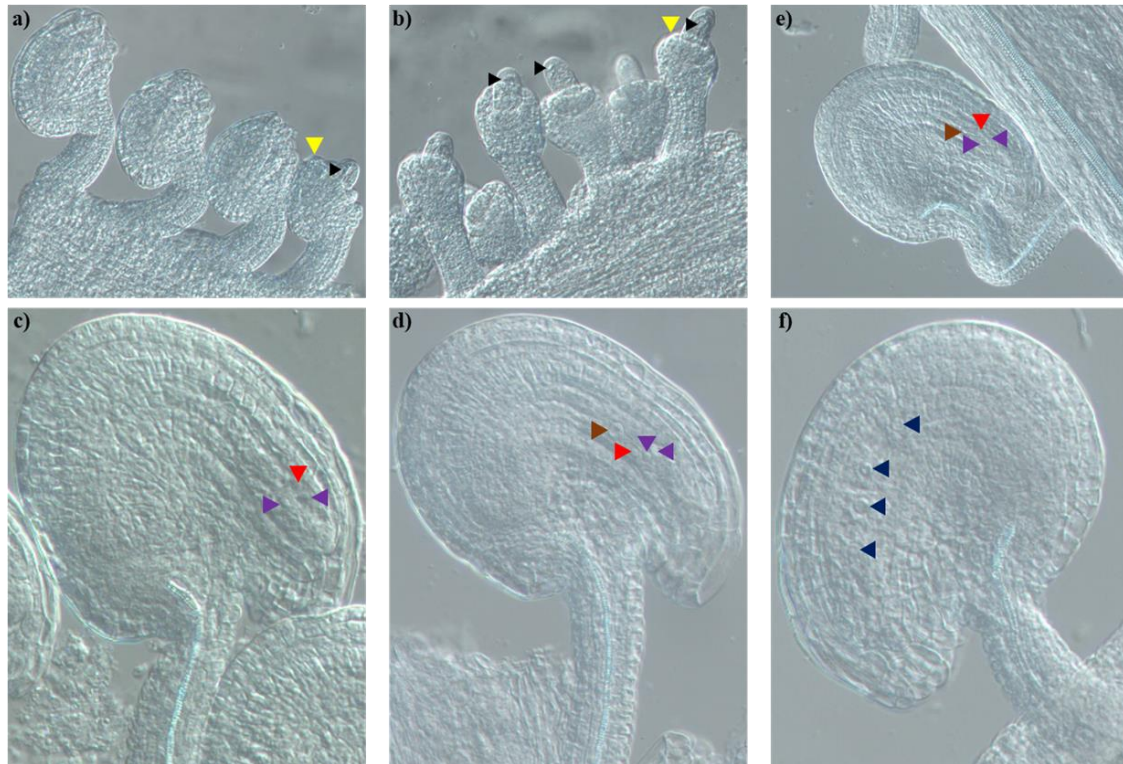


Figure 36. Clearing with chloral hydrate of WT (**a**) and **c**) and *stk* (**b**), **d**) – **f**) ovules. **a**) and **b**) ovules at stage 2-IV of ovule development (during megasporogenesis) extracted from flower buds, presenting a nucellus (black triangles) and integuments in the beginning of their development (yellow triangles); **c**) to **e**) ovules at stage 3-VI of ovule development (during megagametogenesis) extracted from newly opened flowers, presenting two synergids (purple triangles), an egg cell (red triangles) and a central cell (brown triangles); **f**) ovule at stage 4-IV (in the post-fertilisation development) extracted from a silique collected at stage 15 of flower development, presenting four nuclei (blue triangles) that may have resulted from the second nuclear division in the endosperm. The flower developmental stages are the ones described by Smyth *et al.* (1990) and the ovule developmental stages were reported by Schneitz *et al.* (1995).

However, *stk* seeds (Figure 37. **c**) – **d**)) seem rounder, smaller and present a bigger funiculus when compared to WT seeds (Figure 37. **a**) – **b**)). Finally, seeds of *stk* siliques collected at stage 17 of flower development do not show a constriction in the abscission zone unlike what happens with WT seeds at the same stage (Figure 37. **c**)).

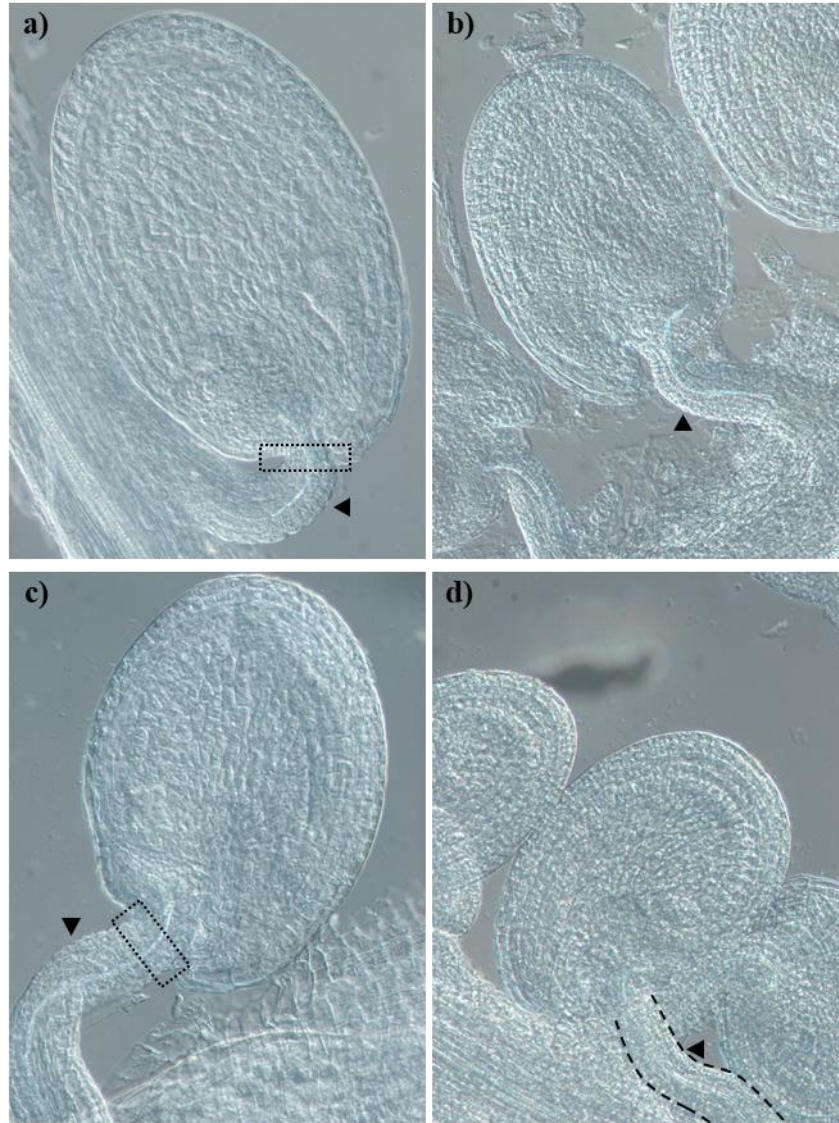


Figure 37. Clearing with chloral hydrate of seeds of WT (a) and b)) and *stk* (c) and d)) siliques extracted at stage 15 of flower development (b) and d)) and stage 17 of flower development (a) and c)). It can be observed that, apparently, *stk* seeds present a smaller size, a rounder shape and a bigger funiculus than WT seeds. Furthermore, seemingly, seeds that belong to *stk* siliques collected around stage 17 of flower development do not present a well differentiated abscission zone (dotted lines). Black arrows and dashed lines, funiculus. The flower developmental stages are according to Smyth *et al.* (1990).

4.5.1.2. *In vivo* PT growth assay

Another phenotype analysed was the PT growth through the pistil of WT (Figure 38. a) – e)) and *stk* (Figure 38. f) – h)) plants, under self-pollination. This process was observed *in vivo* using aniline blue staining. In both cases, the pollen grains observed germinated normally in the stigma (Figure 38. b) and g)) and the PTs grew properly through the transmitting tissue until reaching the ovules (Figure 38. b), c), e) and g)), being each ovule penetrated by only one PT (Figure 38. b) – e), g) – h)). The PT growth can be observed since the callose present at the PT wall and plugs is stained by the DAB

solution used in this technique. Moreover, the WT and *stk* ovules observed at an initial developmental stage also presented labelling in a punctate pattern (Figure 38. **a**) and **f**).

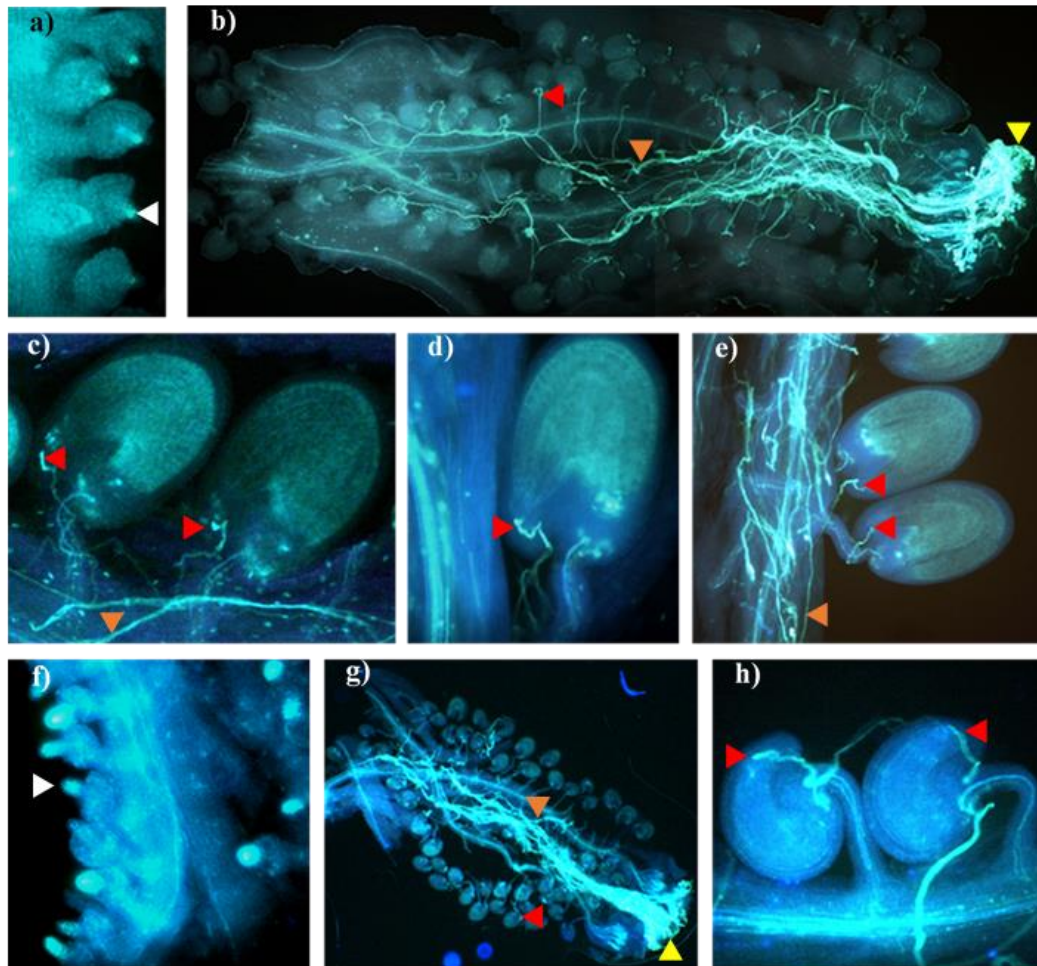


Figure 38. Aniline blue staining of PTs (PTs) in WT (**a** – **e**) and *stk* (**f** – **h**) pistils. **a**) and **f**) ovules extracted from flower buds, presenting a small punctate pattern in the nucellus (white triangles); **b**) and **g**) pistils collected from newly opened flowers, presenting pollen grains germinating in the stigma (yellow triangles), PTs growing normally through the transmitting tissue (orange triangles) and PTs penetrating the ovules (red triangles); **c**) to **e**) seeds collected from siliques at stage 17 of flower development according to Smyth *et al.* (1990), presenting normal PT growth through the transmitting tissue and PTs penetrating the ovules; **h**) ovules extracted from newly opened flowers, presenting PTs normally penetrating them.

4.5.1.3. Seed set analysis

WT and *stk* plants were also compared in what concerns their siliques' seed set. In a first approach, studying entire groups of plants as units, ten siliques were randomly collected from a group of WT plants and ten siliques were also randomly obtained from a group of *stk* plants. Then, these siliques' phenotypes were characterised, measuring the length of each silique, counting their seeds and considering five other factors according to the appearance of the ovules or seeds, which was related to the stage of embryo development, or the condition and integrity of each. The average results are displayed

in the graphics of Figure 39. The total number of seeds and the number of white and green seeds are the only results that present homogeneity of variances, one of the assumptions of the ANOVA, and, from these, only the total number of seeds produced per silique is significantly different between the group of *stk* plants and the group of WT plants, enabling to conclude that, across the group of *stk* plants analysed, the total seed production per silique is statistically lower than each silique's seed production in the group of WT plants.

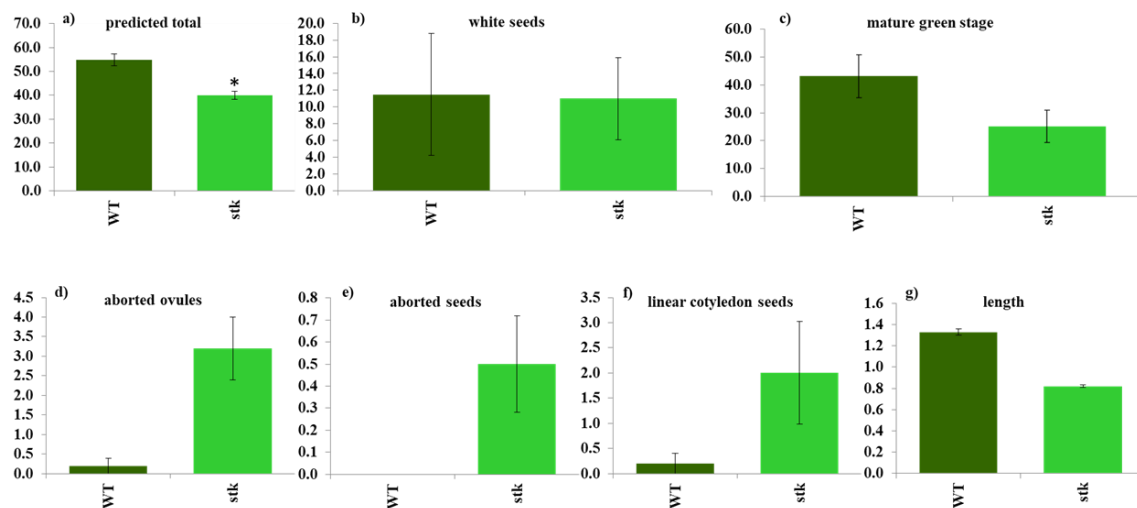


Figure 39. Graphics showing the results of seed set and silique length when using ten random siliques from a group of WT plants and another ten random siliques from a group of *stk* plants as the samples being studied. These graphics display, per silique, the average **a)** total number of seeds produced; **b)** number of white seeds found; **c)** number of mature green seeds produced; **d)** number of aborted ovules found; **e)** number of aborted seeds detected; **f)** number of seeds whose embryos were arrested at the linear-cotyledon stage; **g)** length. Graphics **a)**, **b)** and **c)** are the only ones that present data with homogeneous variances and, from these three graphics, only graphic **a)** exhibits significantly different results. Cochran's C test, ANOVA statistic technique and Tukey's test were performed, being $\alpha=0.05$. Error bars, standard error; *, results statistically different from the results obtained with the WT siliques for the same factor (total number of seeds produced per silique in this case).

In a second approach, the seed set analysis considered three biological replicates (using 3 WT and 3 *stk* mutant plants) and ten technical replicates (using 10 siliques per plant) in order to properly compare the seed set of a WT plant with the one of a *stk* mutant plant. The analysis was performed according to the previous approach and the results obtained are shown in Figure 40. The total number of seeds, the number of green seeds and of seeds whose embryos were arrested at the linear-cotyledon stage and the siliques' length are the only results that present homogeneity of variances. Relatively to the number of seeds that presented embryos arrested at the linear-cotyledon stage, the differences observed between WT and *stk* mutant plants are not statistically significant. As for the total number of seeds, despite the significant differences observed, nothing can be concluded, since the differences detected and their statistical significance vary

according to the pair of plants being analysed. Nevertheless, coherent results can be noticed when comparing WT and *stk* mutant plants regarding the siliques' length and the number of green seeds produced, enabling to affirm that a WT plant produces, on average, more green seeds per silique and longer siliques than a *stk* mutant plant.

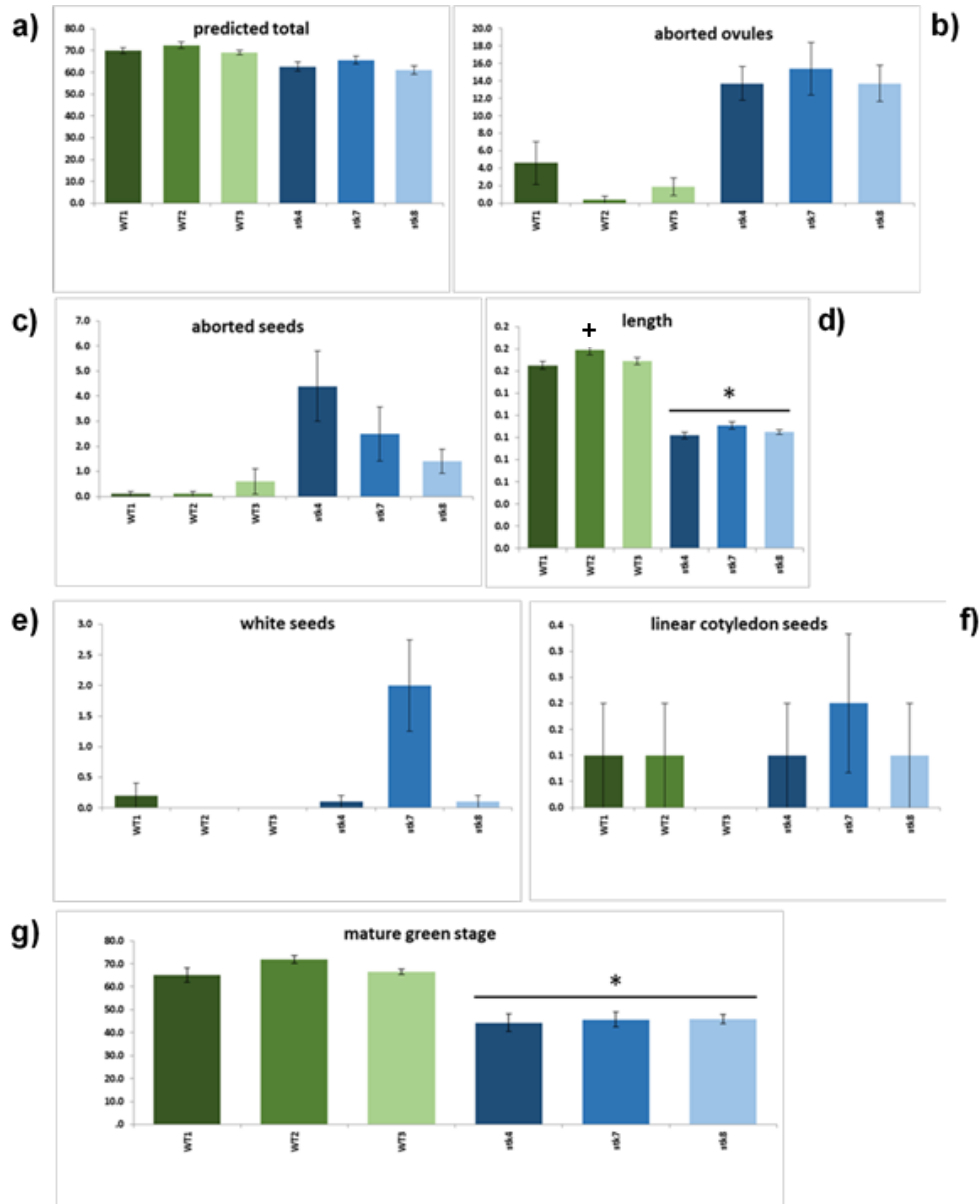


Figure 40. Graphics showing the results of seed set and silique length when using three WT plants and three *stk* mutant plants, analysing ten random siliques from each plant. These graphics display, per silique, the average **a)** total number of seeds produced; **b)** number of aborted ovules found; **c)** number of aborted seeds detected; **d)** length; **e)** number of white seeds found; **f)** number of seeds whose embryos were arrested at the linear-cotyledon stage; **g)** number of mature green seeds produced. Graphics **a)**, **d)**, **f)** and **g)** present data with homogeneous variances and, from these three graphics, only graphic **a)**, **d)** and **g)** exhibit significantly different results. Cochran's C test, ANOVA statistic technique and Tukey's test were performed, being $\alpha=0.05$. Error bars, standard error; *, results statistically different from the ones obtained with the WT plants for the same factor (number of green seeds produced per silique and silique length in this case); +, results significantly different from the ones obtained in the other WT plants for the same factor (siliques' length in this case).

4.5.2. Putative AGPs target genes expression pattern

In order to understand the influence that the knockout of the *stk* gene had in the expression pattern of the putative target genes under study, the expression pattern of each gene was analysed. With regards to the *AGP9* gene, flowers and siliques of plants transgenic for the *pAGP9:NLS:3GFP* construct, collected at stages 13 (Figure 41. **a**) – **c**), 15 (Figure 41. **d**) – **f**) and 17 (Figure 41. **g**) – **i**) of flower development, were dissected and observed in a fluorescence microscope. Figure 41 shows some of the expression patterns observed. At all the stages analysed, GFP signal was detected in the vascular tissues of the pistil transmitting tract and funiculus, as well as in the chalazal pole of ovules and seeds. As it is characteristic of transgenic lines for constructs containing the *NLS* gene, the expression patterns exhibits small dots, representing the nuclei.

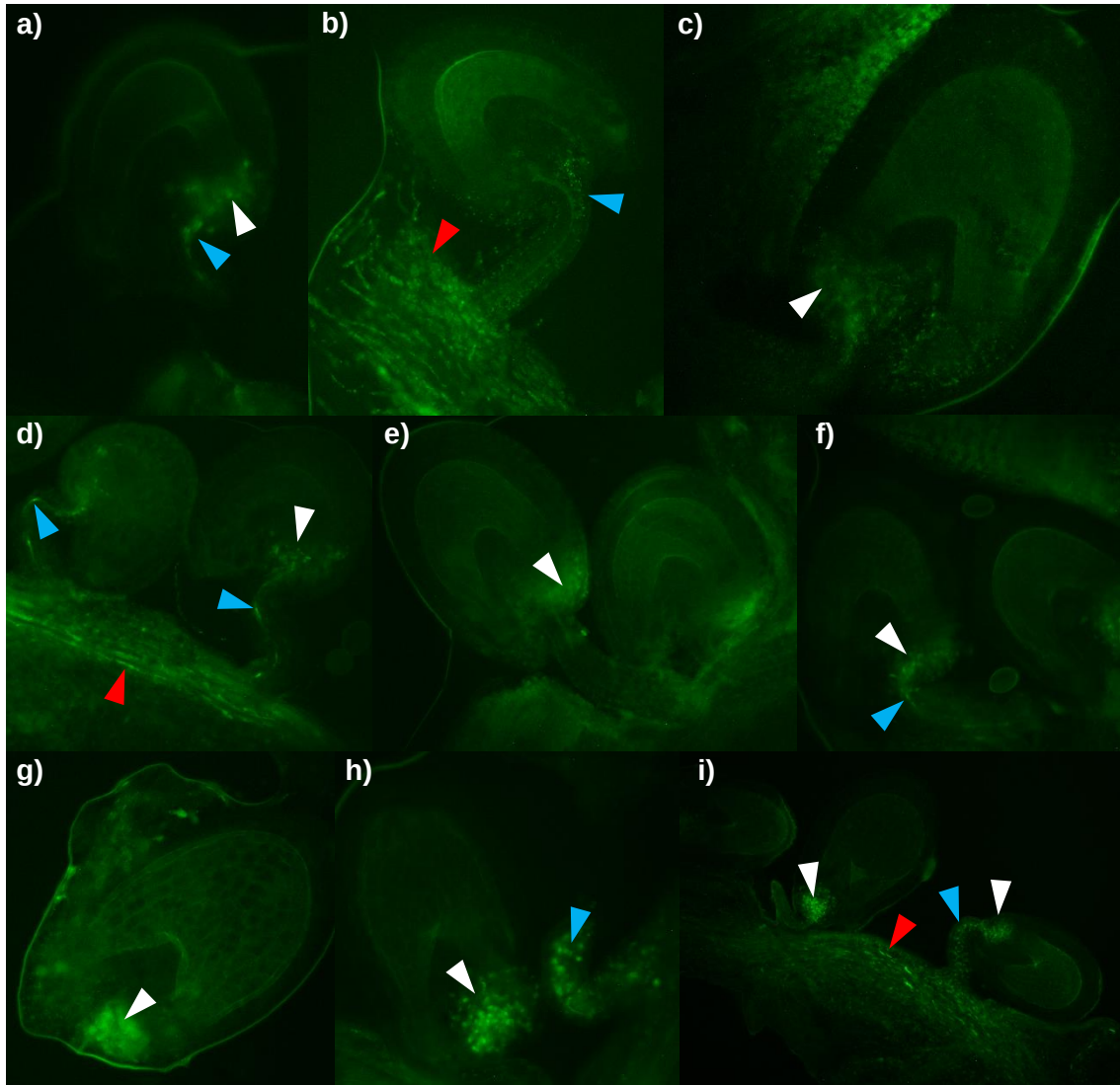


Figure 41. Fluorescence microscopy of ovules and seeds of plants transgenic for the pAGP9:NLS:3GFP construct. The flowers and siliques used were collected at stages 13 (a) to c)), 15 (d) to f)) and 17 (g) to i)) of flower development according to Smyth *et al.* (1990). GFP signal can be observed in the chalazal pole of ovules and seeds (white triangles) and vascular tissues of the pistil transmitting tract (red triangles) and funiculus (blue triangles).

The histochemical detection of the *GUS* gene expression was performed for the transgenic plants of pAGP21:*GUS*, pAGP24:*GUS* and pAGP31:*GUS* constructs. In what concerns the expression pattern of the *AGP21* gene, *GUS* gene expression guided by the *AGP21* gene promoter was observed in the micropylar pole of ovules and seeds and in the septum of pistils collected at stages 13 (Figure 42. a) – c)) and 15 (Figure 42. d) – f)) flower development. At stage 17 (Figure 42. g) – i)) the expression pattern tends to be more diffuse and spreads across the seed endosperm, maintaining expression in the septum.

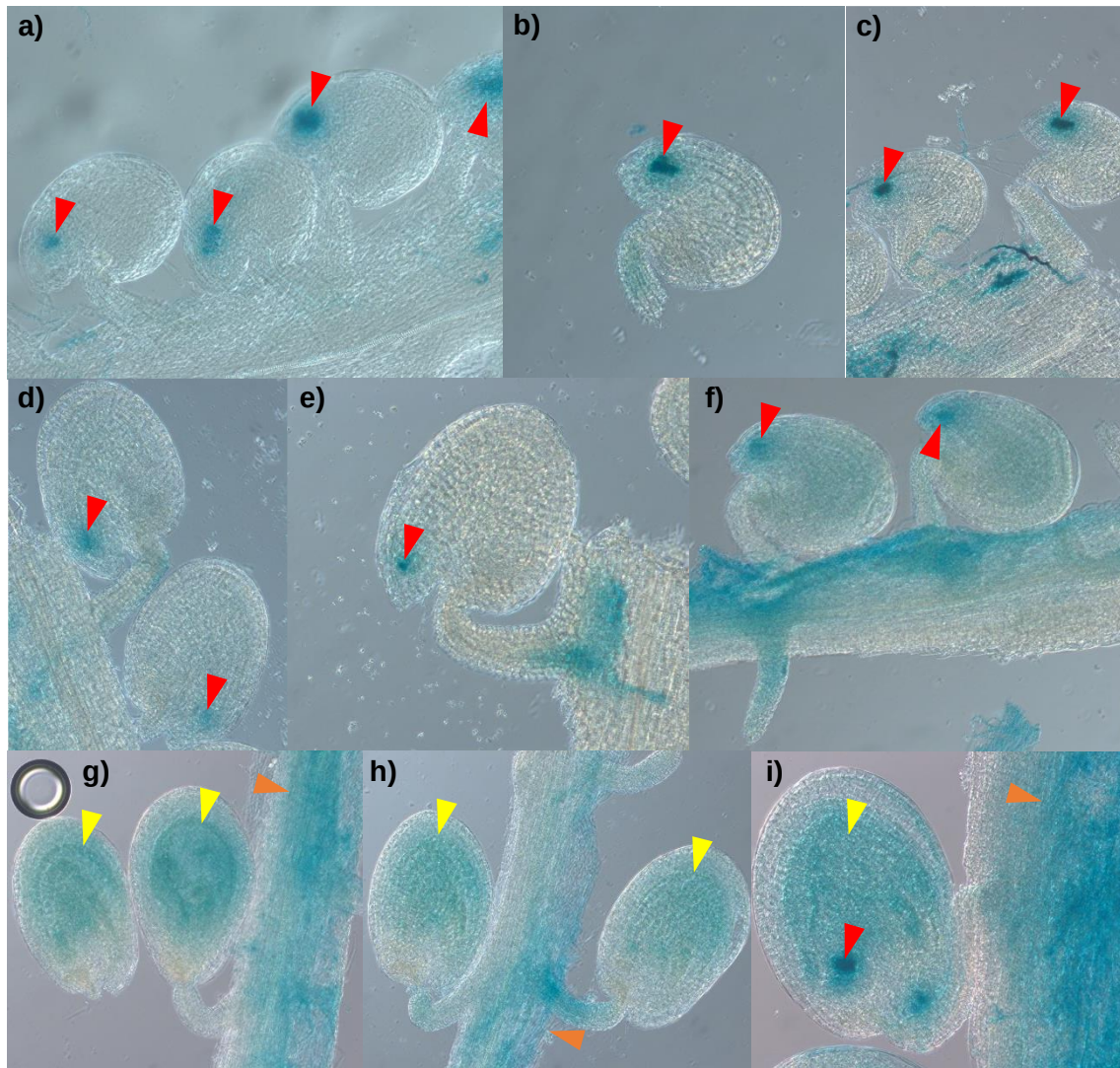


Figure 42. Histochemical detection of GUS activity regulated by the *AGP21* gene promoter in ovules and seeds. In flowers and pistils collected at stages 13 (a) to c) and 15 (d) to f) of flower development, according to Smyth *et al.* (1990), GUS activity was detected in the micropylar pole (red triangles) and in the pistil septum (orange triangles) of ovules and seeds. At stage 17 (g) to i), GUS activity was still observed in the septum, but the expression in the seeds became more diffuse and spread across the endosperm (yellow triangles).

In regards to the expression pattern of the *AGP24* gene, *GUS* gene expression guided by the *AGP24* gene promoter was detected in the micropylar and chalazal poles of ovules and seeds present in pistils collected at stages 13 (Figure 43. a) – c)) and 15 (Figure 43. d) – f)) flower development. However, at stage 17 (Figure 43. g) – i)), GUS activity changes and a strong expression is detected in the seeds' chalazal pole and across the endosperm.

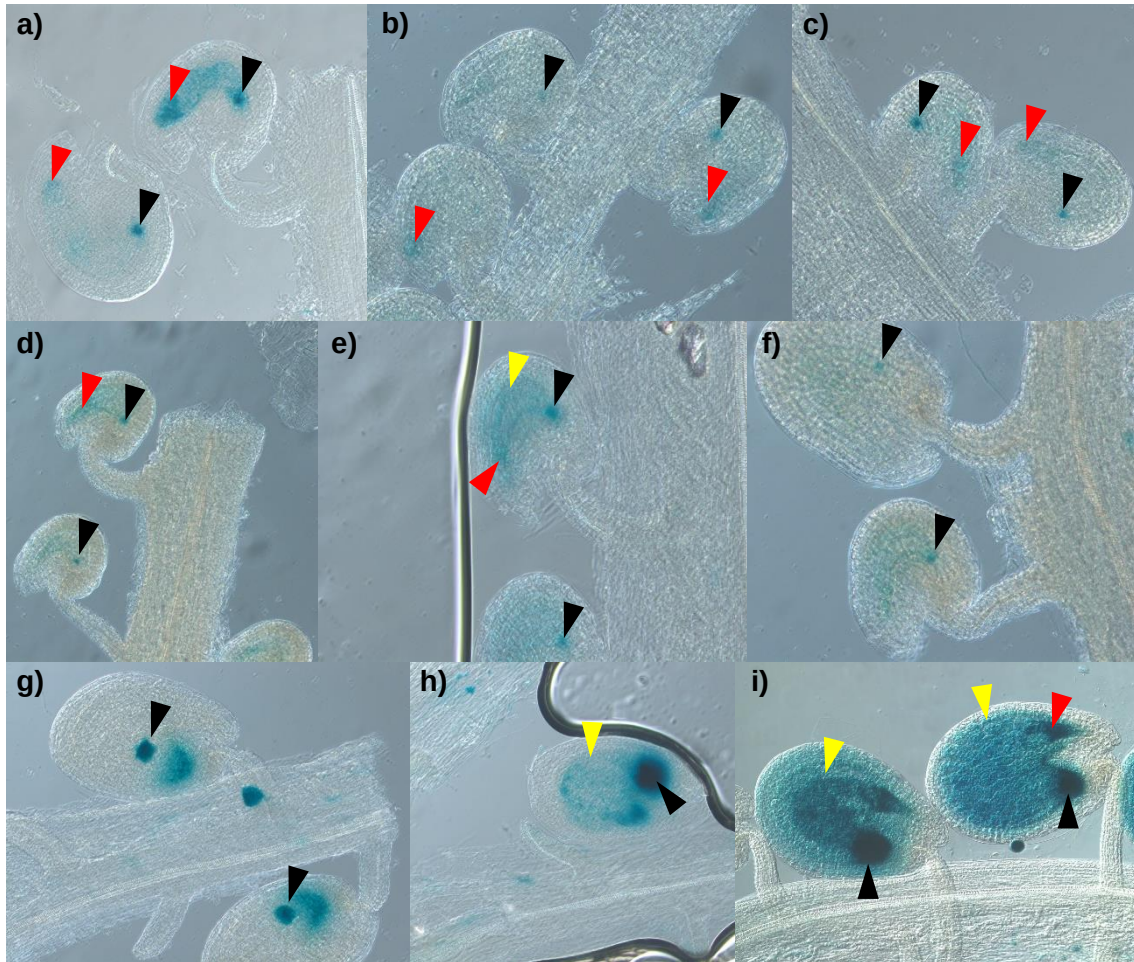


Figure 43. Histochemical detection of GUS activity regulated by the *AGP24* gene promoter in ovules and seeds. In flowers and pistils collected at stages 13 (**a**) to **c**) and 15 (**d**) to **f**) of flower development according to Smyth *et al.* (1990), GUS activity was detected in the micropylar (red triangles) and chalazal pole (black triangles) of ovules and seeds. At stage 17 (**g**) to **i**), GUS activity was observed in the seeds' chalazal pole and across the endosperm (yellow triangles) with a strong expression.

Finally, relatively to the expression pattern of the *AGP31* gene, *GUS* gene expression conducted by the *AGP31* gene promoter was noticed in the ovules' chalazal pole and in the transmitting tissue of pistils collected at stage 13 (Figure 44. **a**) – **c**) of flower development. However, at stages 15 (Figure 44. **d**) – **f**) and 17 (Figure 44. **g**) – **i**), there is almost no *GUS* activity in the seeds' chalazal pole, but the expression detected in the transmitting tissue only disappears at stage 17.

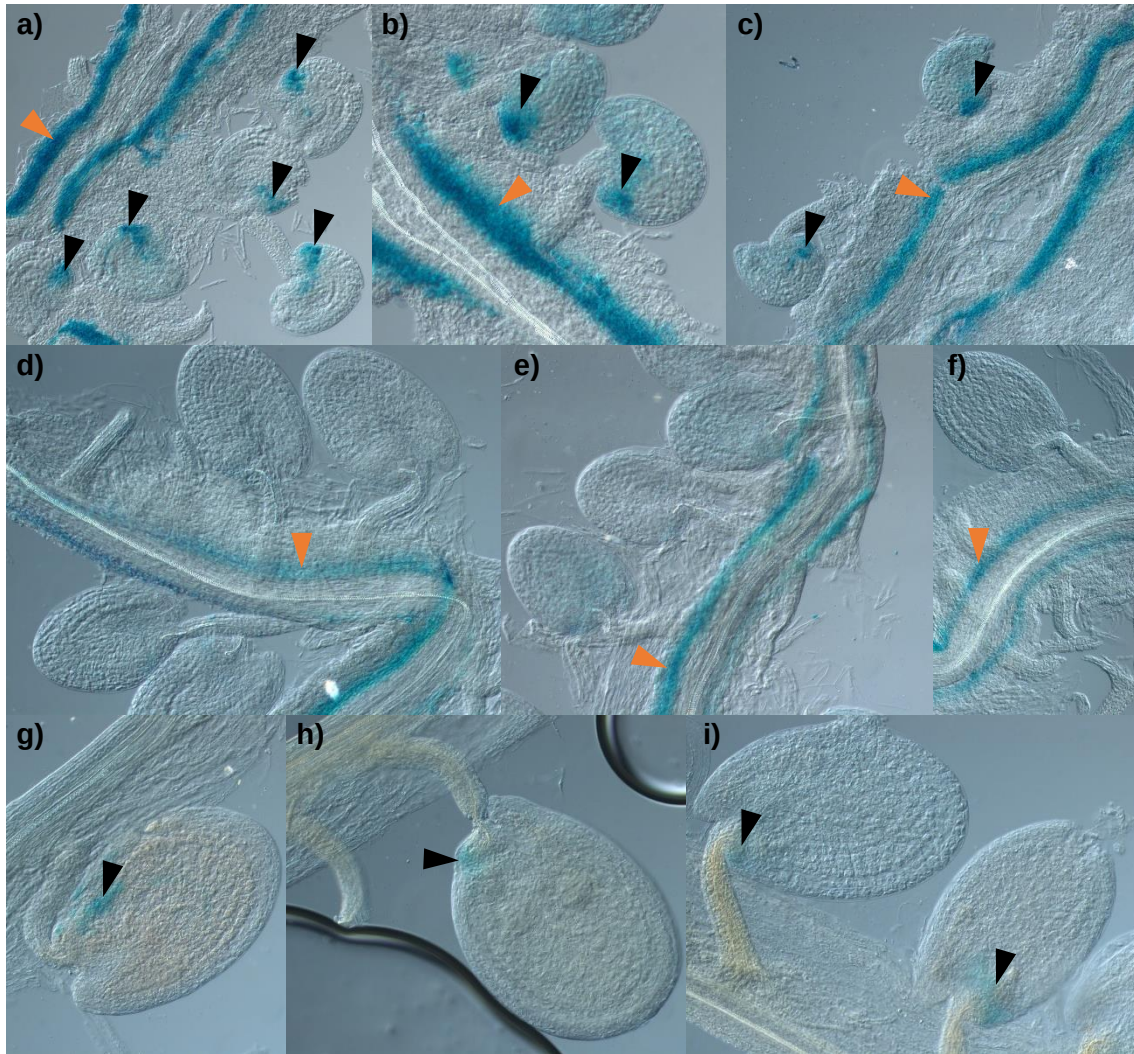


Figure 44. Histochemical detection of GUS activity regulated by the *AGP31* gene promoter in ovules and seeds. In flowers and pistils collected at stage 13 (**a**) to **c**) of flower development according to Smyth *et al.* (1990), GUS activity was detected in the ovules' chalazal pole (black triangles) and in the transmitting tissue (orange triangles). At stage 15 (**d**) to **f**), GUS expression is limited to the transmitting tissue and, at stage 17 (**g**) to **i**), GUS activity almost disappeared, being only observed a tenuous expression in the seeds' chalazal pole.

4.5.3. Putative AGPs target genes expression pattern in *stk* mutants

After performing controlled crosses performed between *stk* mutant plants and transgenic lines for the *pAGP9:NLS:3GFP*, *pAGP21:GUS*, *pAGP24:GUS*, and *pAGP31:GUS* constructs, the seeds collected from the siliques formed after these crosses were sowed. Seeds collected from siliques originated as consequence of analogous crossings previously performed in the laboratory were also used. Only seeds formed in siliques used in the crossings between *stk* mutant plants and transgenic lines for the *pAGP24:GUS*, and *pAGP31:GUS* constructs germinated. The germinated plants were genotyped and the respective phenotypes were analysed, in order to understand if the *STK* gene has any influence upon the expression patterns of the putative target genes. The histochemical detection of the *GUS* gene expression was performed with flowers

and siliques collected at stages 13, 15 and 17 of flower development according to Smyth *et al.* (1990). At all stages and in flowers and pistils collected following both crossings that originated seeds that germinated (Figures 45 and 46), no *GUS* activity was detected in the female tissues. *GUS* activity was detected in the nectaries of *stk* mutant plants transgenic for the pAGP31:*GUS* construct (Figure 47), proving that these plants were indeed transgenic for the pAGP31:*GUS* construct and that the histochemical detection of the *GUS* gene expression protocol worked and did not influence the results. However, relatively to the plant material collected from *stk* mutant plants transgenic for the pAGP24:*GUS* construct, no *GUS* expression was detected at all. Furthermore, only one *stk* mutant plant transgenic for this construct germinated. Thus, additional plants with the same genotype would be needed to confirm the results in this last case.



Figure 45. Histochemical detection of *GUS* activity regulated by the *AGP24* gene promoter in ovules and seeds of *stk* mutant plants. In flowers and pistils collected at stages 13 (a)), 15 (b)) and 17 (c)) of flower development according to Smyth *et al.* (1990), no *GUS* expression was detected.

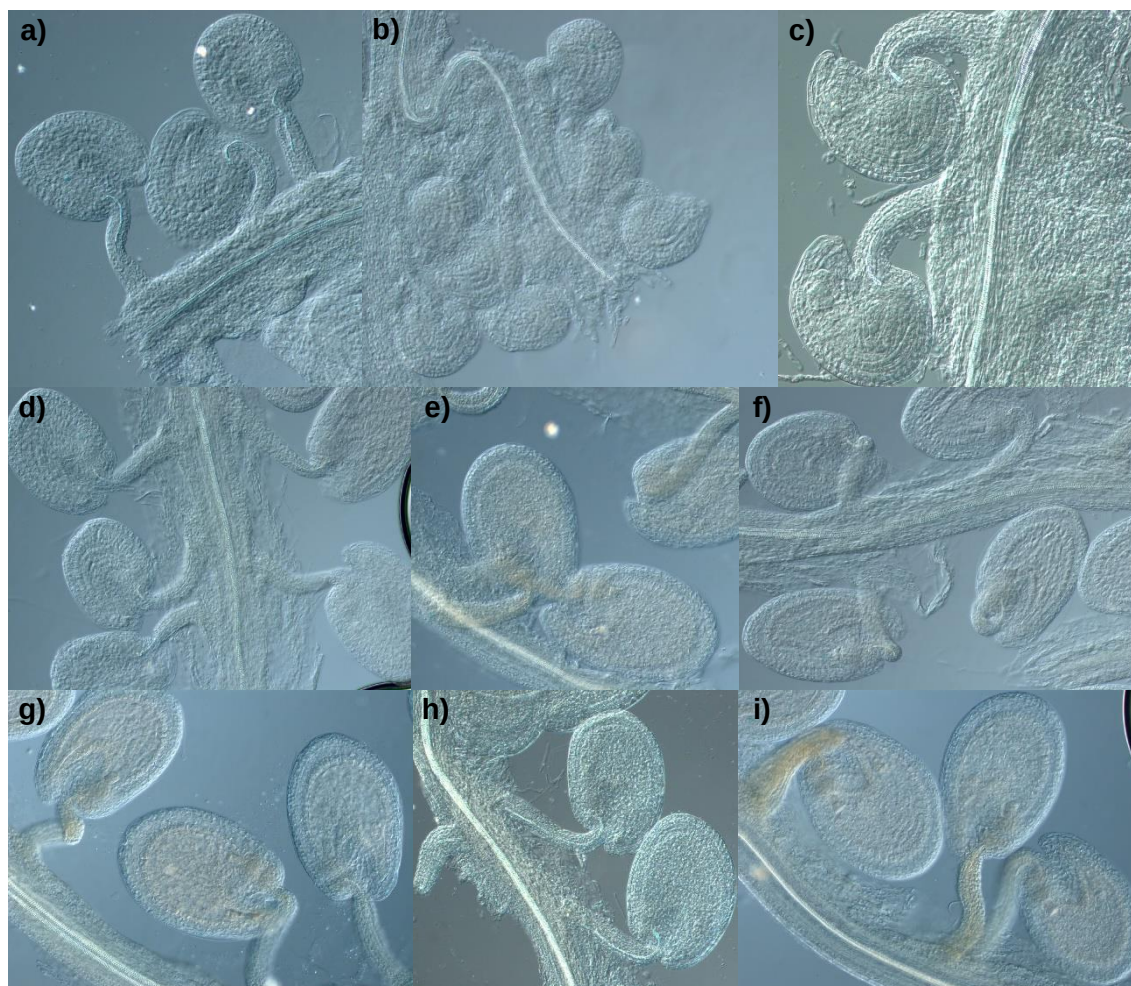


Figure 46. Histochemical detection of GUS activity regulated by the *AGP31* gene promoter in ovules and seeds of *stk* mutant plants. In flowers and pistils collected at stages 13 (a), 15 (b) and 17 (c) of flower development according to Smyth *et al.* (1990), no GUS expression was detected.

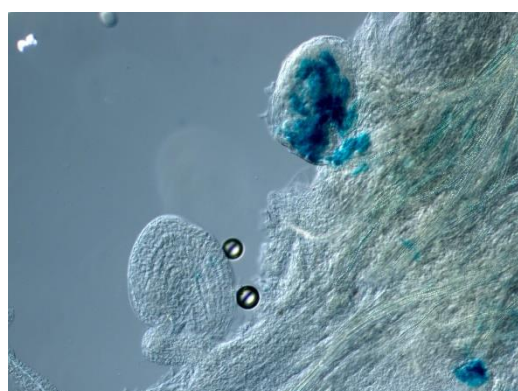


Figure 47. Histochemical detection of GUS activity regulated by the *AGP31* gene promoter in nectaries of *stk* mutant plants.

5. Discussion

The world's increasing population requires an even greater growth in food production. However, this production amplification cannot depend on an expansion of the arable land. So, only the study of plant reproduction and the molecular mechanisms behind it will enable the needed increase in seed production.

Sexual plant reproduction in angiosperms is a highly sophisticated process that involves several steps from the pollen grain germination in the stigma until double fertilisation and seed formation. Ovule and seed development have been intensively studied in the last decades, employing new genetic approaches that enabled the discovery of new gene expression networks.

5.1. Characterise SEEDSTICK functions in ovule and seed development

STK, which is the focus of the SEXSEED project, is a TF that controls one complex molecular system, the ovule and seed development.

Pinyopich *et al.* (2003) study revealed important information about the rule of this TF, discovering that it has an influence in seed dehiscence, silique and seed size and seed shape. In *Arabidopsis* plants, seed dehiscence happens when a constriction occurs in the funiculus abscission zone. In *stk* mutants, this zone contains more cells and these cells are bigger, resulting in an abnormal growth of the funiculus, inhibiting seed dehiscence and decreasing the available space for each seed to grow. Furthermore, *stk* mutants' siliques are smaller than WT *Arabidopsis* siliques and seeds are smaller and rounder than the WT ones (Pinyopich *et al.*, 2003). In the present work, intending to later compare with the phenotypes of double mutant plants for *STK* and its putative target genes, WT and *stk* mutant plants ovules and seeds were cleared and observed using Differential Interference Contrast (DIC) microscopy. The results obtained confirm the ones described by Pinyopich *et al.* (2003). Indeed, the *stk* seeds observed were rounder, smaller and presented a bigger funiculus when compared to WT seeds. Furthermore, *stk* siliques collected at stage 17 of flower development, according to Smyth *et al.* (1990), did not show a constriction in the abscission zone unlike what happens with WT seeds at the same stage. Once again, it was confirmed that STK is responsible for regulating the abscission zone constriction, as well as the seed size and shape.

Moreover, seed set was also compared between WT and *stk* mutant plants. In a first approach, the seed set was analysed using random siliques collected from a group of plants. Although this method is not totally wrong, since it enables to confirm if the average

seed set of a plant's silique is also preserved among the siliques of a group of plants with the same genotype, it neglects the need for biological replicates to infer if the seed set of a certain plant with a given genotype does indeed differ from the one of another plant with a different genotype, inhibiting to conclude about the real influence that the difference in the genotype has in the seed set. So, in a second approach, the seed set of random siliques collected from each one of three random WT plants was analysed, doing the same with *stk* mutant plants. As for the first approach, it could be concluded that, across the group of *stk* plants analysed, the total seed production per silique is statistically lower than each silique's seed production in the group of WT plants. Relatively to the second attempt, a WT plant produces, on average, more green seeds per silique and longer siliques than a *stk* mutant plant. A previous work by Malabarba *et al.* (2017) reported a higher number of mature seeds present in the siliques of WT plants comparatively to *stk* mutant plants' siliques. Pinyopich *et al.* (2003) also described *stk* mutant plants' siliques as being shorter than WT plants' siliques. Malabarba *et al.* (2017) also noticed a higher number of aborted ovules and aborted seeds in the *stk* mutant plants' siliques than in WT plants' siliques. Although this trend was also observed in the present work, a firm conclusion cannot be drawn, since the variances obtained for these two factors (number of aborted ovules and aborted seeds) were not homogeneous, inhibiting to conduct an ANOVA with these results in order to check their statistical significance. Thus, more replicates would need to be analysed to confirm whether there are indeed more aborted ovules and aborted seeds in *stk* mutant plants' siliques than in WT plants' siliques. Nevertheless, in the present study, the STK regulation of seed production and silique size was confirmed.

Finally, PT growth was also observed in WT and *stk* mutant plants by aniline blue staining. This technique enables to observe the PT growth across the female tissues, from germination until entering the embryo sac, because aniline blue stains the callose present at the PT wall and at the plugs that are formed along PT growth. This technique also enables to conclude whether a mutation in a certain gene affects the female and male tissues, including the stigma receptivity and the ovules viability (Dumas and Knox, 1983). The phenotype observed in *stk* plants was quite like to the one of WT plants. PTs of *stk* plants grew across pistils of the same plants without any defect and each ovule was reached by only one PT. It can be concluded that the knockout of the STK gene does not affect PT growth. Nevertheless, this technique should be performed following reciprocal crosses between *stk* mutant plants and WT plants to understand the real interference of the STK gene knockout in PT growth and PT attraction, from the male and female parts, separately. Furthermore, only performing a semi-*in vivo* assay one

could truly conclude whether the knockout of this gene has an influence on PT attraction and the time it takes to reach the ovule. In addition, as mentioned in section 1.4. of Introduction, STK is part of a quartet that contains other TFs with redundant functions (Favaro *et al.*, 2003; Pinyopich *et al.*, 2003). Thus, in *stk* plants, these other TFs might be masking the lack of STK and regulating the PT growth, indicating that only multiple mutants for the TFs' genes would enable to obtain a phenotype in this process.

5.2. Identify STK target genes

Some AGPs were recently identified as potential STK direct targets by ChIP-sequencing analyses (unpublished data). As referred in section 1.5.2. of the Introduction, AGPs are also involved in many processes related to reproduction. This increases the interest in studying the gene expression network that involves STK and these molecules.

To confirm that STK regulates the expression of four AGP genes, *AGP9*, *AGP21*, *AGP24* and *AGP31*, firstly, a Real-Time PCR was performed using cDNA from WT and *stk* mutant plants and choosing these genes as the target genes. Two attempts of RNA extraction were conducted following two slightly different protocols, always using tissues where *STK* is expressed.

In the first attempt, the plant material used were seeds from two WT and two *stk* mutant plants. The cDNA synthesis was performed twice using these RNA samples, since the RT-qPCR performed using the cDNA samples originated in the first synthesis did not present coherent results. In fact, not only the results obtained with the *stk* samples diverged a lot between each other, when comparing the expression levels of *AGP21*, *AGP24* and *AGP31* of *stk* plants with the ones of each WT plant, but it can also be observed that there was practically no consistency between selecting WT1 or WT2 cDNA as the control sample. In what concerns the PCR later performed to amplify the *STK* coding sequence (section 3.5.2.), that used the WT cDNA samples made in this first synthesis, did not lead to the formation of any band in the electrophoresis that used this PCR's products, the inconsistency of these first RT-qPCR results was associated with a possible poor condition of the cDNA samples. Therefore, the cDNA synthesis and the RT-qPCR were repeated. Furthermore, unlike the first RT-qPCR in which none of the melting temperatures related to the *AGP9* target gene reached the threshold of the melt peak graphic, inhibiting to properly analyse the relative normalised expression of this gene, in the second RT-qPCR a new pair of optimised primers were used in the reactions performed for this gene, trying to include this target gene in the RT-qPCR results.

Nevertheless, in the second RT-qPCR the inconsistencies maintained. Furthermore, when using the WT1 cDNA sample to amplify the *ACT8* gene in the semi-quantitative RT-

PCR, the band formed in the respective electrophoresis was tenuous, indicating again that this cDNA sample could be fragmented. Despite the contradictions in the results obtained in both RT-qPCRs between selecting WT1 or WT2 cDNA as the control sample, the results were practically the same when comparing both RT-qPCR attempts using the same WT cDNA sample, indicating that the inconsistencies found resided in the RNA samples, upstream of the cDNA synthesis.

In order to try to overcome this problem, RNA extraction was repeated, but using only one WT sample and three *stk* samples, to avoid differences between WT results and enable to discard one *stk* sample in case it diverged from the other two. These RNA extractions used a different plant material and in a bigger amount. This time RNA was extracted from pistils, excluding the valves, to obtain higher concentrations of cDNA necessary for the EMSA technique. Two of the RNA samples did not present a good quality when submitted to an electrophoresis, WTI and *stkII*. These samples were also used in a cDNA synthesis performed later, as explained in section 4.2., which could create discrepancies in the cDNA quality. Although all the cDNA samples amplified the *ACT8* gene in the semi-quantitative RT-PCR, the *stkII* sample results had to be discarded from the final RT-qPCR, since they diverged a lot from the other two *stk* samples, maybe because of the poor condition of the RNA samples and/or the difference in the cDNA synthesis process. However, once again, the two remaining *stk* samples presented quite different results in the RT-qPCR.

Overall, analysing the three RT-qPCRs, even though they all presented very inconsistent results, the first and second attempts seemed a bit more coherent when comparing the results of both *stk* samples with the ones of the WTI sample relatively to the comparison with the WTII sample. In order to summarise these results, the *AGP9* gene seems down-regulated in the *stk* mutant samples when considering the second RT-qPCR, but this is contradicted by the third RT-qPCR, whose results seem to indicate an up-regulation. The *AGP21* gene expression, when looking at the comparisons made with the WTII sample in the first two RT-qPCRs and considering the third RT-qPCR, seems to be down-regulated in the *stk* mutant samples, but only one of the *stk* samples (*stkII*) shows a lower expression of these genes when compared to the WTI sample in the first two RT-qPCRs. As for the results for the *AGP24* gene expression, the first two indicate a down-regulation of this gene in the *stk* mutant samples when compared to the WTI sample, but an up-regulation when compared to the WTII sample. The RT-qPCR indicates a possible up-regulation of this gene at least in one of the *stk* samples. The expression levels of the *AGP31* gene presents a pattern quite like to the ones described for the *AGP24* gene, but the third RT-qPCR points towards a down-regulation of this gene in the *stk* mutants.

To sum up this analysis of the expression quantification of each putative STK target *AGP* gene, this last approach would need to be repeated and improved, repeating the RNA extractions in the same conditions for all the samples used and repeating the RT-qPCR studies in order to create a case study that would enable to draw coherent conclusions about the STK influence in the expression levels of each one of the selected target genes. As for the *ACT8* and *RUB1* genes, these genes are a common choice in the laboratory as reference genes in the RT-qPCR protocol. However, new genes may need to be chosen, since the instability of these genes' expression might be causing the inconsistencies observed in this work. For instance, while Czechowski *et al.* (2005) identify *ACT8* as a stable reference gene, Dekkers *et al.* (2012) described this gene as being one of the genes with a less stable expression from a group of 42 putative reference genes for seed research, based on microarray expression data.

Another way to evaluate the influence of STK in the expression of putative target genes is to determine if the expression pattern of these genes is altered in *stk* mutants. To access this, firstly, the expression patterns of each gene in *Arabidopsis* ovules and seeds were determined by analysing the expression of reporter genes guided by the promoters of the target genes. Relatively to the *AGP9* gene, a transgenic line for a construct containing the *GFP* gene regulated by the *AGP9* gene promoter was analysed. This analysis allowed to conclude, as it was described previously by Pereira *et al.* (2014), that *AGP9* is expressed in the vascular tissues of the transmitting tract and funiculus and in ovules and seeds chalazal pole. This might indicate that the *AGP9* gene is involved in the nutrition or signalling between the pistil and the ovule, especially the endosperm, since the main nutrient absorption into the endosperm coming from the maternal parent happens through the chalazal pole (Debeaujon *et al.*, 2003; Ingram, 2010). As for the other *AGPs* genes, transgenic lines for constructs with the *GUS* gene guided by the promoter of each putative target gene were analysed. As regards to *AGP21* gene, its expression pattern is present in ovules and seeds micropylar pole and in the pistil septum at earlier stages of flower development, which may suggest that this gene is involved in PT attraction and/or entrance in the female gametophyte, given that synergids, which are present in the micropylar pole, are responsible for PT attraction and that the PT enters the embryo sac through them, as referred in section 1.5.2. of the Introduction. At later stages, the expression pattern of the *AGP21* gene becomes more diffuse across the endosperm, indicating that this gene can be implicated in embryo nutrition. The presence of the *AGP21* gene expression in the septum can also be related to nutrient supply coming from the maternal tissues. With regards to the expression pattern of the *AGP24* gene, although it is present in ovules and seeds micropylar and chalazal poles

at earlier stages of flower development, at later stages, it spreads across the endosperm, maintaining the expression in seeds chalazal pole. For the same reasons referred before, it can be suggested that the function of the *AGP24* gene is associated with nutrition purposes, due to its expression in the chalazal pole and endosperm, and with PT attraction/entrance in the embryo sac, given its expression in the micropylar pole. Finally, as for the expression of the *AGP31* gene, it is located in ovules chalazal pole and in the transmitting tissues, disappearing progressively at the following stages, which indicates that this gene might be involved in the embryo sac nutrition, given its expression in the chalazal pole, and in PT guidance across the transmitting tissue.

These results were compared to the expression patterns of the same genes, but in *stk* mutant lines transgenic for the same constructs. The only plants that germinated from the crossings performed to obtain these lines were *stk* mutant plants transgenic for the p*AGP24*:GUS and p*AGP31*:GUS constructs. Relatively to the first ones, the expression pattern of the *AGP24* gene in *stk* mutants was completely absent. These results are in accordance with the ones of the first two RT-qPCRs when comparing the *AGP24* expression level in the *stk* samples to the WTI sample, which indicate a down-regulation of this gene in the *stk* mutant samples. However, only one plant was observed with no GUS expression in other tissues of the plant. Thus, more transgenic plants of this line would need to be analysed in order to confirm these results. As for the *stk* mutant plants transgenic for the p*AGP31*:GUS construct, the expression of the *AGP31* gene was not present in any female tissue. These results coincide with the ones of the third RT-qPCR performed, but also with the results of the first two RT-qPCRs when comparing the *AGP31* expression level in the *stk* samples to the WTI sample, which indicate a down-regulation of this gene in the *stk* mutant samples. In this case, GUS activity was detected in other tissues of the plant and more than one plant of this line showed this pattern, leading to conclude that the *AGP31* gene may indeed be a target of the STK TF.

5.3. STK direct interaction with the putative AGPs target genes

In order to unravel how the gene regulatory network in which STK is involved works, two techniques were employed to discover if STK regulates and binds directly to its putative target genes under study. The first was the Electrophoretic Mobility Shift Assay (EMSA). As described in section 1.6.1. of the Introduction, this procedure was already used in investigations related to plant reproduction and it is extensively used to study DNA-protein interactions. This method evaluates if the presence of a given protein in a sample of a certain DNA fragment delays the electrophoretic run of that fragment, indicating that

the protein binds to it. If the interaction under study is between a gene and a TF, most probably the sequence of the DNA fragment that needs to use in the assay corresponds to the binding site motif associated with the TF family. Therefore, in the present work, possible CARG boxes were identified in the *AGP9*, *AGP24* and *AGP31* genes *in silico* as described in section 3.4.1.. Then, the *STK* coding sequence was amplified using WT cDNA, to be cloned into the pET28a expression vector. In the ligation process, this vector was digested, as well as the amplified fragment. However, right after this step and in the following steps of the ligation procedure, the electrophoreses performed did not reveal the expected bands, making it impossible to obtain the final recombinant vector. As explained in section 4.3.1., this may have happened due to low DNA concentrations in the amplified fragment sample and in the linearised plasmid sample, since not even the band corresponding to the non-recombined vector was formed. This way, this procedure would need to be repeated, improving each step, to increase the amounts of DNA amplified and avoid DNA losses.

The second technique performed was the Protoplast Transactivation Assay (PTA). The protocol performed in this work followed what was described by Ueda *et al.* (2011), who used it for the study of a DNA-protein interaction involved in embryo development. This assay tests the binding of a TF to a certain gene *in planta*. Ueda *et al.* (2011) inserted constructs containing a reporter gene linked to each *cis* element of the target gene and a construct with the TF sequence into protoplasts. This way, if the TF interacts with the *cis* elements of the target gene, the reporter gene expression will be detected. Firstly, a section of the promoter, the 5' UTR and the 3' UTR segments were selected as potential *cis* elements of the *AGP24* and *AGP31* genes. Then, to clone them into the pBluescript II SK (+) vector, each segment was amplified. Nevertheless, only the potential *cis* elements of the *AGP31* gene and the 5' UTR of the *AGP24* gene were amplified, requiring an optimisation of the amplification of the promoter segment of this last gene. As for the pBluescript II SK (+) plasmid, its digestion reaction, necessary for the ligation process, led to the formation of a band in the respective electrophoresis that might correspond to the linearised vector. However, as the band corresponding to the non-linearised plasmid was also formed, the restriction reaction needs to be adjusted, to completely digest the plasmid miniprep sample.

Finally, the MU1110 plasmid, used by Ueda *et al.* (2011) for assembling the construct that contained the TF gene, was also digested together with an amplified fragment that corresponded to the *STK* sequence, in order to join both parts. However, once again, after the ligation step and restriction analysis of the product, the electrophoresis did not reveal any band. As described for the EMSA technique, this may have occurred due to

low DNA concentrations in the amplified fragment sample and in the linearised plasmid sample and each step would need to be improved in order to maximise the amounts of DNA obtained.

Although, none of these two assays was completed due to the lack of time available, they should not be underestimated. The ligation protocols are complex and need continuous improvements of each specific step. Furthermore, PTA requires several protoplast transfection steps, making it a complex and time-consuming technique. Nonetheless, PTA is an *in planta* technique, enabling to get a reliable picture of DNA-protein interactions in plant cells. In its turn, although being an *in vitro* technique, EMSA is a faster and easier approach. In conclusion, both techniques can be helpful in the understanding of gene regulatory networks, like the one controlled by STK. In a near future, it would also be interesting to perform the Chromatin ImmunoPrecipitation (ChIP), an efficient experimental technique used for accurately characterise bound for the functional protein to DNA molecule.

5.4. Create double mutants for STK and each one of its putative AGPs target genes

CRISPR-Cas9 is a recent technique that establishes a fast way to create new plant mutant lines and higher order mutant lines (Hsu *et al.*, 2014). Wang *et al.* (2015) developed an approach using the CRISPR-Cas9 technique in *Arabidopsis thaliana* that leads to the transmission of the mutation from one generation to the next. This approach uses a vector (pHEE401E+Blii) that contains the Cas9 gene regulated by a constitutive promoter, which enables the expression of Cas9 in ovules, or at an initial embryonic stage. Therefore, the mutations caused in these cells will be passed to most of the progeny.

In the present work, the protocol used by Wang *et al.* (2015) was followed in order to create mutations in STK putative target genes (*AGP1*, *AGP4*, *AGP9*, *AGP15*, *AGP21*, *AGP24* and *AGP31*), either in WT plants, either in *stk* mutant plants, creating double mutants. As for the first transformation attempt, only one seed, collected after the transformation with constructs for the *AGP31* gene knockout, germinated. This seed presented red fluorescence when observed in the inverted microscope. The explanation behind this low number of plants that germinated may relay in the fact that the plants used in this first transformation were too old, at the end of their life cycle. So, only a few siliques were formed following the transformation procedure and few seeds were collected and sowed.

With regards to the second transformation attempt, six plants germinated from the seeds that had been produced by the plants used in the transformation with the constructs for the *AGP9* gene knockout. One of these plants germinated from the only seed that had not presented red fluorescence from all the seeds that germinated, maybe because the mCherry gene expression in this seed was not high enough to be detected.

Only two seeds produced after the transformation with the vectors for the *AGP21* gene knockout germinated, both with red fluorescence. As for the transformation with the constructs for the *AGP31* gene knockout, one seed germinated, also with red fluorescence.

None of the seeds from the transformation with the constructs for the *AGP24* gene knockout germinated, although some of them presented red fluorescence. This may have happened because the pHEE401E+BliI plasmids that recombined with the fragments containing the sgRNAs specific cassettes contained a mutation in the resistance gene that blocked its expression. Another possible explanation is that the knockout of this gene inhibits germination. The plasmids containing the constructs assembled for the *AGP24* gene knockout would need to be created again and the resistance gene of the pHEE401E+BliI used should be sequenced.

Only the WT plants used in the transformation procedures produced seeds that germinated. This may have happened because *stk* mutant plants used in the transformation procedures produced less seeds. Nevertheless, these transformations would need to be repeated using a higher number of younger plants.

Finally, the germinated plants genomic DNA was extracted and a fragment embracing the two sgRNAs specific cassettes for the knockout of the potentially mutated gene was amplified and sequenced in order to detect any mutations when comparing the obtained sequence with the published one. The germinated plants were not mutants, since no significant differences were detected between the obtained and the original sequences, mainly in the segments next to the sgRNAs specific cassettes, unlike what was expected. As referred in the Introduction, the Cas9 enzyme is expected to recognise the PAM sequence, which is located next to the sgRNAs specific cassettes in the plant genome, and cut the target DNA. However next to the sgRNAs specific cassettes sequenced there are no mutations. Furthermore, like each fragment contained two sgRNAs specific cassettes, if the Cas9 worked and cut the target genes, the target genes should contain a big deletion next to the sgRNAs specific cassettes, which did not happen.

The amplified fragments that embrace the sequences homologous to the four sgRNAs specific cassettes designed for the *AGP21* gene were not properly sequenced, which

would require adjustments of the previous steps in order to conclude whether the respective plants were mutants or not.

With regards to the plant that germinated from the single seed that did not present red fluorescence and that was produced after the transformation for the *AGP9* gene knockout, the fragment that would be amplified with the DNA of this plant was not even amplified, which could be explained by the poor condition of its genomic DNA, since the other plants used in this specific transformation did not present this problem.

Relatively to the knockout of the *AGP1*, *AGP4*, *AGP15*, *CKX5* and *CKX7* genes, the transformation process should be performed using these genes' knockout constructs in a nearby future, given their involvement in several mechanisms related to ovule and seed development.

It cannot be neglected that this technique also requires the performance of several attempts and replicates, since the Cas9 enzyme may not always operate. Nevertheless, the assembly of the final recombinant vectors proved to be a very efficient and fast protocol.

6. Conclusions

This study intended to unravel some of the mechanisms and gene interactions controlling sexual plant reproduction and seed development, more specifically the ones related to STK. The phenotypes of *stk* mutant plants' ovules and seeds were described so that they could be later compared to double mutants for *STK* and its putative target genes. Seeds from *stk* present a smaller size and a rounder shape than WT seeds and the funiculus abscission zone does not constrict. As for seed production and siliques' length, *stk* plants produce less mature green seeds per silique and shorter siliques.

AGP9, *AGP21*, *AGP24* and *AGP31*, putative STK target genes, expressions levels and patterns were analysed in WT and *stk* plants. *AGP9* expression was observed in the ovules' and seeds' chalazal pole and in the vascular tissues of the transmitting tissue and funiculus. *AGP21* gene expression was detected in the septum and micropylar pole of ovules. The *AGP24* expression was observed in the ovules' micropylar and chalazal pole at early stages of flower development and an intense expression in the endosperm at later stages. With regards to the expression pattern of *AGP31* gene, it was detected in ovules' chalazal pole and in the transmitting tissue. The expression levels obtained by Real-time PCR were not conclusive and need to be improved. Nevertheless, they point towards a down-regulation of the *AGP31* gene in the *stk* mutant plants, which is supported by the absence of an expression of this gene in the *stk* mutants. This absence

of expression in the *stk* mutant was also verified for the *AGP24* gene, but these results need validation.

To verify if STK binds directly to these potential target genes *in vitro* and *in vivo*, Electrophoretic Mobility Shift Assay (EMSA) and Protoplast Transactivation Assay (PTA) techniques were initiated. Still some improvements need to be done to properly complete these methods.

Finally, the CRISPR-Cas9 technique was performed to create double mutants for *STK* and the putative target genes. Although no mutant plant has yet been obtained by this technique, the protocol used to assemble the recombinant plasmids employed in this technique proved to be efficient and fast.

Concluding, although the results obtained have not yet been conclusive, they point towards a down-regulation of the *AGP31* gene and eventually of the *AGP24* gene in *stk* mutant plants. However, as it was noted, almost every technique should be repeated and solid results should be obtained to confirm this and to unravel a little bit more about the gene regulatory network controlled by STK. Furthermore, in a near future it would also be interesting to perform this study using the constructs assembled for the knockout of the *AGP1*, *AGP4*, *AGP15*, *CKX5* and *CKX7* genes, since this putative *STK* targets are involved in many processes related to ovule and seed development.

7. References

- Alting-Mees M, Short J** (1989) pBluescript II: gene mapping vectors. *Nucleic acids research* **17**: 9494
- Baker SC, Robinson-Beers K, Villanueva JM, Gaiser JC, Gasser CS** (1997) Interactions among genes regulating ovule development in *Arabidopsis thaliana*. *Genetics* **145**: 1109-1124
- Bartrina I, Otto E, Strnad M, Werner T, Schmulling T** (2011) Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*. *Plant Cell* **23**: 69-80
- Battaglia RB, V.; Colombo. L.** (2008) Morphological analysis of female gametophyte development in the *bel1 stk shp1 shp2* mutant. *Plant Biosyst* **142**: 643-649
- Becker MG, Hsu SW, Harada JJ, Belmonte MF** (2014) Genomic dissection of the seed. *Front Plant Sci* **5**: 464
- Berger F** (2003) Endosperm: the crossroad of seed development. *Curr Opin Plant Biol* **6**: 42-50
- Blamey MG-W, C.** (1989) Illustrated flora of Britain and Northern Europe. Hodder and Stroughton, Sevenoaks, Kent, UK
- Brambilla V, Battaglia R, Colombo M, Masiero S, Bencivenga S, Kater MM, Colombo L** (2007) Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in *Arabidopsis*. *Plant Cell* **19**: 2544-2556
- Brambilla V, Kater M, Colombo L** (2008) Ovule integument identity determination in *Arabidopsis*. *Plant Signaling & Behavior* **3**: 246-247
- Cocking EC** (1960) A method for the isolation of plant protoplasts and vacuoles. *Nature* **187**: 962-963
- Colombo L, Franken J, Koetje E, van Went J, Dons H, Angenent GC, van Tunen AJ** (1995) The petunia MADS box gene *FBP11* determines ovule identity. *The Plant Cell* **7**: 1859-1868
- Costa M, Nobre MS, Becker JD, 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. *BMC Plant Biology* **13**: 7
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R** (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant physiology* **139**: 5-17

- Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L** (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. *The Plant Cell* **15**: 2514-2531
- Dekkers BJ, Willems L, Bassel GW, van Bolderen-Veldkamp R, Ligterink W, Hilhorst HW, Bentsink L** (2012) Identification of reference genes for RT-qPCR expression analysis in *Arabidopsis* and tomato seeds. *Plant and Cell Physiology* **53**: 28-37
- Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE** (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature biotechnology* **32**: 1262-1267
- Doudna JA, Charpentier E** (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**
- Dresselhaus T, Coimbra S** (2016) Plant Reproduction: AMOR Enables Males to Respond to Female Signals. *Curr Biol* **26**: R321-323
- Dresselhaus T, Franklin-Tong N** (2013) Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Mol Plant* **6**: 1018-1036
- Dumas C, Knox R** (1983) Callose and determination of pistil viability and incompatibility. *Theoretical and Applied Genetics* **67**: 1-10
- 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**: 1349
- Engler C, Gruetzner R, Kandzia R, Marillonnet S** (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PloS one* **4**: e5553
- Engler C, Kandzia R, Marillonnet S** (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS one* **3**: e3647
- Ezquer I, Mizzotti C, Nguema-Ona E, Gotté M, Beauzamy L, Viana VE, Dubrulle N, de Oliveira AC, Caporali E, Koroney A-S** (2016) The developmental regulator SEEDSTICK controls structural and mechanical properties of the *Arabidopsis* seed coat. *The Plant Cell* **28**: 2478-2492
- Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L** (2003) MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* **15**: 2603-2611
- Ferrándiz C, Pelaz S, Yanofsky MF** (1999) Control of carpel and fruit development in *Arabidopsis*. *Annual review of biochemistry* **68**: 321-354
- Fried M, Crothers DM** (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic acids research* **9**: 6505-6525

- Galuszka PF, I.; Šebela, M.; Peč, P.** (2000) Degradation of cytokinins by cytokinin oxidases in plants. *Plant Growth Regulation* **32**: 315–327
- Garner MM, Revzin A** (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic acids research* **9**: 3047-3060
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO** (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**: 343-345
- Guille MJ, Kneale GG** (1997) Methods for the analysis of DNA-protein interactions. *Molecular biotechnology* **8**: 35-52
- Hanahan D** (1983) Studies on transformation of *Escherichia coli* with plasmids. *Journal of molecular biology* **166**: 557-580
- Hellman LM, Fried MG** (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature protocols* **2**: 1849
- Higashiyama T, Yabe S, Sasaki N, Nishimura Y, Miyagishima S, Kuroiwa H, Kuroiwa T** (2001) Pollen tube attraction by the synergid cell. *Science* **293**: 1480-1483
- Hijazi M, Durand J, Pichereaux C, Pont F, Jamet E, Albenne C** (2012) Characterization of the arabinogalactan protein 31 (AGP31) of *Arabidopsis thaliana*: new advances on the Hyp-O-glycosylation of the Pro-rich domain. *J Biol Chem* **287**: 9623-9632
- Hsu PD, Lander ES, Zhang F** (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278
- Ingram GC** (2010) Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195-214
- Jameson PE, Song J** (2016) Cytokinin: a key driver of seed yield. *J Exp Bot* **67**: 593-606
- Kass J, Artero R, Baylies MK** (2000) Non-radioactive electrophoretic mobility shift assay using digoxigenin-ddUTP labeled probes. *Drosophila Information Service* **83**: 185-188
- Köllmer I, Novak O, Strnad M, Schmulling T, Werner T** (2014) Overexpression of the cytosolic cytokinin oxidase/dehydrogenase (CKX7) from *Arabidopsis* causes specific changes in root growth and xylem differentiation. *Plant J* **78**: 359-371
- Komor AC, Badran AH, Liu DR** (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**: 20-36
- Kooiker M, Airoidi CA, Losa A, Manzotti PS, Finzi L, Kater MM, Colombo L** (2005) BASIC PENTACYSINE1, a GA binding protein that induces conformational changes

in the regulatory region of the homeotic *Arabidopsis* gene *SEEDSTICK*. *The Plant Cell* **17**: 722-729

Lalanne E, Honys D, Johnson A, Borner GH, Lilley KS, Dupree P, Grossniklaus U, Twell D (2004) *SETH1* and *SETH2*, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in *Arabidopsis*. *Plant Cell* **16**: 229-240

Lamport DT, Varnai P (2013) Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. *New Phytol* **197**: 58-64

Lee JH, Skowron PM, Rutkowska SM, Hong SS, Kim SC (1996) Sequential amplification of cloned DNA as tandem multimers using class-II restriction enzymes. *Genetic analysis: biomolecular engineering* **13**: 139-145

Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen L-L (2017) CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Molecular plant* **10**: 530-532

Lopez-Hernandez F, Tryfona T, Rizza A, Xiaolan LY, Harris MO, Webb AA, Kotake T, Dupree P (2020) Calcium Binding by Arabinogalactan Polysaccharides Is Important for Normal Plant Development. *The Plant Cell* **32**: 3346-3369

Losa A, Colombo M, Brambilla V, Colombo L (2010) Genetic interaction between *AINTEGUMENTA* (*ANT*) and the ovule identity genes *SEEDSTICK* (*STK*), *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*). *Sexual plant reproduction* **23**: 115-121

Lowder LG, Zhang D, Baltes NJ, Paul JW, Tang X, Zheng X, Voytas DF, Hsieh T-F, Zhang Y, Qi Y (2015) A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant physiology* **169**: 971-985

Malabarba J, Buffon V, Mariath JE, Gaeta ML, Dornelas MC, Margis-Pinheiro M, Pasquali G, Revers LF (2017) The MADS-box gene *Agamous-like 11* is essential for seed morphogenesis in grapevine. *Journal of experimental botany* **68**: 1493-1506

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826

Mansfield SGB, L. G. (1991) Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo. *Canadian journal of botany* **69**: 461-476

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

McCormick S (2004) Control of male gametophyte development. *Plant Cell* **16 Suppl**: S142-153

- Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M** (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* **282**: 662, 679-682
- Mendes MA, Guerra RF, Berns MC, Manzo C, Masiero S, Finzi L, Kater MM, Colombo L** (2013) MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *The Plant Cell* **25**: 2560-2572
- Mendes MA, Guerra RF, Castelnovo B, Silva-Velazquez Y, Morandini P, Manrique S, Baumann N, Groß-Hardt R, Dickinson H, Colombo L** (2016) Live and let die: a REM complex promotes fertilization through synergid cell death in *Arabidopsis*. *Development* **143**: 2780-2790
- Mizukami AG, Inatsugi R, Jiao J, Kotake T, Kuwata K, Ootani K, Okuda S, Sankaranarayanan S, Sato Y, Maruyama D** (2016) The AMOR arabinogalactan sugar chain induces pollen-tube competency to respond to ovular guidance. *Current Biology* **26**: 1091-1097
- 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* seed coat. *PLoS Genet* **10**: e1004856
- Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L** (2012) The MADS box genes *SEEDSTICK* and *ARABIDOPSIS B_{sister}* play a maternal role in fertilization and seed development. *Plant J* **70**: 409-420
- Moreira D, Pereira AM, Lopes AL, Coimbra S** (2020) The best CRISPR/Cas9 versus RNA interference approaches for Arabinogalactan proteins' study. *Molecular Biology Reports*: 1-11
- Murashige T, Skoog F** (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia plantarum* **15**: 473-497
- Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L** (2002) The *TRANSPARENT TESTA16* locus encodes the *ARABIDOPSIS BSISTER* MADS domain protein and is required for proper development and pigmentation of the seed coat. *Plant Cell* **14**: 2463-2479
- Okuda S, Tsutsui H, Shiina K, Sprunck S, Takeuchi H, Yui R, Kasahara RD, Hamamura Y, Mizukami A, Susaki D** (2009) Defensin-like polypeptide LUREs are pollen tube attractants secreted from synergid cells. *Nature* **458**: 357-361
- Oñate-Sánchez L, Vicente-Carbajosa J** (2008) DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC research notes* **1**: 93
- Padgett KA, Sorge JA** (1996) Creating seamless junctions independent of restriction sites in PCR cloning. *Gene* **168**: 31-35

- Pajoro A, Biewers S, Dougali E, Leal Valentim F, Mendes MA, Porri A, Coupland G, Van de Peer Y, Van Dijk AD, Colombo L** (2014) The (r) evolution of gene regulatory networks controlling *Arabidopsis* plant reproduction: a two-decade history. *Journal of Experimental Botany* **65**: 4731-4745
- Palanivelu R, Tsukamoto T** (2012) Pathfinding in angiosperm reproduction: pollen tube guidance by pistils ensures successful double fertilization. *Wiley Interdiscip Rev Dev Biol* **1**: 96-113
- Pereira AM, Lopes AL, Coimbra S** (2016a) Arabinogalactan Proteins as Interactors along the Crosstalk between the Pollen Tube and the Female Tissues. *Front Plant Sci* **7**: 1895
- Pereira AM, Masiero S, Nobre MS, Costa ML, Solis MT, Testillano PS, Sprunck S, Coimbra S** (2014) Differential expression patterns of arabinogalactan proteins in *Arabidopsis thaliana* reproductive tissues. *J Exp Bot* **65**: 5459-5471
- Pereira AM, Nobre MS, Pinto SC, Lopes AL, Costa ML, Masiero S, Coimbra S** (2016b) "Love Is Strong, and You're so Sweet": JAGGER Is Essential for Persistent Synergid Degeneration and Polytubey Block in *Arabidopsis thaliana*. *Mol Plant* **9**: 601-614
- Pereira AM, Pereira LG, Coimbra S** (2015) Arabinogalactan proteins: rising attention from plant biologists. *Plant Reprod* **28**: 1-15
- Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF** (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**: 85-88
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F** (2013) Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**: 2281-2308
- Reece-Hoyes JS, Walhout AM** (2012) Yeast one-hybrid assays: a historical and technical perspective. *Methods* **57**: 441-447
- Riggs AD, Bourgeois S, Newby RF, Cohn M** (1968) DNA binding of the lac repressor. *Journal of molecular biology* **34**: 365-368
- Roeder AH, Yanofsky MF** (2006) Fruit development in *Arabidopsis*. *The Arabidopsis Book/American Society of Plant Biologists* **4**
- Rosenberg AH, Lade BN, Dao-shan C, Lin S-W, Dunn JJ, Studier FW** (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**: 125-135
- Rounsley SD, Ditta GS, Yanofsky MF** (1995) Diverse roles for MADS box genes in *Arabidopsis* development. *The Plant Cell* **7**: 1259-1269

- Schneitz K, Hülskamp, M. and Pruitt, R. E.** (1995) Wild-type ovule development in *Arabidopsis thaliana*: A light microscope study of cleared whole-mount tissue. *The Plant Journal* **7**: 731-749
- Schultz CG, P.; Oxley, D.; Youl, J.; Bacic, A.** (1998) GPI-anchors on Arabinogalactan proteins: implications for signalling in plants. *Trends Plant Sci* **3**: 1360–1385
- Serino GM, D.** (2018) *Arabidopsis thaliana* as an Experimental Organism. eLS, John Wiley & Sons, Ltd, Chichester, UK
- Showalter AM** (2001) Arabinogalactan-proteins: structure, expression and function. *Cell Mol Life Sci* **58**: 1399-1417
- Smyth DR, Bowman JL, Meyerowitz EM** (1990) Early flower development in *Arabidopsis*. *The Plant Cell* **2**: 755-767
- Studier FW** (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**: 60-89
- Studier FW, Moffatt BA** (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **189**: 113-130
- Su S, Higashiyama T** (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. *Plant reproduction* **31**: 67-75
- Takeuchi H, Higashiyama T** (2012) A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol* **10**: e1001449
- Tang X, Zheng X, Qi Y, Zhang D, Cheng Y, Tang A, Voytas DF, Zhang Y** (2016) A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Molecular plant* **9**: 1088-1091
- Tröbner W, Ramirez L, Motte P, Hue I, Huijser P, Lönig W, Saedler H, Sommer H, Schwarz-Sommer Z** (1992) *GLOBOSA*: a homeotic gene which interacts with *DEFICIENS* in the control of *Antirrhinum* floral organogenesis. *The EMBO journal* **11**: 4693-4704
- Tsai SQ, Joung JK** (2016) Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nature Reviews Genetics* **17**: 300-312
- Tsukamoto T, Qin Y, Huang Y, Dunatunga D, Palanivelu R** (2010) A role for LORELEI, a putative glycosylphosphatidylinositol-anchored protein, in *Arabidopsis thaliana* double fertilization and early seed development. *Plant J* **62**: 571-588
- Tucker MR, Koltunow AM** (2014) Traffic monitors at the cell periphery: the role of cell walls during early female reproductive cell differentiation in plants. *Curr Opin Plant Biol* **17**: 137-145

- Tucker MR, Okada T, Hu Y, Scholefield A, Taylor JM, Koltunow AM** (2012) Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in *Arabidopsis*. *Development* **139**: 1399-1404
- Ueda M, Zhang Z, Laux T** (2011) Transcriptional activation of *Arabidopsis* axis patterning genes *WOX8/9* links zygote polarity to embryo development. *Developmental cell* **20**: 264-270
- van Hengel AJ, Tadesse Z, Immerzeel P, Schols H, van Kammen A, de Vries SC** (2001) N-acetylglucosamine and glucosamine-containing arabinogalactan proteins control somatic embryogenesis. *Plant Physiol* **125**: 1880-1890
- Wang Z-P, Xing H-L, Dong L, Zhang H-Y, Han C-Y, Wang X-C, Chen Q-J** (2015) Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous
- Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmulling T** (2003) Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* **15**: 2532-2550
- Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, Wang X-C, Chen Q-J** (2014) A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC plant biology* **14**: 327
- Yan W, Chen D, Kaufmann K** (2016) Molecular mechanisms of floral organ specification by MADS domain proteins. *Curr Opin Plant Biol* **29**: 154-162
- Yariv J, Lis H, Katchalski E** (1967) Precipitation of arabic acid and some seed polysaccharides by glycosylphenylazo dyes. *Biochem J* **105**: 1C-2C
- Yoo S-D, Cho Y-H, Sheen J** (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature protocols* **2**: 1565

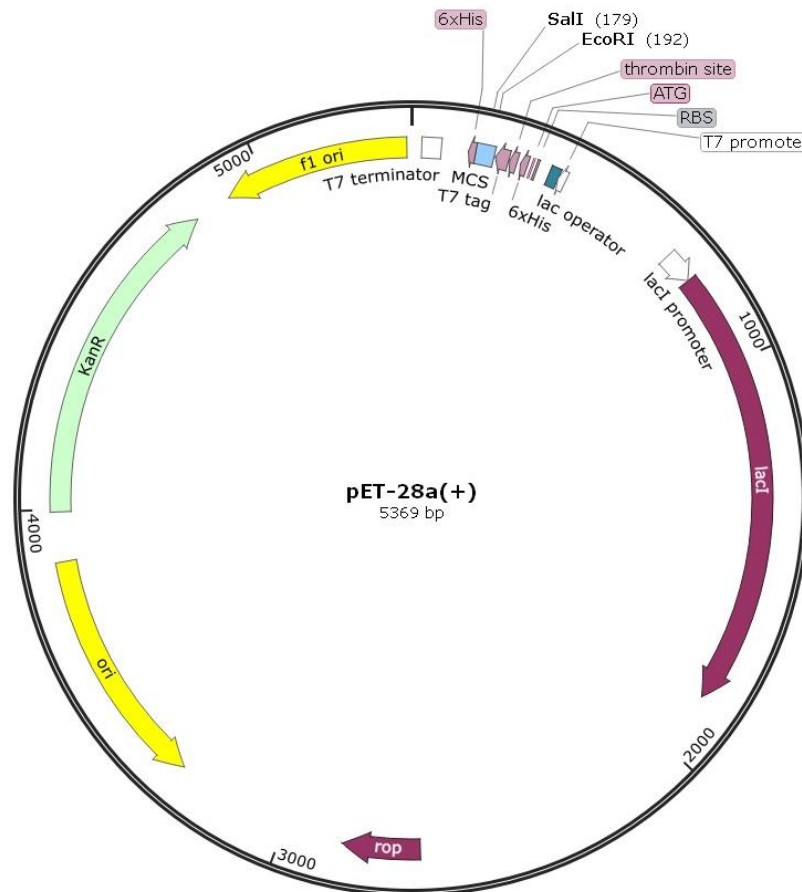
8. Annexes

Annex I

Supplemental table 1. Sequences of the oligonucleotides designed for the EMSA technique containing the CArG boxes of the *AGP9*, *AGP24* and *AGP31* genes.

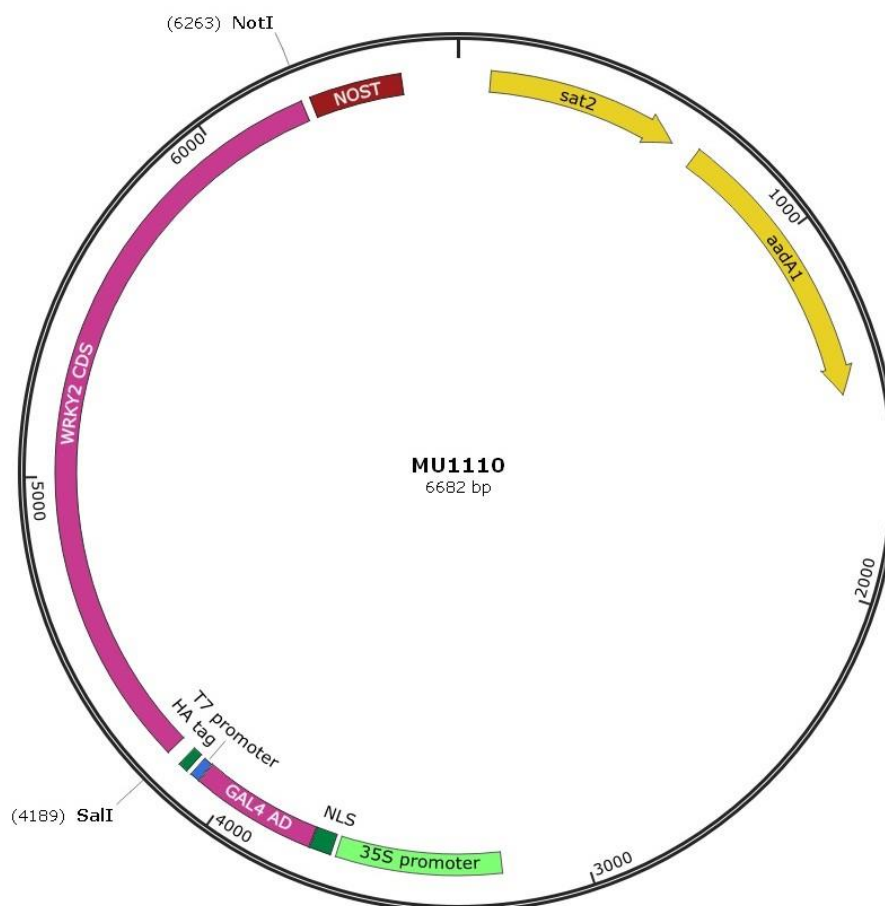
CArG box	Gene	Sequence (5'-3')	Gene	Sequence (5'-3')
C1	<i>AGP9</i>	TTGTTTTCAAATAAGGAAAAAGCT	<i>AGP24</i>	GAGCCCCATAATTAAGGTTAA
C2		CTCTATTGACTTTGAAGGTTTGT		AATTTTACCTTCATTGCCTACC
C3		CGTTGAAATCTAGATATTTGATATGATA		CCTACCCCATAGAGTCCA
C4		GATACGTAAACTTATTTGAAAACAAAA		CAATAACTCACTTTATGACAAAGATT
C5		AAACTATAAACAAATAAAGGTAACAAAAAG		CCTTCCATATAATTGGCCAA
C6		AAACTATAAACAAATAAAGGTAACAAAAAG		TAATTGGCCAATTATAGTAACGTG
C7		TTGTATTACATGTTTGGAATTAGAG		ACATATAATCTTTAACAGCTGTCTTGA
C8		ATTAGAGTAACTTAAAGCTTTATAGTTTATATTG		AATGCATCAAAAGCCGGGCCAATTCAACAAAAA
C9		CATCCCATCATAATTTTGATAAAAT		GAAACAAATCTTATTTTAGTCCTAAATT
C10		AACCTCTCACATGGCATCA		TTAAAATAAAATCTTATTTAAGATATTCAATAAA
C11		CATGGCATCATGGGTATT		TCTCATTCAATCAAAAGAAAAAAC
C12		CCAACCTCAATACGACGC		AAACTCAAACGTTTTTGAAGATA
C1	<i>AGP31</i>	AAATGTTTACTAGATGGGTAAGATT		
C2		TGTGACTTTTGTGGTCGTA		
C3		TTGTTTGGTCGTAATAGTAGAAAAAT		
C4		CATATCACATTTTTCAGATGTGTA		
C5		CAATGACCATTTTAATGTTGTTG		
C6		CTAAAGTTTTCCATCTTGAATATCTATC		
C7		AAAGAGTAACAAAAAAGAGAGTAACA		
C8		AGAGAGTAACAAATGGTACTAGTGACT		
C9		TTGGGCACAAATGGTACTA		
C10		TACTAGTAACTAATAATGGGGTAATTCATG		
C11		TGTCGATCGATTTTGATTTACT		
C12		TTACTATTTACAAATAAAGGGAGTAATTC		

Annex II



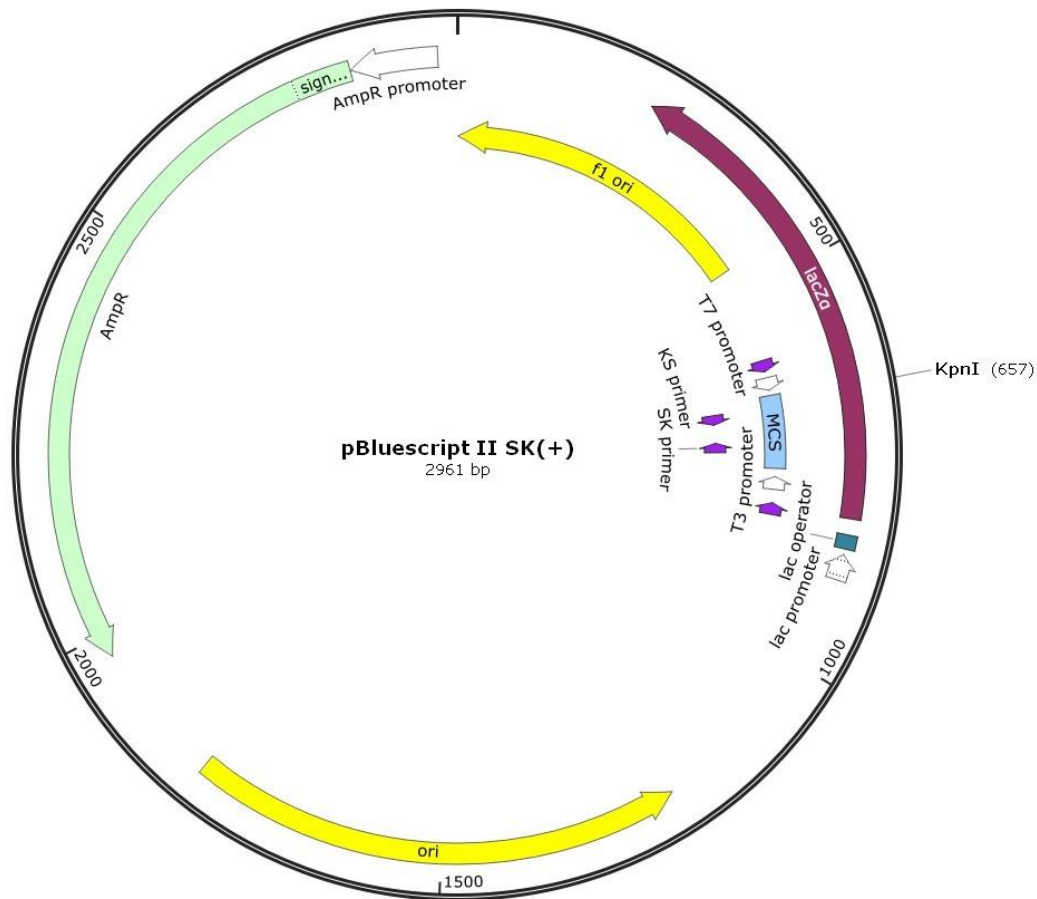
Supplemental figure 1. Schematic representation of the pET28a (+) vector. This vector contains: a gene that confers resistance to kanamycin (*KanR*); a standard origin of replication (*ori*); an origin for replication inside a phage (*f1 ori*); the *ROP* gene (*rop*), which regulates plasmid DNA replication by modulating the processing of the precursor of the replication primer; a multiple cloning site (MCS), where there are several restriction enzymes' recognition sequences (SalI and EcoRI recognition sequences are indicated in the map); a T7 promoter, which is the region where the T7 polymerase links to the vector to activate the transcription of the gene of interest; a *lac* operator, which is where the repressor links to the vector to block the transcription; a T7 terminator, which indicates the end of the transcription; a *lacI* coding sequence and the respective promoter (*lacI* promoter), that originate a repressor that inhibits the transcription of the gene of interest by blocking the binding of the T7 RNA polymerase to the T7 promoter; a hexa-histidine tag (6×His), used to purify the protein of interest using nickel affinity chromatography; a thrombin cleavage site, which is used in the cleavage of the tag regions by thrombin; a T7 tag coding sequence, which is used in immunoprecipitation techniques and Western blot; and a ribosome-binding site (RBS), which is where the ribosomes link for the translation process (SnapGene® Viewer 5.0.7).

Annex III



Supplemental figure 2. Schematic representation of the MU1110 vector. This plasmid contains the activation domain of GAL4 (GAL4 AD) upstream of the TF gene, so that if the TF links to the target gene, it will always activate the expression of the target gene. The expression of the TF is regulated by a 35S promoter, a T7 promoter and a Nuclear Localisation Sequence (NLS), which conducts the TF to the nucleus. This vector also contains a human influenza hemagglutinin tag (HA tag), which is used for purification purposes, a nopaline synthase transcription termination (NOST) signal sequence, downstream of the TF gene, a *STREPTOTHRICIN ACETYLTRANSFERASE2* selection marker gene (*sat2*) and a streptomycin and spectinomycin resistance gene (*aadA1*). The TF sequence present in the MU1110 vector provided by Professor Ueda was the coding sequence of the WRKY2 TF, which is surrounded by the recognition sequences of SalI and NotI restriction enzymes, facilitating the removal of this coding sequence.

Annex IV



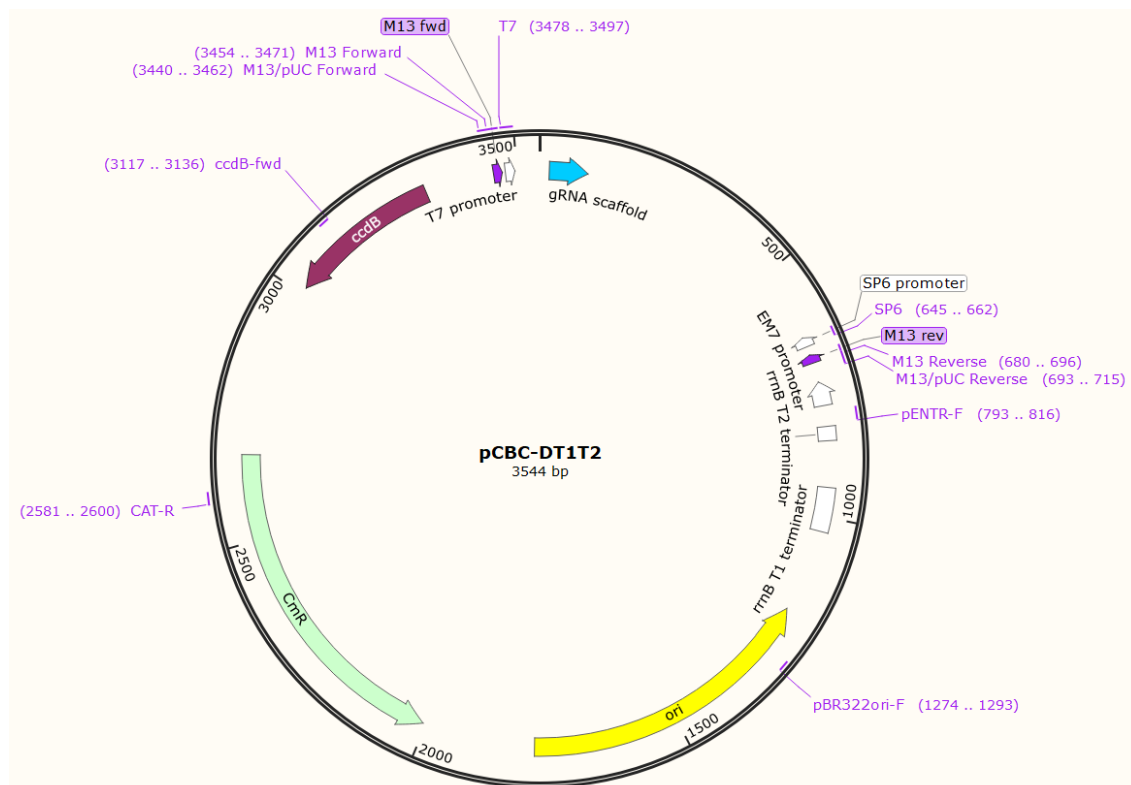
Supplemental figure 3. Schematic representation of the pBluescript II SK (+) vector. This vector contains a gene that confers resistance to ampicillin (*Amp^R*), a standard origin of replication (*ori*), an origin for replication inside a phage (*f1 ori*) and a *lacZα* gene for alpha-complementation, within which there is a T7 and a T3 promoter sequences, for transcription with T7 and T3 RNA polymerases, respectively, surrounding a multiple cloning site (MCS) with a wide variety of restriction enzymes' recognition sequences and where a gene of interest can be cloned into. The *lacZα* gene enables to conclude if the MCS contains an insert. Since *lacZα* codifies for the α subunit of the β-galactosidase enzyme, in case an insert was not cloned inside the MCS and if the culture medium contains IPTG, the IPTG will bind to the *lac* promoter, also present in this vector upstream of the *lacZα* gene, activating the synthesis of the α subunit of the β-galactosidase. This subunit will, consequently, complement the other β-galactosidase subunit produced by the *lacZΔM15*, which, in turn, is present in the genome of certain *E. coli* strains like DH5α, leading to the formation of the β-galactosidase enzyme that will metabolise X-Gal, that must also be present in the medium, conferring a blue colour to the transformed colonies. In the opposite case, if the MCS contains an insert, the *lacZα* will not be transcribed and the transformed colonies will present a white colour. This map also contains the location of the recognition sequences of the KpnI restriction enzyme (SnapGene® Viewer 5.0.7).

Annex V

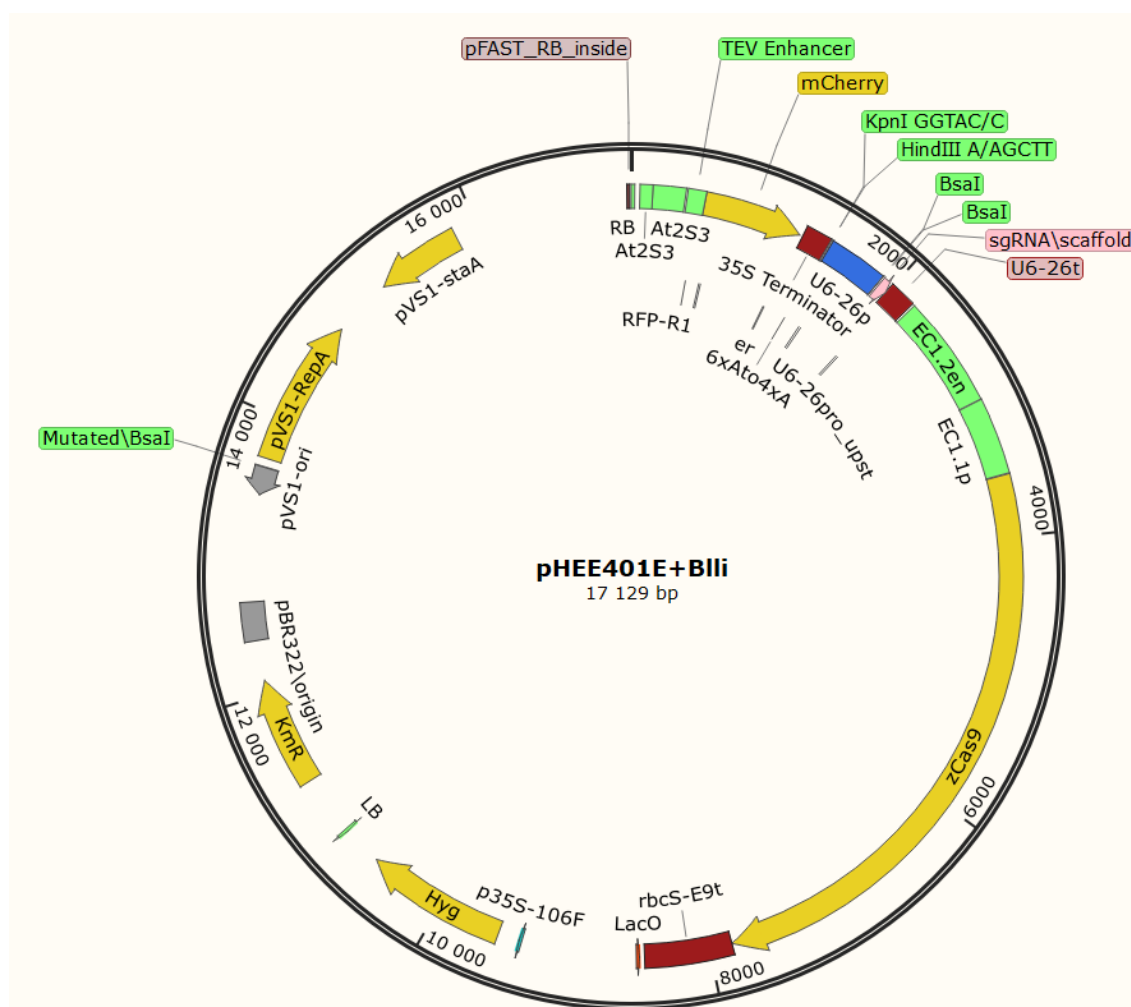
Supplemental table 2. PCR conditions applied in the amplification of the *AGP24* (promoter and 5' UTR) and *AGP31* (promoter, 5' UTR and 3' UTR) predicted *cis* elements.

Step		Temperature (°C)		Time		Expected band size	
10 cycles	Initial denaturation		95		2 minutes		1073 bp [Promoter] 120 bp [5' UTR]
	Denaturation	95		30 seconds			
		51.2 [Promoter] <i>AGP24</i>					
	Annealing	48.4 [5' UTR]					
		47.3 [Promoter]	30 seconds				
		48.3 [5' UTR] <i>AGP31</i>					
		51.7 [3' UTR]					
	Extension	72		1 min 10 sec [Promoter] <i>AGP24</i>			
				15 seconds [5' UTR]			
				1 min 10 sec [Promoter] <i>AGP31</i>			
		15 seconds [5' UTR]					
25 cycles	Denaturation	95		30 seconds		1043 bp [Promoter] 189 bp [5' UTR] 250 bp [3' UTR]	
		Annealing	64.0 [Promoter] <i>AGP24</i>				
	63.0 [5' UTR]						
	57.3 [Promoter]		30 seconds				
	61.5 [5' UTR] <i>AGP31</i>						
	Extension	63.8 [3' UTR]					
				1 min 10 sec [Promoter] <i>AGP24</i>			
				15 seconds [5' UTR]			
				1 min 10 sec [Promoter] <i>AGP31</i>			
			15 seconds [5' UTR]				
			15 seconds [3' UTR]				
	Final extension		72		5 minutes		
	Pause/cooling		8		+ ∞		

Annex VI

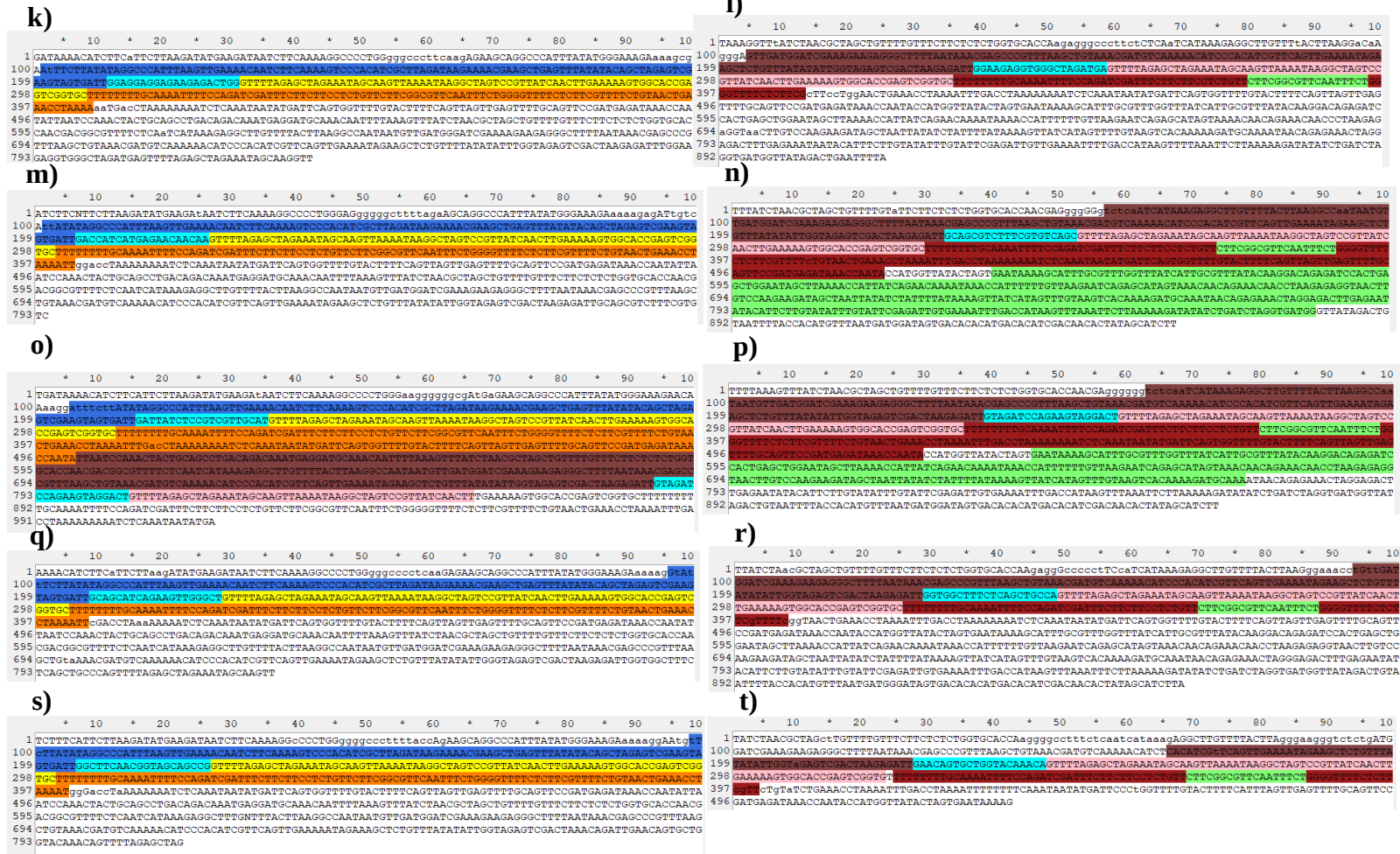


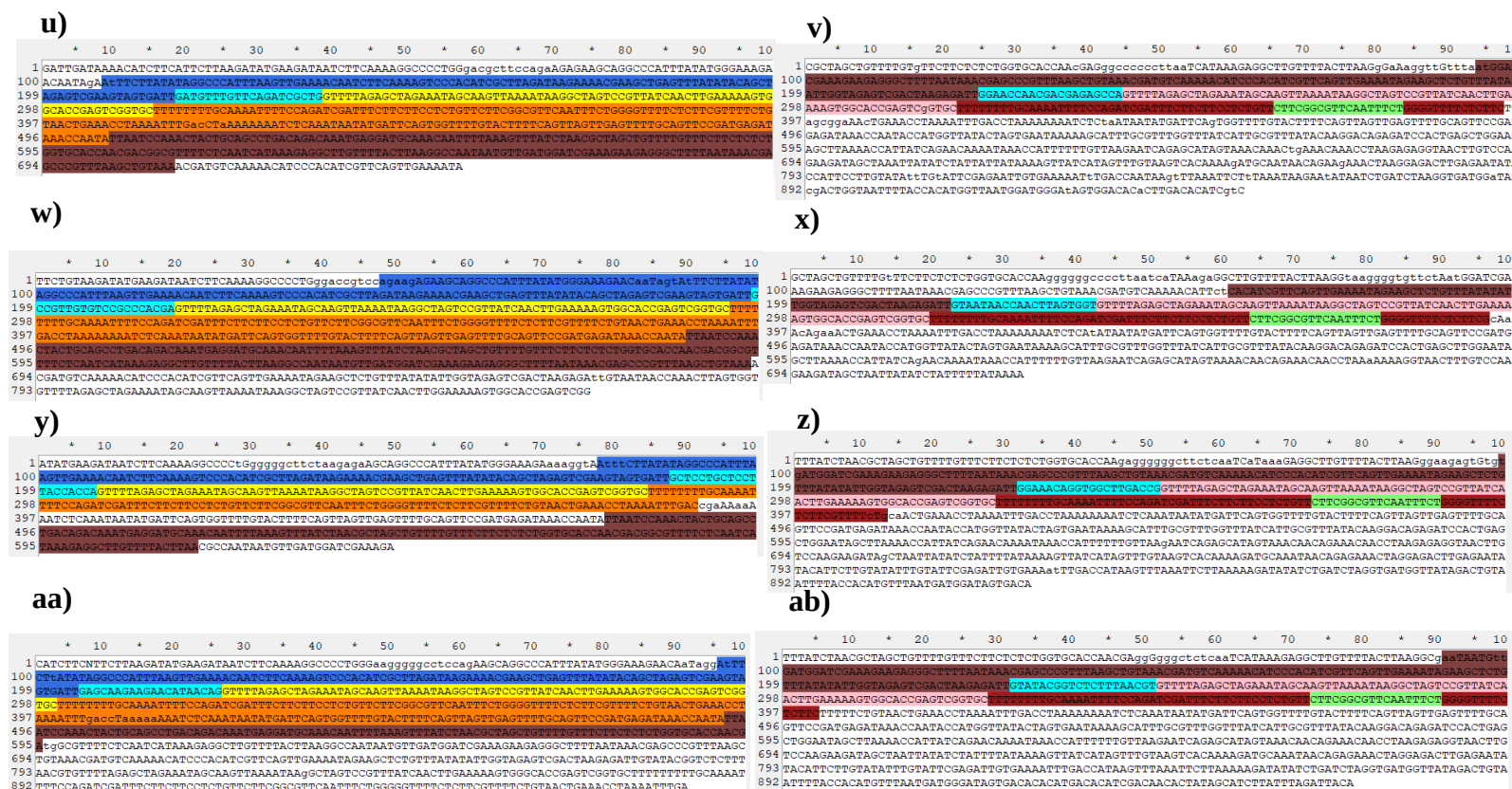
Supplemental figure 4. Schematic representation of the template plasmid pCBC-DT1T2 (Xing *et al.*, 2014). This vector contains a T7 promoter and an origin of replication (ori) among other features. This vector enables the expression of a maximum of 2 sgRNA cassettes (SnapGene® Viewer 5.0.7).



Supplemental figure 5. Schematic representation of the backbone vector pHEE401E+BllI. This plasmid contains two recognition sequences for the BsaI restriction enzyme next to each other, which enables the insertion of the sgRNAs cassettes previously included in the amplified template. This area is surrounded upstream by the Arabidopsis U6-26 small nuclear RNA promoter (U6-26p) and downstream by a sgRNA scaffold and the Arabidopsis U6-26 small nuclear RNA terminator (U6-26t). Next to this region, this plasmid contains the Cas9 gene (zCas9 – *Zea mays* codon-optimised Cas9), whose expression is regulated upstream by the egg cell specific promoter EC1.1 (EC1.1p), which leads to its expression in ovules and one-cell stage embryos, and the egg cell specific enhancer (EC1.2en). Downstream of the Cas9 gene, there is the terminator of the *rbcS-E9* (*Pea sativum* ninth element of the Rubisco small subunit) (rbcS-E9t) (Morelli *et al.*, 1985). The presence of these last three elements increases the CRISPR-Cas9 technique efficiency, since they direct the Cas9 expression to germline cells, improving the genome editing for the generation of non-mosaic mutants. In order to select the transformed seeds, this vector also contains the mCherry gene followed by the 35S terminator. Furthermore, to select the transformed bacteria, this plasmid contains the kanamycin resistance gene (*KmR*) and, to select the transformed plants, it also incorporates the hygromycin resistance gene (Hyg) (Wang *et al.*, 2015) (SnapGene® Viewer 5.0.7).

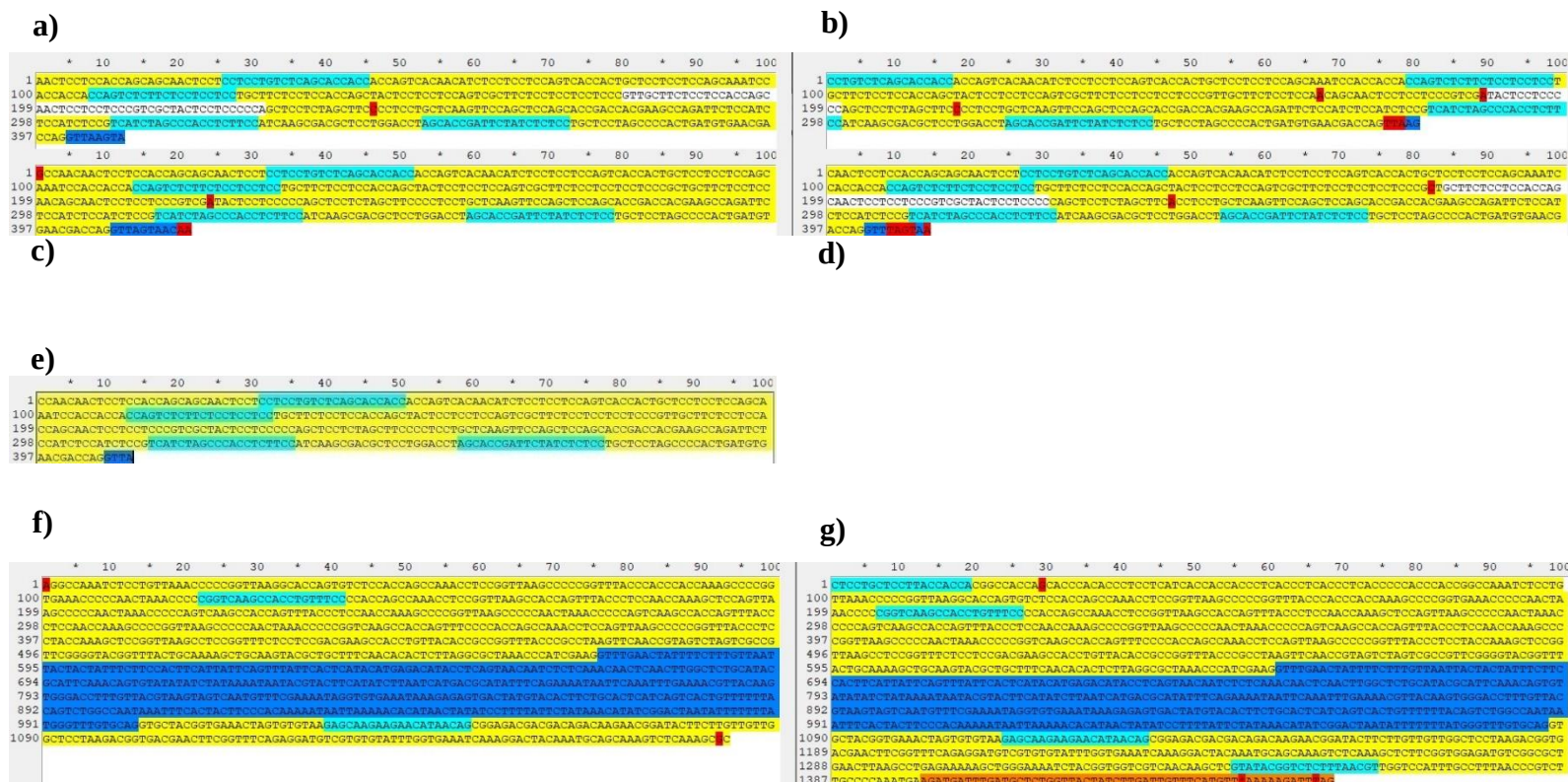
[illegible]





Supplemental figure 6. Results of the sequencing of the plasmid DNA extracted from colonies transformed with the recombinant plasmid transformed with the vectors assembled for the knockout of the *AGP1* (a) – d)), *AGP4* (e) – h)), *AGP9* (i) – l)), *AGP15* (m) – p)), *AGP21* (q) – t)), *AGP24* (u) – x)) and *AGP31* (y) – ab)) genes. a), e), i), m), q), u) and y), region of the vector containing the 1st pair of sgRNAs specific cassettes sequenced with the U6-26p primer; b), f), j), n), r), v), and z), region of the same vector sequenced with the U6-29p primer; c), g), k), o), s), w) and aa), region of the vector containing the 2nd pair of sgRNAs specific cassettes sequenced with the U6-26p primer; d), h), l), p), t), x) and ab), region of the same vector sequenced with the U6-29p primer. Sequences labelled: dark blue, U6-26 promoter; light blue, sgRNAs specific cassettes; yellow, 1st scaffold; orange and dark red, U6-26 terminator; brown, U6-29 promoter; pink, 2nd sgRNA scaffold; 1st green region, U6-26 terminator upstream region; 2nd green region, egg cell specific enhancer; light red, mismatches between the sequencing results and the vectors' predicted sequences; sequences not labelled do not belong to the DNA template, being located in the endings of the sequencing products thus presenting several mismatches in comparison to the predicted sequences.

Annex VIII



Supplemental figure 7. Sequencing of fragments that comprise the sequences homologous to the four sgRNAs specific cassettes designed for the *AGP9* (a) to e) and *AGP31* (f) to g)) genes, amplified using genomic DNA of the plants that germinated following the transformation processes. Only the sequence shown in f) was obtained following the first transformation process. **Yellow**, sequences corresponding to exons of the respective gene; **dark blue**, sequences corresponding to introns of the respective gene; **light blue**, sequences homologous to the four sgRNAs specific cassettes designed for the respective gene; **orange**, sequences corresponding to UTR of the respective gene; **red**, mismatches between the obtained and the published sequences; *, insertion/deletion of a nucleotide in the sequence obtained relatively to the original one; **white**, fragments of the original genes that were not present in the sequences obtained.