

Arabinogalactan proteins: deciphering their function while looking at the sugars

Jessy Ariana Faria da Silva

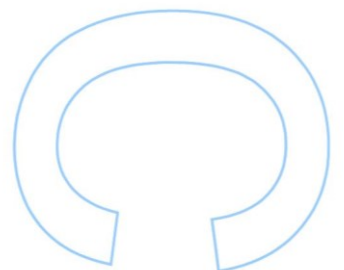
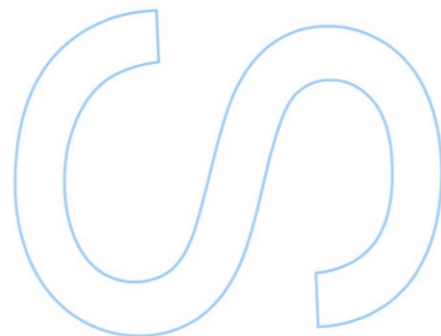
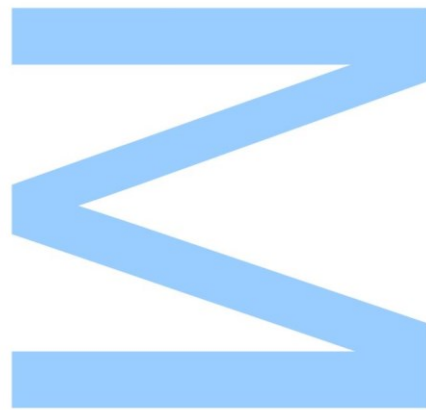
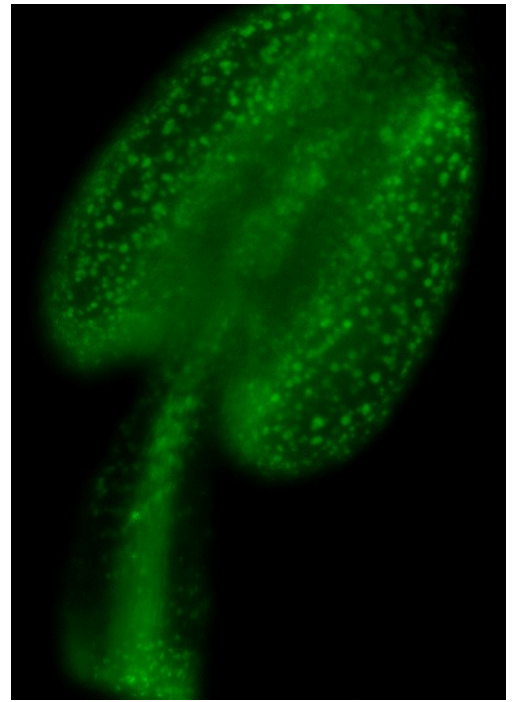
Biologia Funcional e Biotecnologia de Plantas
Departamento de Biologia
2019

Orientador

Sílvia Coimbra, Professor auxiliar, FCUP

Coorientador

Matthew Tucker, Professor auxiliar, AFW, University of Adelaide

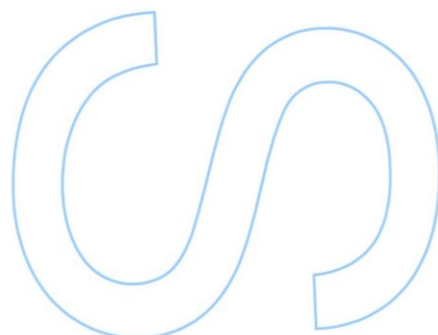
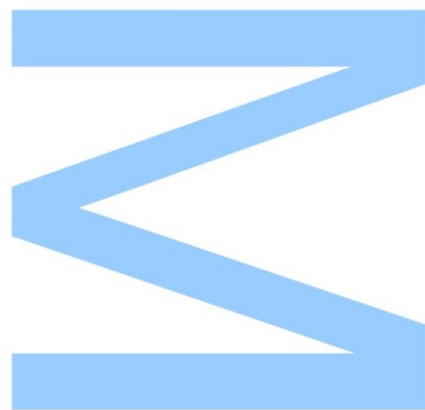




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



*To mum and dad,
who inspired me to pursue my true passion.*

*If you know you are on the right track,
if you have this inner knowledge,
then nobody can turn you off...
no matter what they say.*

Barbara McClintock

ACKNOWLEDGMENTS

"No one who achieves success does so without acknowledging the help of others. The wise and confident acknowledge this help with gratitude."

Alfred North Whitehead

Undertaking this master's has been a truly life-changing experience for me. Throughout this year I faced many challenges that I would not be able to overcome without the help of the wonderful people surrounding me. Here's my chance to express gratitude to those who have supported me.

Firstly, I would like to express my deepest gratitude to my supervisor Prof. Sílvia Coimbra for the continuous support and encouragement throughout my learning process. Thank you for receiving me in your lab and giving me the opportunity to fall in love with "our" mysterious proteins, AGPs. The door to your office was always open whenever I ran into with my enigmas. You allowed this thesis to be my own work but steered me in the right direction whenever you thought I needed it.

My special thanks to my SPReD lab colleagues for helping me grow scientifically and keeping me motivated in discouraging times (with all the available chocolate in the lab... Shhh!).

I would like to express my very great appreciation to my co-supervisor Prof. Matthew Tucker for receiving me at his lab in Adelaide, Australia. Thank you for trusting in my work, for your valuable and insightful comments but also for the hard questions which incited me to widen my research from various perspectives.

I thank my aussie lab mates: Dayton, Chao and Mia for making me feel like home (and taking me home safely!), for introducing me to all the safety stuff, for helping me shopping and for providing KFC to my microscopy nights. Thank you for all the fun you provided me.

My sincere thanks also go to Dr. Neil Shirley for technical assistance of qPCR analysis and for sharing his knowledge and technical know-how.

To Prof. Paul Dupree for providing me the material used in this research and for all the previous work on these amazing enzymes. Assistance provided by Dr. Federico López Hernández was also greatly appreciated. Thank you for always being available to help me with my doubts.

This research work wouldn't have been possible without funding from the SEXSEED project (H2020-MSCA-RISE-2015).

A warm word to my lab mate and friend, Sara, for helping with my first baby steps in science. Thank you for always...I really mean always... being available for me (in any kind of microscopy emergency). Thanks for sharing your (our) workbench and pipettes. Thank you for taking me off the lab to a morning coffee break while my gel solidified in the Erlenmeyer. For all the jokes, tender, philosophical and stupid playful moments that light up my daily life. A heartfelt thank you!

To Sofia and Leo, for the immense support, patience, help and company at lunches and snacks.

Some special words of gratitude to all my Coyote Girls who have always been a major source of support. Thanks for always being there for me and being who you are.

Finally, I must express my very profound gratitude to my family whose love and guidance are with me in whatever dream I pursue. I am what I am today thanks to you. Thank you for believing in me and in my capacities to do this work.

Most importantly, I wish to thank my boyfriend, Miguel, for the unconditional love, strength, patience and unfailing support. Thank you for following me to the other side of the world. For everything and more, thank you!

Thank you all for the encouragement!

RESUMO

Estima-se que a população humana atingirá nove bilhões em 2050. Para atender às necessidades de sustentar esta crescente população, o rendimento agrícola terá que ser duplicado sem aumentar a área de terra arável. O processo de reprodução sexuada das espermatófitas culmina na formação das sementes e estas são unidades essenciais para a propagação das plantas. Assim, para melhorar a produtividade, temos primeiro que entender os mecanismos e os fatores que regulam o complexo processo de formação da semente.

Nas angiospérmicas, a formação da semente envolve a germinação do grão de pólen e o crescimento do tubo polínico através dos tecidos do pistilo. O tubo polínico leva os dois gâmetas masculinos até ao gametófito feminino onde ocorre a dupla fecundação: um gâmeta masculino funde-se com a oosfera, originando o embrião, e o outro gâmeta masculino com a célula central diploide, dando origem ao endosperma triploide.

As proteínas arabinogalactânicas (AGPs) são proteínas altamente glicosiladas envolvidas na reprodução das plantas. As AGPs são extensivamente decoradas por açúcares. 90% das moléculas são constituídas por polissacarídeos arabinogalactânicos compostos por arabinose, galactose e açúcares menores como ácido glucurónico (GlcA), fucose e ramnose. A glicosilação das AGPs é catalisada por várias glicosiltransferases (GTs) na via secretora. Ao longo dos últimos anos, foram realizados vários esforços para caracterizar estas enzimas, mas muitas delas continuam por identificar. Recentemente, foi proposto que as AGPs ligam e armazenam cálcio (Ca^{2+}) através dos resíduos de GlcA na membrana plasmática.

O objetivo global desta dissertação foi tentar decifrar um pouco mais o papel das AGPs durante a dupla fecundação na planta modelo *Arabidopsis thaliana*. Para fazer isso, foram selecionadas duas abordagens. Primeiro, foi realizado um estudo para revelar o papel duma AGP possivelmente envolvida durante este processo: a proteína arabinogalactânica do tipo fasciclina 5 (FLA5). Este trabalho mostrou que *FLA5* é altamente expressa no pólen e por isso pode ter um papel importante no desenvolvimento do gametófito masculino. No entanto, os mutantes *fla5* não mostraram nenhum fenótipo visível provavelmente devido à redundância funcional associada às AGPs. Tanto os gametófitos femininos como os masculinos desenvolveram-se normalmente e, conseqüentemente, a dupla fecundação ocorreu originando as sementes. Apesar disso, a ausência de *FLA5* nos mutantes levou a uma regulação negativa da *FLA14* e uma regulação positiva da *FLA3* e do *SEEDSTICK*.

Em segundo lugar e atendendo que a fração de açúcares das AGPs parece ser crucial para a função das mesmas, foi realizado um estudo sobre um grupo específico de GTs: as glucuronosiltransferases (GlcATs). As GlcATs parecem ter um importante papel em assegurar a presença de GlcA nas AGPs que posteriormente ligarão o Ca^{2+} . Anteriormente, foi demonstrado que cinco das GlcATs transferem GlcA especificamente para as AGPs e as plantas mutantes nestas enzimas mostraram uma redução na substituição de GlcA metilado ([Me] GlcA) em extratos enriquecidos com AGPs. Além disso, os extratos de AGPs de plantas mutantes retêm menos Ca^{2+} do que os das plantas de tipo selvagem *in vitro*. O qPCR mostrou que as GlcATs eram amplamente expressas nos tecidos de *Arabidopsis* mas tinham padrões de expressão distintos. A análise dos promotores usando proteínas amarelas fluorescentes múltiplas (YFP), fundidas com um sinal de localização nuclear (NLS) foi obtida para *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* e *At1g03520* para confirmar o padrão de expressão destas GlcATs. As fusões de proteínas fluorescentes de *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* e *At1g03520* foram também expressas transitoriamente em *Nicotiana benthamiana* e verificou-se que se localizaram no Golgi. Os mutantes de GlcATs foram cultivados em meio deficiente em Ca^{2+} e verificou-se que as plantas possuíam um tamanho reduzido, siliques pequenas, óvulos abortados, pólen não viável e óvulos não fecundados, levando à hipótese de que as GlcATs estão envolvidas no desenvolvimento do pólen e na atração do tubo polínico pelo óvulo. Finalmente, este trabalho realça a importância da porção de açúcares para a função das AGPs, fornecendo uma compreensão mais profunda sobre as AGPs durante a reprodução.

PALAVRAS-CHAVE

Arabidopsis thaliana, Proteínas Arabinogalactânicas, Proteínas Arabinogalactânicas do Tipo Fasciclina, Biossíntese das AGPs, Glucuronosiltransferases, Sinalização pelo Cálcio, Reprodução Sexual.

ABSTRACT

By 2050, the human population is estimated to reach nine billion. To achieve the needs to sustain this growing population, we will have to double agriculture yields without increasing arable land. Seeds are the final product of sexual reproduction process in spermatophytes and are essential units for plant propagation. As agricultural activities rely on this event, in order to improve productivity, we must first understand the mechanisms and the factors regulating the complex process of seed formation.

In angiosperms, seed formation involves a pollen tube growth journey along the female tissues. The pollen tube delivers two sperm cells into the female gametophyte where double fertilization takes place: one sperm cell fertilizes the egg cell, originating the embryo, and the other sperm cell fertilizes the diploid central cell, generating the endosperm.

Arabinogalactan proteins (AGPs) are highly glycosylated proteins deeply involved in plant reproduction. AGPs are extensively decorated by sugars. 90% of the molecules are constituted by arabinogalactan polysaccharides, composed of arabinose, galactose and minor sugars such as glucuronic acid (GlcA), fucose and rhamnose. AGP glycosylation is catalyzed by various glycosyltransferases (GTs) in the secretory pathway. Over the past few years, several efforts have been made to characterize these enzymes but many of them remain to be identified. Recently, it was proposed that AGPs bind and store calcium (Ca^{2+}) through GlcA residues at the plasma membrane.

The overall objective of this thesis was to seek further insights to decipher the role of AGPs during double fertilization in the plant model species *Arabidopsis thaliana*. In order to do that two approaches were selected. Firstly, we performed a study to unravel the role of an AGP possibly involved during this process: fasciclin-like arabinogalactan 5 (FLA5). This work showed that *FLA5* is highly expressed in the pollen and therefore may play a significant role in the male gametophyte development. However, *fla5* mutants showed no visible phenotype probably due to the redundancy associated to AGPs. Both female and male gametophytes developed normally and consequently, double fertilization occurred giving rise to seeds. Nevertheless, the absence of *FLA5* in the mutants led to a downregulation of *FLA14* and an upregulation of *FLA3* and *SEEDSTICK*.

Secondly, as the sugar moiety of AGPs seems crucial for their function, we took a step towards the disclosure of the role of a specific group of GTs: the glucuronosyltransferases (GlcATs). GlcATs appear to have an important mission on

ensuring the presence of GlcA on AGPs for posterior binding to Ca^{2+} . Previously, five GlcATs were shown to transfer GlcA specifically onto AGPs and mutant plants showed a reduction in methyl GlcA ([Me]GlcA) substitution in AGP-enriched preparations. Besides, AGP extracts from mutant plants held less Ca^{2+} than wild-type plants *in vitro*. qPCR showed that GlcATs were widely expressed in Arabidopsis tissues but have distinct patterns of expression. Promoter analysis using multiple yellow fluorescent proteins (YFP) fusions to a nuclear localization signal (NLS) were obtained for *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* to confirm the GlcATs expression patterns. *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* fluorescent protein fusions were transiently expressed in *Nicotiana benthamiana* and found to be localized in the Golgi.

Available T-DNA insertion mutants of GlcATs were grown in Ca^{2+} -deficient and non-deficient mediums. Plants with reduced size, smaller siliques, aborted ovules, non-viable pollen and non-fertilized ovules were found leading to the hypothesis that GlcATs are involved in two processes signaled by Ca^{2+} : pollen development and pollen tube attraction to the ovule.

Finally, this work highlights the importance of the sugar portion for AGPs' function, providing a deeper understanding of AGPs during reproduction.

KEYWORDS

Arabidopsis thaliana, Arabinogalactan Proteins, Fasciclin-Like Arabinogalactan Proteins, AGPs biosynthesis, Glucuronosyltransferases, Calcium Signaling, Sexual Reproduction.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	VII
RESUMO	IX
ABSTRACT.....	XI
LIST OF TABLES	XIX
LIST OF FIGURES	XXI
LIST OF ABBREVIATIONS.....	XXV
CHAPTER 1. GENERAL INFORMATION	1
1.GENERAL INTRODUCTION.....	3
1.1. Feeding the world without destroying the planet	3
1.2. Double Fertilization - an angiosperms' phenomenon	3
1.2.1. Male gametophyte development - ready, steady, go!	3
1.2.2. Female gametophyte development - made to be loved	4
1.2.3. Finding its way through the female tissues.....	5
1.2.4. PT arrival and synergid cell death - the death of the guardians	6
1.2.5. Successful double fertilization - love is a powerful force.....	7
1.3. The plant cell wall - the protector of the realm	8
1.4. Arabinogalactan proteins - all you need is sugar	10
1.4.1 AGPs biosynthesis - the sugar factory	11
1.4.1.1. Glypiation - anchoring.....	11
1.4.1.2. Pro hydroxylation - getting ready	12
1.4.1.3. Glycosylation - becoming sweeter	13
1.4.1.4. AG polysaccharides - sugar link	13
1.4.1.5. AGPs in the extracellular space - facing the challenge	14
1.4.2. AGPs: what do they do and where can they be found?	14
1.5. General objective	15
2. GENERAL METHODOLOGY	21

2.1. Growth <i>in vitro</i>	21
2.1.1. Sterilization of seeds	21
2.1.2. Growth <i>in vitro</i>	21
2.2. Genotyping	21
2.3. Gene expression analysis	22
2.3.1. RNA extraction and quantification	22
2.3.2. cDNA synthesis	22
2.3.3. Quantitative polymerase chain reaction (qPCR)	22
2.4. Bacteria screening	23
2.5. Arabidopsis transformation	23
2.6. Developmental analysis	24
2.6.1. Clearing	24
2.6.2. PT staining with aniline blue	24
2.6.3. Alexander staining	24
2.6.4. Seed set analysis	24
2.7. Statistical analysis	25
2.8. Bioinformatic analysis	25
2.8.1. Sequence and phylogenetic analysis	25
2.8.2. Gene expression data	26
3. REFERENCES	27
CHAPTER 2. UNRAVELLING THE ROLE OF FASCICLIN-LIKE ARABINO GALACTAN PROTEIN 5 IN ARABIDOPSIS REPRODUCTION	37
ABSTRACT	39
1. INTRODUCTION	41
1.1. Fasciclin-like arabinogalactan proteins - cell adhesion proteins	41
1.2. Why FLA5? - the chosen one	43
1.3. Objectives	44
2. MATERIAL AND METHODS	45

2.1. Plant material and growth conditions	45
2.2. Plant marker lines generation	45
2.2.1. DNA amplification and purification	45
2.2.2 Cloning	46
2.2.3. Bacteria transformation	46
2.2.3.1. Preparation of electrocompetent cells	46
2.2.3.2. Electroporation	47
2.2.4. Restriction digestion analysis	47
2.3. <i>In vitro</i> pollen germination assay	48
2.4. Detection of GUS activity	48
2.5. Reciprocal crosses	48
2.6. RNA extraction, cDNA synthesis and qPCR	48
2.7. Co-expression data	49
2.8. Protein and genomic sequences	49
2.9. 3D protein structure	50
3. RESULTS	51
3.1. Phylogenetic analysis of FLAs family	51
3.2. Bioinformatic analysis of FLA5	57
3.3. Expression pattern of <i>FLA5</i>	59
3.4. Isolation and characterization of the <i>FLA5</i> mutant lines	65
4. DISCUSSION	70
4.1. Phylogenetic analysis of FLAs family	70
4.2. Bioinformatic analysis of FLA5	71
4.3. Expression pattern of <i>FLA5</i>	71
4.4. Isolation and characterization of the <i>FLA5</i> mutant lines	72
5. CONCLUSIONS	74
6. REFERENCES	75

CHAPTER 3. IDENTIFICATION OF THE ROLE OF METHYL GLUCURONIC ACID OF ARABINO GALACTAN PROTEINS DURING ARABIDOPSIS SEXUAL REPRODUCTION.....	81
ABSTRACT.....	83
1. INTRODUCTION	85
1.1. Glycosyltransferases - meet the sugar workers	85
1.1.1. Galactosyltransferases - the first ones to arrive.....	89
1.1.2. Glucuronosyltransferases - adding the special sugar	91
1.1.3. Glucuronic acid methyltransferases - changing GlcA	92
1.1.4. Fucosyltransferases - continuing the hard work	92
1.1.5. Rhamnosyltransferases - the ghost workers	93
1.2. Glycoside hydrolases - cutting all the sugar.....	93
1.3. Calcium in sexual reproduction - anywhere, anytime... ..	94
1.3.1. GlcA and Ca^{2+} - a deep connection	94
1.3.2. PT growth and Ca^{2+} - a co-dependent relationship.....	95
1.4. Objectives	96
2. MATERIAL AND METHODS.....	97
2.1. Plant material and growth conditions	97
2.1.1. Plant lines	97
2.1.2. Growth in soil	97
2.1.3. Hydroponic Growth	97
2.2. Constructs generation	99
2.2.1. DNA amplification and purification.....	99
2.2.2. Cloning	100
2.2.3. Bacteria chemical transformation	100
2.2.4. Restriction analysis	100
2.3. Transient expression and subcellular localization in <i>N. benthamiana</i>	101
2.4. Protein and genomic sequences.....	102

3. RESULTS	103
3.1. <i>In silico</i> analysis of GT14 family.....	103
3.2. Expression profile of GT14 genes.....	108
3.3. Plant marker lines.....	112
3.4. Subcellular localization	117
3.5. Functional analysis of glucuronosyltransferase genes	119
4. DISCUSSION	136
4.1. <i>In silico</i> analysis of GT14 family.....	136
4.2. Expression profile of GT14 genes.....	136
4.3. Plant marker lines.....	137
4.4. Subcellular localization	138
4.5. Functional analysis of glucuronosyltransferase genes	138
5. CONCLUSIONS.....	141
6. REFERENCES	143
APPENDIX	153

LIST OF TABLES

Chapter 1 - General Introduction

Table 1 - Information about the AGPs expressed in the reproductive tissues and presumed functions during sexual reproduction in <i>A. thaliana</i>	16
Table 2 - Description of the samples used to extract RNA.	22

Chapter 2 - Unravelling the role of fasciclin-like arabinogalactan protein 5 in Arabidopsis reproduction

Table 1 - List of <i>A. thaliana</i> T-DNA insertion lines for <i>FLA5</i> used.....	44
Table 2 - List of antibiotics used in the constructs generation in <i>E. coli</i> and <i>A. tumefaciens</i>	46
Table 3 - FLAs coding region information in Arabidopsis.....	50
Table 4 - List of the 24 genes with the most similar expression pattern to <i>FLA5</i> , considering only floral tissues.	64

Chapter 3 - Identification of the role of methyl glucuronic acid of arabinogalactan proteins during Arabidopsis sexual reproduction

Table 1 - Information about the characterized transferases involved in the biosynthesis of AGPs glycans in <i>A. thaliana</i>	86
Table 2 - List of <i>A. thaliana</i> T-DNA insertion lines used.	97
Table 3 - Nutrient solutions composition.	99
Table 4 - Enzymes used in the restriction analyses of constructs and expected fragments sizes.	101
Table 5 – Arabidopsis GT14 coding region information.	103
Table 6 - Arabidopsis GT14 predicted domains and sites.	104
Table 7 - Predicted subcellular localization of GT14 proteins in Arabidopsis.	117

Appendix

Supplementary table 1 - List of primers used to genotype T-DNA insertion lines.	152
Supplementary table 2 - List of primers used in gene expression analysis.	153
Supplementary table 3 - List of primers used in <i>FLA5</i> promoter cloning.....	154
Supplementary table 4 - List of primers used in GT14 promoter cloning.	154
Supplementary table 5 - List of primers used in GT14 coding sequence cloning.....	155

LIST OF FIGURES

Chapter 1 - General Introduction

Fig. 1 - Arabidopsis male gametophyte development.....	4
Fig. 2 - Arabidopsis female gametophyte development.....	5
Fig. 3 - Pollen adhesion to the stigma and PT growth in an Arabidopsis pistil.....	6
Fig. 4 - Schematic model of synergid cells degeneration.....	7
Fig. 5 - Arabidopsis double fertilization and seed development.....	8
Fig. 6 - Schematic representation of a primary cell wall.....	9
Fig. 7 - Schematic representation of the predicted structures of the classes of AGPs.....	11
Fig. 8 - GPI-anchored proteins composition.....	12

Chapter 2 - Unravelling the role of fasciclin-like arabinogalactan protein 5 in Arabidopsis reproduction

Fig. 1 - Schematic representation of Arabidopsis FLAs.....	41
Fig. 2 - Expression pattern of <i>pSTK::STK-GFP</i> after fertilization.....	43
Fig. 3 - Schematic representation of Invitrogen™ Gateway Recombination Cloning Technology.....	45
Fig. 4 - Percent of identity matrix between aligned proteins sequences of Arabidopsis FLAs.....	51
Fig. 5 - Phylogenetic tree of Arabidopsis FLAs.....	52
Fig. 6 - FLAs gene positions on <i>A. thaliana</i> chromosomes.....	53
Fig. 7 - Phylogenetic tree of Arabidopsis FLAs and putative homologues in <i>P. patens</i> , <i>S. moellendorffii</i> , <i>H. vulgare</i> and <i>P. trichocarpa</i>	55
Fig. 8 - Schematic representation of FLA5 before and after post-translational modifications.....	57
Fig. 9 - Representation of FLA5 tridimensional structure predicted by the software PyMOL (Schrödinger, 2015).	58
Fig. 10 - Expression pattern of <i>FLA5</i>	59
Fig. 11 - Restriction digestion analyses of <i>pDONR207:FLA5prom</i> with EcoRV enzyme.....	60
Fig. 12 - GUS expression analysis under <i>FLA5</i> promoter in during Arabidopsis flower development.....	61
Fig. 13 - Histochemical localization of GUS activity in transgenic Arabidopsis tissues expressing the <i>FLA5prom:GUS</i> fusion gene.....	62

Fig. 14 - RNA-seq data on <i>FLA5</i> expression (Klepikova et al., 2016).....	63
Fig. 15 - Analysis of <i>FLA5</i> T-DNA insertion lines.....	65
Fig. 16 - Seed set and silique length analysis of Arabidopsis wt and <i>fla5-1</i> mutant siliques.	66
Fig. 17 - <i>FLA5</i> mutant phenotypic characterization.	67
Fig. 18 - Relative expression of <i>FLA5</i> , <i>FLA3</i> , <i>FLA14</i> and <i>STK</i> in wt and <i>fla5-1</i> inflorescences.....	68

Chapter 3 - Identification of the role of methyl glucuronic acid of arabinogalactan proteins during Arabidopsis sexual reproduction

Fig. 1 - AGP type II glycan structure and sites of action of known GTs acting on AGPs of Arabidopsis.....	89
Fig. 2 - Three-dimensional molecular model simulating an Hyp-AG with bound Ca ²⁺ . 95	
Fig. 3 - Representation of the <i>in vitro</i> and hydroponic growth.....	98
Fig. 4 - Percent of identity matrix between aligned GT14 protein sequences from Arabidopsis.....	105
Fig. 5 - Phylogenetic tree of Arabidopsis GT14 proteins sequences.	105
Fig. 6 - GT14 genes distribution on <i>A. thaliana</i> chromosomes.	106
Fig. 7 - Phylogenetic tree of GT14 proteins and putative homologues in <i>P. patens</i> , <i>S. moellendorffii</i> , <i>H. vulgare</i> and <i>P. trichocarpa</i>	107
Fig. 8 - Expression pattern of GT14 genes in different tissues in Arabidopsis.	109
Fig. 9 - Expression pattern of GT14 genes in different tissues in Arabidopsis.	110
Fig. 10 - Expression pattern of GT14 genes in different tissues in Arabidopsis.	111
Fig. 11 - Preparation of plant marker lines with YFP gene under regulation of GT14 promoters.	113
Fig. 12 - <i>GlcAt14Bprom:YFP</i> expression pattern in Arabidopsis.....	114
Fig. 13 - <i>GlcAt14Eprom:YFP</i> expression pattern in Arabidopsis.....	115
Fig. 14 - <i>At4g27480Eprom:YFP</i> expression pattern in Arabidopsis.....	116
Fig. 15 - Preparation of plant marker lines with GT14 genes and GFP gene under regulation of 35S promotor.....	118
Fig. 16 - Subcellular localization of six GT14 proteins in <i>N. benthamiana</i> leaf epidermal cells.	119
Fig. 17 - Analysis of <i>GlcAT14A</i> T-DNA insertion lines.	120
Fig. 18 - Analysis of <i>GlcAT14B</i> T-DNA insertion lines.	121
Fig. 19 - Analysis of <i>GlcAT14D</i> T-DNA insertion lines.....	122
Fig. 20 - Analysis of <i>GlcAT14E</i> T-DNA insertion lines.	123

Fig. 21 - Analysis of <i>At4g27480</i> T-DNA insertion lines.	124
Fig. 22 - Five-week old Arabidopsis wt and mutant plants grown using hydroponic solutions.	126
Fig. 23 - Seed set analysis of Arabidopsis wt and mutant siliques in plants grown using hydroponic solutions.	127
Fig. 24 - Silique length of six-week old Arabidopsis mutant and wt plants grown using hydroponic solutions.	128
Fig. 25 - Percentage of seeds per silique of six-week old Arabidopsis mutant and wt plants grown using hydroponic solutions.	129
Fig. 26 - Cleared ovules of Arabidopsis mutant and wt plants grown using hydroponic solutions.	130
Fig. 27 - Pollen from Arabidopsis mutant and wt plants grown using hydroponic solutions stained with Alexander's staining.	131
Fig. 28 - Aniline blue staining of Arabidopsis mutant and wt pistils of plants grown using hydroponic solutions.	133
Fig. 29 - Percentage of fertilized and unfertilized ovules per pistil of Arabidopsis mutant and wt plants grown using hydroponic solutions.	134
Fig. 31 - Wt ovules stained with aniline blue.	135

Supplementary data

Supplementary fig. 1 - Normalized levels of <i>FLA5</i> gene transcripts (arbitrary units) in a range of tissues.	156
Supplementary fig. 2 - Chromatogram of <i>FLA5</i> amplified fragment by qPCR and purified by HPLC.	156
Supplementary fig. 3 - Alignment of the sequencing results of <i>FLA5</i> amplified fragment by qPCR.	156
Supplementary fig. 4 - Alignment of the sequencing results of <i>FLA5prom:GUS</i> construct.	157
Supplementary fig. 5 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from <i>fla5-1</i> (A) and <i>fla5-2</i> (B) mutants.	158
Supplementary fig. 6 - Tissue-specific absolute expression of GT14 genes in Arabidopsis obtained by qPCR.	159
Supplementary fig. 7 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.	160
Supplementary fig. 8 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.	161

Supplementary fig. 9 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.....	162
Supplementary fig. 10 - Alignment of the sequencing results of <i>At5g39990</i> (A), <i>At5g15050</i> (B), <i>At2g37585</i> (C) and <i>At3g24040</i> (D) amplified fragments by qPCR.	163
Supplementary fig. 11 - Alignment of the sequencing results of <i>At3g15350</i> (A), <i>At4g27480</i> (B), <i>At1g71070</i> (C) and <i>At1g53100</i> (D) amplified fragments by qPCR.	164
Supplementary fig. 12 - Alignment of the sequencing results of <i>At1g03520</i> (A), <i>At3g03690</i> (B) and <i>At4g03340</i> (C) amplified fragments by qPCR.	165
Supplementary fig. 13 - Alignment of the sequencing results of <i>GlcAT14Bprom:YFP</i> (A), <i>GlcAT14Eprom:YFP</i> (B) and <i>At4g27480prom:YFP</i> (C) constructs.	166
Supplementary fig. 14 - Alignment of the sequencing results of <i>At1g71070prom:YFP</i> (A), <i>At1g53100prom:YFP</i> (B) and <i>At1g03520prom:YFP</i> (C) constructs.	167
Supplementary fig. 15 - Alignment of the sequencing results of <i>GlcAT14B:GFP</i> (A), <i>GlcAT14E:GFP</i> (B) and <i>At4g27480:GFP</i> (C) constructs.	168
Supplementary fig. 16 - Alignment of the sequencing results of <i>At1g71070:GFP</i> (A), <i>At1g53100:GFP</i> (B) and <i>At1g03520:GFP</i> (C) constructs.	169
Supplementary fig. 17 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from <i>atglcat14a-2</i> (A) and <i>atglcat14b-1</i> (B) mutants.	170
Supplementary fig. 18 - qPCR analysis to detect the presence of <i>GlcAT14A</i> (A), <i>GlcAT14B</i> (B), <i>GlcAT14D</i> (C), <i>GlcAT14E</i> (D) and <i>At4g27480</i> (E) transcripts in the inflorescences of wt and mutants. The final concentrations of mRNAs are expressed as arbitrary units that represent the numbers of copies per microliter of cDNA, normalized against the geometric means of three control genes that vary the least with respect to each other. Error bars indicate SDs for each mRNA.	171
Supplementary fig. 19 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from <i>atglcat14d-1</i> (A) and <i>atglcat14d-2</i> (B) mutants.	172
Supplementary fig. 20 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from <i>atglcat14e-1</i> (A) and <i>at4g27480-1</i> (B) mutants.	173
Supplementary fig. 21 - Cleared ovules of Arabidopsis mutant and wt plants grown using hydroponic solutions.	174
Supplementary fig. 22 - Pollen from Arabidopsis mutant and wt plants grown using hydroponic solutions stained with Alexander's staining.	175
Supplementary fig. 23 - Aniline blue staining of Arabidopsis mutant and wt pistils of plants grown using hydroponic solutions.	176

LIST OF ABBREVIATIONS

[Ca ²⁺]	Concentration of Calcium
[Me]GlcA	[Methyl]Glucuronic Acid
4-O-Me-GlcA	4-O-methylglucuronosyl
AG	Arabinogalactan
AGP	Arabinogalactan Protein
Ala	Alanine
AmiRNA	Artificial microRNA
Ara	Arabinose
Asn	Asparagine
Asp	Aspartic Acid
BNS	Basal Nutrient Solution
bp	Base Pair
CAZy	Carbohydrate-Active Enzyme Database
cDNA	Complementary DNA
CDS	Coding Sequence
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
Col-0	Columbia-0
CWP	Cell Wall Protein
DIC	Differential Interference Contrast
DNA	Desoxiribonucleic Acid
DNase	Desoxirribonuclease
ENODL	Early Nodulin-like AGP
ER	Endoplasmic Reticulum
EXT	Extensin
FAS1	Fasciclin-like
FLA	Fasciclin-like Arabinogalactan Protein
Fuc	Fucose
FUT	Fucosyltransferase
Gal	Galactose
GalA	Galacturonic Acid
GALT	Galactosyltransferase
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GH	Glycoside Hydrolase
GlcA	Glucuronic Acid
GlcAT	Glucuronosyltransferase/Glucuronic Acid Transferase
Glu	Glutamine
GPI	Glycosylphosphatidylinositol
GT	Glycosyltransferase
GUS	β-glucuronidase
HAE	Hybrid AGP/EXT
HG	Homogalacturonan
His	Histidine

HPLC	High-performance Liquid Chromatography
HRGP	Hydroxyproline-rich Glycoprotein
Hyp	Hydroxyproline
Hyp-O-GalT	Hydroxyproline-O- β -galactosyltransferases
LB	Luria Broth
LCNS	Low Calcium Nutrient Solution
Leu	Leucine
mAu	Mili Absorbance Unit
MEGA X	Molecular Evolutionary Genetics Analysis X
MES	2-(N-Morpholino)ethanesulfonic Acid
MMC	Megaspore Mother Cell
mRNA	Messenger RNA
MS	Murashige Skoog
NASC	Nottingham Arabidopsis Stock Centre
NCBI	National Center for Biotechnology Information
NLS	Nuclear Localization Signal
nsLTP-like	Non-specific Lipid Transfer Protein-like
nsLTP-like AGP	Non-specific Lipid Transfer Protein-like AGP
OD ₆₀₀	Optical Density measured at 600 nm
ORF	Open Reading Frame
p	Probability Value
PACE	Polysaccharide Analysis by Carbohydrate Electrophoresis
PCD	Programmed Cell Death
PCNL	Plastocyanin-like
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric Point
PMC	Pollen Mother Cell
PME	Pectin Methylesterases
Pro	Proline
PRP	Proline-rich Protein
PT	Pollen Tube
qPCR	Quantitative PCR
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
Rha	Rhamnose
RNA	Ribonucleic Acid
RNAi	RNA interference
RNA-seq	RNA-sequencing
SB	Sodium Boric Acid
SD	Standard Deviation
SEM	Standard Error of the Mean
Ser	Serine
SIGnAL	Salk Institute Genomic Analysis Laboratory
T-DNA	Transfer-DNA
TAIR	The Arabidopsis Information Resource

Thr	Threonine
T _m	Melting Temperature
TMD	Transmembrane Domain
TPM	Transcripts Per Million
UTR	Untranslated Region
v/v	Volume per Volume
w/v	Weight per Volume
wt	Wild-type
XGA	Xylogalacturonan
YFP	Yellow Fluorescent Protein

CHAPTER 1

GENERAL INFORMATION

1. GENERAL INTRODUCTION

1.1. Feeding the world without destroying the planet

Our need for food is one of the biggest threats to the environment. By 2050, the human population will increase by two or three billion. To meet the growing worldwide need for food to feed more than nine billion of people, we must double the number of crops we grow until then without expanding agriculture. Therefore, we must “be more efficient about where we grow, what we grow and how we grow”. A balance between meeting society’s growing food needs and sustaining the planet for future generations must be found (Foley, 2014). To increase agricultural production, it is critical to understand the mechanisms involved in sexual reproduction that lead to the formations of seeds.

1.2. Double Fertilization - an angiosperms’ phenomenon

Angiosperms, the flowering plants, are the most diverse and the largest group (nearly 400,000 species) of the spermatophytes, the seed plants (Soltis et al., 2018). Double fertilization is an exclusive process of angiosperms which initiates the development of a new seed and the establishment of a new plant generation (Pereira et al., 2016a).

During evolution, plants developed a life cycle that alternates between a diploid sporophyte generation, achieved by double fertilization, and a haploid gametophyte generation accomplished by meiotic division (Drews and Yadegari, 2002; Yadegari and Drews, 2004). The sporophyte generates two types of spores: microspores and megaspores, that give rise to male and female gametophytes, respectively (Yadegari and Drews, 2004). During double fertilization, two male sperm cells are delivered by a pollen tube (PT) to the female gametophyte where one fuses with the egg cell and the other fuses with the central cell, giving rise to the embryo and the endosperm, respectively (Berger et al., 2008; Pereira et al., 2015).

1.2.1. Male gametophyte development - ready, steady, go!

The angiosperm male gametophyte, or pollen grain, is a three-celled organism formed in the anthers locules (McCormick, 2004). The anther is composed by four layers: epidermis, endothecium, middle layer and tapetum (Bedinger, 1992). Pollen development can be divided into two phases: microsporogenesis and microgametogenesis (Borg et al., 2009). During microsporogenesis, a diploid pollen mother cell (PMC) undergoes meiosis to produce a tetrad of haploid microspores enclosed in an impermeable callose wall (Bedinger, 1992; McCormick, 2004; Borg and

Twell, 2010). The unicellular microspores in each tetrad are released by the action of callase, an enzyme secreted by the tapetum. The tapetum is a nutritive layer that lines the locule containing the developing microspores and that disintegrates in the later stages of pollen development (McCormick, 2004; Scott et al., 2004). During microgametogenesis, the released uninucleate microspores enlarge and undergo an asymmetric mitosis originating two cells (bicellular pollen grain): the vegetative cell and the generative cell (McCormick, 2004; Borg and Twell, 2010). The larger vegetative cell, which will develop into a PT, exits the cell cycle, whereas the small generative cell, which is engulfed within the cytoplasm of the vegetative cell, undergoes a further mitotic division to form two sperm cells: the male gametes (Fig. 1) (McCormick, 2004; Borg and Twell, 2010).

The synthesis of the pollen wall starts after meiosis. The mature pollen wall is formed by two layers: an outer layer named exine composed of sporopollenin and an inner layer termed intine which contains pectocellulose (Bedinger, 1992; McCormick, 1993). The origin of the two layers is different: the sporopollenin is synthesized predominantly by the tapetum (sporophytic origin) while the intine is secreted by the microspores (gametophytic origin) (Owen and Makaroff, 1995; Vizcay-Barrena and Wilson, 2006).

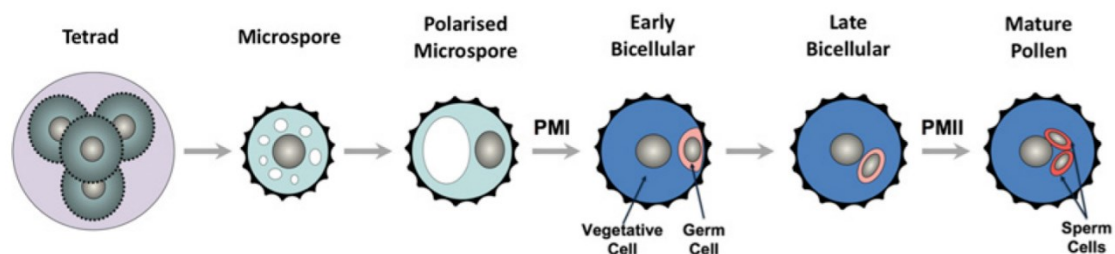


Fig. 1 - Arabidopsis male gametophyte development.

A diploid pollen mother cell (PMC) undergoes meiotic division to produce a tetrad of haploid microspores. The released microspores enlarge and undergo an asymmetric mitosis (PMI) to produce a bicellular pollen grain with a small generative cell engulfed within the cytoplasm of a large vegetative cell. The vegetative cell exits the cell cycle and the generative cell undergoes a further mitotic division (PMII) to form two sperm cells. Abbreviations: PMI, pollen mitosis I; PMII, pollen mitosis II. Taken from Borg and Twell (2010).

1.2.2. Female gametophyte development - made to be loved

The angiosperm female gametophyte development occurs in the ovule and can be divided into two phases: megasporogenesis and megagametogenesis (Yadegari and Drews, 2004). Early in ovule development, the nucellus is composed of two layers: an epidermal (L1) and a subepidermal (L2) layers. During megasporogenesis, from the L2 layer cells, a diploid archesporial cell is selected (Pinto et al., 2019). The archesporial

cell expands and differentiates into a megaspore mother cell (MMC) which undergoes meiosis and gives rise to four haploid megaspores (Yadegari and Drews, 2004; Pereira et al., 2016a; Pinto et al., 2019). One of these megaspores will survive and become the functional megaspore while the remaining three megaspores will undergo programmed cell death (PCD) (Pereira et al., 2016a). During megagametogenesis, the functional megaspore after three rounds of mitosis and cellularization gives rise to the mature female gametophyte, the embryo sac (Yadegari and Drews, 2004). The embryo sac is a seven celled, eight-nucleated structure, composed by: three antipodal cells positioned in the chalazal pole of the ovule, one diploid central cell and two synergid cells surrounding the egg cell, at the micropylar pole of the ovule. The central cell and the egg cell are the two female gametes (Fig. 2) (Drews et al., 1998; Pereira et al., 2016a).

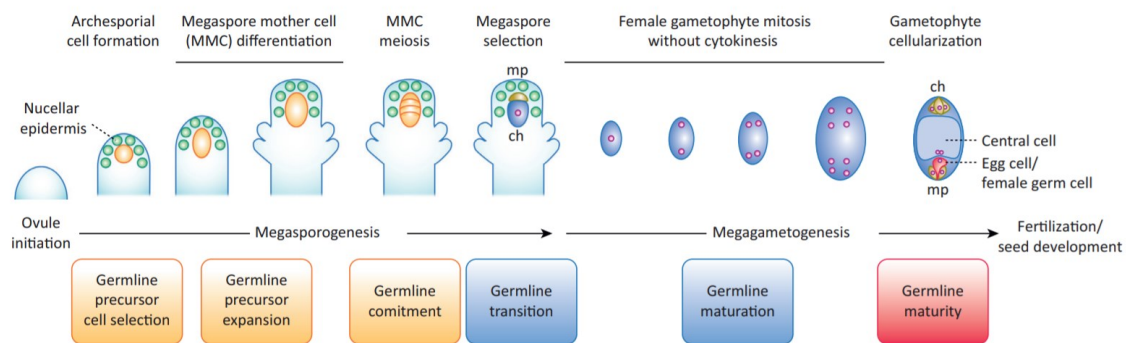


Fig. 2 - Arabidopsis female gametophyte development.

After ovule initiation, a germline precursor cell is selected to form an archesporial cell in the nucellar subepidermal cells. The archesporial cell expands and differentiates into the megaspore mother cell (MMC). The diploid MMC undergoes meiotic division (germline commitment) and gives rise to four haploid megaspores from which: one will become the functional megaspore (germline transition to a haploid phase) while the remaining three undergo programmed cell death (PCD). The functional megaspore after three rounds of mitosis gives rise to the mature female gametophyte (germline maturation). The female gametophyte is formed by a seven celled, eight-nucleated structure, composed by: three antipodal cells positioned in the chalazal pole of the ovule, one diploid central cell and two synergid cells surrounding the egg cell, at the micropylar pole of the ovule (germline maturity). Abbreviations: ch, chalaza; mp, micropyle. Retrieved from Pinto et al. (2019).

1.2.3. Finding its way through the female tissues

The first interaction of the pollen with the pistil is the physical adhesion of the pollen grain to the papilla cells of the stigma (Dresselhaus and Franklin-Tong, 2013). The mature pollen grains are released from the anthers in a partially dehydrated state (McCormick, 2004). Pollen from different sources such as anthers, wind or animal pollinators is captured by the stigma (Palanivelu and Tsukamoto, 2012; Dresselhaus and Franklin-Tong, 2013). When in contact with the stigmatic cells, the pollen coat is extruded of the exine against the stigma surface, forming a 'foot' that strongly sticks to the stigma (Fig.

3) (Pereira et al., 2016a). The adhesion of pollen to the stigma is very strong, selective and rapidly established in order to promote a successful fertilization with the appropriate pollen (Palanivelu and Tsukamoto, 2012). Then, the stigmatic cells provide water and the pollen grains hydrate (McCormick, 2004; Dresselhaus and Franklin-Tong, 2013). The vegetative cell extends a tube that grows via tip growth which is responsible for carrying the two sperm cells into the embryo sac (McCormick, 2004). When PTs reach the style, PT guidance signals are provided by the extracellular matrix (Dresselhaus and Franklin-Tong, 2013). PTs grow through the transmitting tract, a tissue rich in polysaccharides, glycoproteins and glycolipids, which are proposed to nourish, guide and support PT development (Crawford and Yanofsky, 2008).

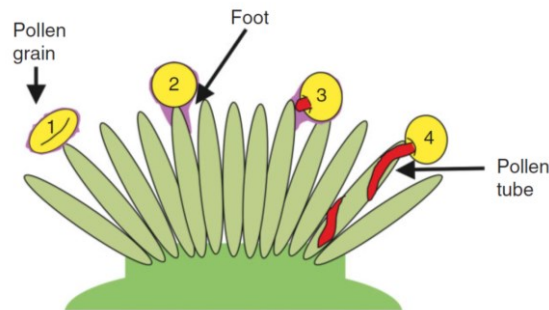


Fig. 3 - Pollen adhesion to the stigma and pollen tube (PT) growth in an *Arabidopsis* pistil.

(1) Dehydrated pollen grain (yellow) adhering to the stigma (green). (2) Pollen coat 'foot formation' (pink area). (3) PT (red) germination and (4) PT growth toward the style. Taken from Palanivelu and Tsukamoto (2012).

1.2.4. PT arrival and synergid cell death - the death of the guardians

The PT tip grows along the transmitting tract tissues until it perceives signals, such as AtLURE1 peptides secreted by the synergid cells, and consequently turns abruptly in the direction of an ovule (Pereira et al., 2016a; Takeuchi and Higashiyama, 2016). The PT slows down, turns into the placenta tissues and enters the funiculus until it reaches the micropyle: the embryo sac entrance (Fig. 4A) (Dresselhaus and Franklin-Tong, 2013; Pereira et al., 2016a). Then, the PT enters through the filiform apparatus of one of the two synergids ceasing its growth and bursting inside it to release the two sperm cells into the chalazal pole of the receptive synergid at the boundary of both female gametes (Pereira et al., 2016a). The synergid cell penetrated by the PT undergoes PCD (Fig. 4B) and the persisting synergid will undergo synergid-endosperm fusion after successful double fertilization (Fig. 4D) (Dresselhaus and Franklin-Tong, 2013; Maruyama et al., 2015; Pereira et al., 2016a). When the sperm cells arrive to the embryo sac, PT development is complete (McCormick, 2004).

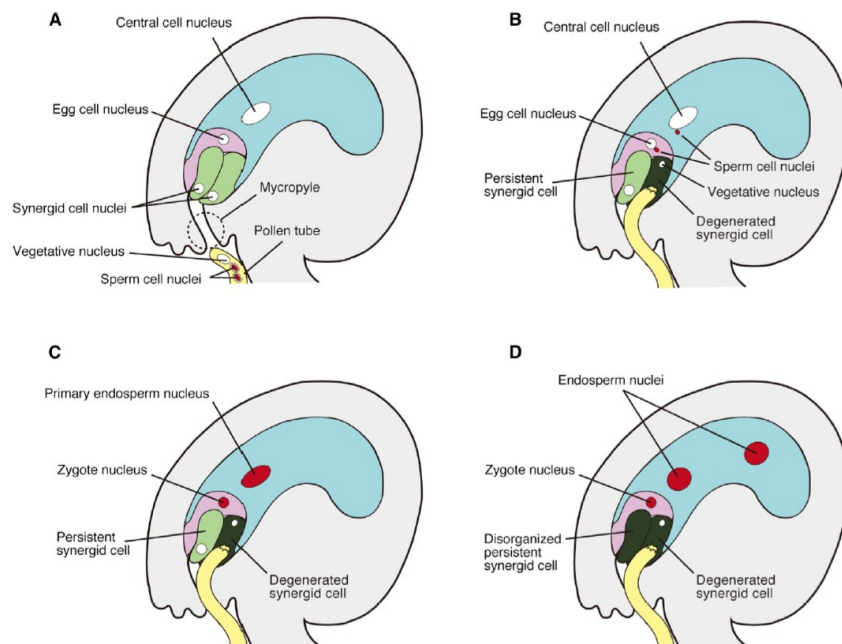


Fig. 4 - Schematic model of synergid cells degeneration.

(A) The pollen tube (PT) is attracted to the micropyle by the signals secreted by the synergids. (B) The PT enters one of the synergids and discharges the sperm cells. The synergid undergoes programmed cell death (PCD) - degenerated synergid cell. (C) Double fertilization takes place: one sperm cell fuses with the egg cell, originating the zygote, while the other fuses with the central cell, giving rise to the endosperm. (D) Degeneration of the persistent synergid cell after double fertilization. In the schematic, antipodal cells were omitted. Taken from Maruyama et al. (2015).

1.2.5. Successful double fertilization - love is a powerful force

During PCD, the cytoskeleton of the synergid reorganizes and forms two actin bands, which are thought to help the migration of the sperm cells towards the boundary between both female gametes (Kasahara et al., 2005; Dresselhaus and Franklin-Tong, 2013). One sperm cell migrates to the egg cell and the other to the central cell, specific recognition signals are likely to take place and the plasma membranes of the sperm cells fuse (plasmogamy) almost simultaneously (Fig. 4B). Then, the sperm cell nuclei fuse with the female gamete nuclei (karyogamy; Fig. 4C) (Dresselhaus and Franklin-Tong, 2013). As a result, when double fertilization takes place, one sperm cell fertilizes the egg cell, giving rise to the embryo, which is the beginning of the sporophyte generation, and the second sperm cell fuses with the diploid central cell, giving rise to the triploid endosperm, which protects and provides nutrients to the developing embryo (Pereira et al., 2016a). Once successful double fertilization has occurred, different barriers are established to prevent the attraction of extra PTs and to prevent the fusion of multiple sperm cells with single gametes (polyspermy) (Beale and Johnson, 2013; Dresselhaus and Franklin-Tong, 2013). The fertilized ovules develop into a seed and the pistil matures

into a fruit (Palanivelu and Tsukamoto, 2012). The effective development of a seed involves the coordinative development of three structures: the embryo, the endosperm and the seed coat. The mature seed is covered by a single layer of endosperm cells surrounding the embryo (Figueiredo and Köhler, 2018). The endosperm contains two maternal and one paternal genome copies (Batista et al., 2019). Both fertilization products are covered and protected by the seed coat derived from the ovule integuments (maternal origin) (Yadegari and Drews, 2004; Figueiredo and Köhler, 2018).

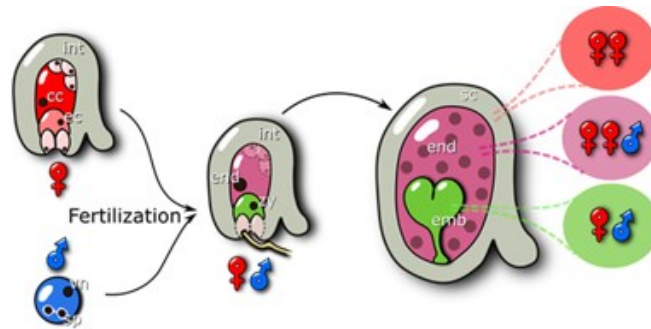


Fig. 5 - Arabidopsis double fertilization and seed development.

Haploid gametes are produced by meiosis. When double fertilization occurs, the female gametes fuse with the male gametes, originating a diploid embryo (green) and a triploid endosperm (pink), which are surrounded by the seed coat (red). Abbreviations: cc, central cell; ec, egg cell; emb, embryo; end, endosperm; int, integuments; sc, seed coat; sp, sperm cell; vn, vegetative nucleus; zy, zygote. Taken from <https://figueiredolab.com/research/>.

1.3. The plant cell wall - the protector of the realm

Plant growth and development are essential processes that crucially rest on the action and assembly of the cell wall (Basu et al., 2016). The cell wall plays different key roles in plant growth and development, cell differentiation, environmental sensing and signaling, intercellular communication, water movement and defense against invading pathogens (Cosgrove, 2005; Jamet et al., 2008). Plant cell walls are classified as primary walls when surrounding growing cells and as secondary walls whenever containing lignin enclosing cells that have ceased growing (Keegstra, 2010; Cosgrove and Jarvis, 2012). The plant cell wall is a dynamic and complex structure composed by different components that all together interact and contribute to cell wall architecture and function (Somerville et al., 2004). The cell wall is a thin (0.1-1µm) and flexible layer constituted by three major classes of polysaccharides: cellulose, hemicelluloses and pectins; and structural proteins (Fig. 6) (Cosgrove, 2005; Cosgrove and Jarvis, 2012).

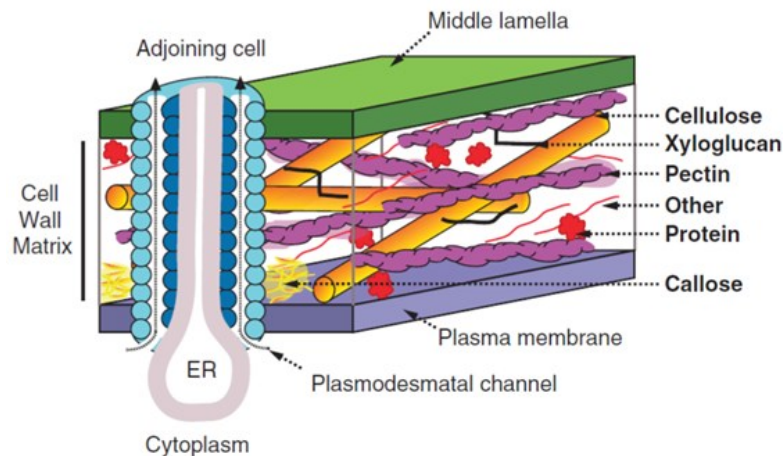


Fig. 6 - Schematic representation of a primary cell wall.

The cell wall comprises various polysaccharides (cellulose, xyloglucan, pectin and callose) and proteins. In the figure, a plasmodesmata that forms a channel between adjoining cells is represented. Adapted from Tucker and Koltunow (2014).

Cellulose is a linear homopolysaccharide composed of β -D-glucose units linked by (1 $\rightarrow$ 4) glycosidic bonds synthesized by a cellulose synthase in the plasma membrane (O'Sullivan, 1997; Höfte and Voxeur, 2017). Cellulose is the most abundant organic polymer on Earth (Klemm et al., 2005) and constitutes 20-30% of the dry weight of the cell wall (Varner and Lin, 1989).

Hemicelluloses are branched polysaccharides that have β -(1 $\rightarrow$ 4)-linked glucose residues (Scheller and Ulvskov, 2010). Hemicelluloses associate with cellulose microfibrils providing a cross-linked matrix (Lerouxel et al., 2006). Hemicelluloses include xyloglucans, xylans, glucuronoxylans, arabinoxylans, mannans, glucomannans and β -(1 $\rightarrow$ 3,1 $\rightarrow$ 4)-glucans (Scheller and Ulvskov, 2010; Zhou et al., 2017). Hemicelluloses are synthesized in the Golgi by glycosyltransferases (GTs) (EC2.4.x.y). These polysaccharides highly contribute to strengthening the cell wall by interacting with cellulose and lignin (Scheller and Ulvskov, 2010).

Pectins are wall complex polysaccharides that are rich in galacturonic acid (GalA) (Willats et al., 2001; Atmodjo et al., 2011). Pectins include homogalacturonan (HG), rhamnogalacturonan I (RGI), rhamnogalacturonan II (RGII) and xylogalacturonan (XGA) (Atmodjo et al., 2011). HG is a linear homopolymer of α -(1 $\rightarrow$ 4)-linked GalA residues which may be methylesterified, acetylated or xylosylated (Willats et al., 2001; Höfte and Voxeur, 2017). GalA residues of HG can be substituted by xylose originating XGA (Willats et al., 2001). RGI contains a backbone of the disaccharide repeat [α -D-GalA-1,2- α -L-Rha-1-4-] $_n$. RGII consists of an HG backbone further decorated with 12 different types of sugars. Sixty seven transferases are required to synthesize pectin (Mohnen, 2008). Pectins control the porosity of the cell wall and are the major adhesive material

between cells (Willats et al., 2001). Cell wall proteins (CWPs) include expansins, xyloglucan endotransglucosylase/hydrolases, endo-(1,4)- β -D-glucanases, involved in wall remodeling and loosening (Cosgrove, 2005) and hydroxyproline-rich glycoproteins (HRGPs) (Showalter, 1993). The HRGPs superfamily is divided into three subgroups: minimally glycosylated proline-rich proteins (PRPs), moderately glycosylated extensins (EXTs) and highly glycosylated arabinogalactan proteins (AGPs) (Hijazi et al., 2014; Johnson et al., 2017). In 2010, Showalter and colleagues conducted a bioinformatics approach to identify HRGPs based on their protein backbones in the Arabidopsis genome. This analysis allowed the identification of 166 HRGPs: 85 AGPs, 59 EXTs, 18 PRPs and four hybrid AGP/EXTs (HAEs) (Showalter et al., 2010). AGP/EXTs contain motifs characteristic of AGPs and EXTs (Showalter et al., 2010; Su and Higashiyama, 2018).

1.4. Arabinogalactan proteins - all you need is sugar

AGPs constitute a highly diverse class of heavily glycosylated proteins that consist of a core-protein backbone surrounded by carbohydrates. Many of these proteins are predicted to be tethered to the plasma membrane by a glycosylphosphatidylinositol (GPI) lipid anchor (Seifert and Roberts, 2007).

According to their polypeptide core based on the presence/absence of particular motif/domains, AGPs can be classified into classical AGPs, AG peptides, lysine-rich AGPs, fasciclin-like AGPs (FLAs), early nodulin-like AGPs (ENODLs) and non-specific lipid transfer protein-like AGPs (nsLTP-like AGPs) (Gaspar et al., 2001; Schultz et al., 2002; Mashiguchi et al., 2009; Showalter et al., 2010). Classical AGPs are characterized by the presence of an N-terminal secretion signal peptide, a central domain rich in proline (Pro), alanine (Ala), serine (Ser) and threonine (Thr), and a hydrophobic C-terminal signal sequence that encodes for the addition of a GPI anchor (Du et al., 1996). AG peptides are short classical AGPs with only 10 to 13 amino acids (Schultz et al., 2000). The other classes are structurally similar to the classical AGPs but have different domains in their polypeptide core such as lysine-rich, fasciclin-like (FAS1), plastocyanin-like (PCNL) and nsLTP domains (Schultz et al., 2002; Johnson et al., 2003; Pereira et al., 2015; Hou et al., 2016).

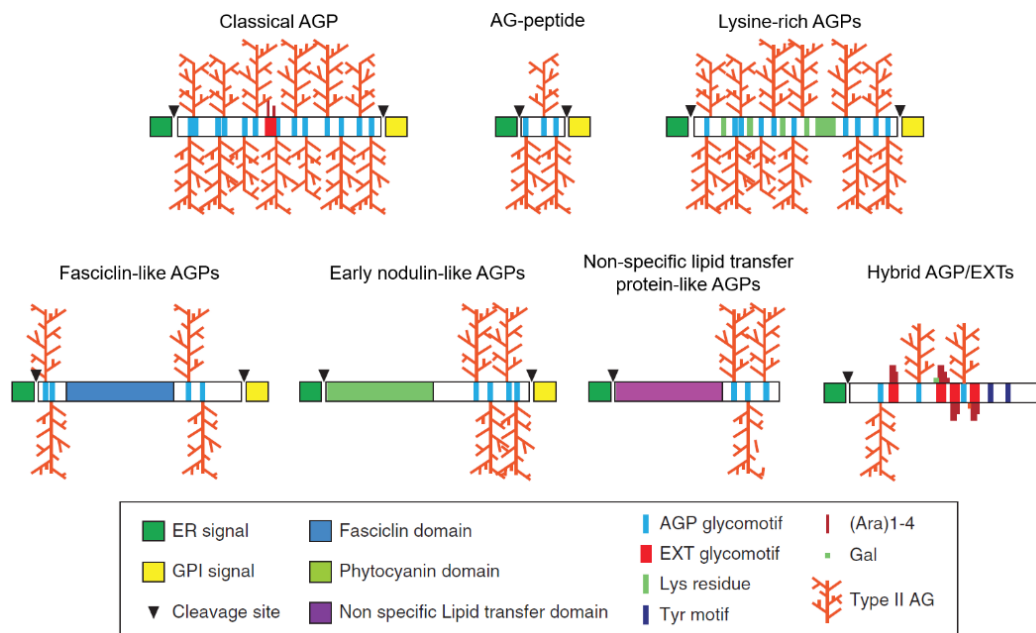


Fig. 7 - Schematic representation of the predicted structures of the classes of AGPs.

The endoplasmic reticulum (ER) (green) and glycosylphosphatidylinositol (GPI) (yellow) signals are cleaved (arrows) in the mature protein form. The protein motifs that direct hydroxylation of proline (Pro) are signaled as light blue bars (AGP glycomotif) where type II arabinogalactan polysaccharide (AG) (orange) chains are added. Arabinogalactan proteins (AGPs) can have atypical domains such as fasciclin-like (FAS1) (blue), phytocyanin (light green) or non-specific lipid transfer protein (nsLTP) (purple). Hybrid AGP/EXTs contain motifs characteristic of AGPs (light blue bars), extensins (EXTs) (red bar) and tyrosine (Tyr) motifs (dark blue bar). The EXT glycomotif directs the addition of short arabinose (Ara) side chains (dark red) on hydroxyproline (Hyp) residues and galactose (Gal) (green) on serine (Ser) residues. Adapted from Ma et al. (2018).

1.4.1 AGPs biosynthesis - the sugar factory

1.4.1.1. Glypiation - anchoring

The N-terminal cleavable signal sequence, recognized by the signal recognition particle, is proteolytically removed, directing AGPs into the endoplasmic reticulum (ER) where post-translational processing begins (Walter and Johnson, 1994; Schatz and Dobberstein, 1996; Ma et al., 2018). The addition of GPI-anchors, also named glypiation, is the first post-translational modification of AGPs attached to the plasma membrane (Duruflé et al., 2017). In Arabidopsis, 55 of the 85 identified AGPs are predicted to contain GPI-anchor motif (Showalter et al., 2010).

The C-terminal signal sequence contains small aliphatic amino acids designated ω , $\omega+1$, and $\omega+2$, followed by a short spacer and terminates in a hydrophobic domain (Schultz et al., 1998; Gaspar et al., 2001; Ellis et al., 2010). The GPI moiety is synthesized on the cytoplasmic surface of the ER via the sequential addition of glucosamine, three α -linked mannose residues, phosphoethanolamine to phosphatidylinositol (Fig. 8) (Schultz et al.,

1998; Yeats et al., 2018). The hydrophobic C-terminal peptide is proteolytically cleaved between the ω and $\omega+1$ sites and the GPI anchor is attached to the ω amino acid by a transamidation reaction (Schultz et al., 1998; Ellis et al., 2010).

Lalanne et al. (2004) demonstrated that mutants of two homologs of phosphatidylinositol N-acetylglucosaminyltransferase subunit C (PIG-C) and A (PIG-A), components of GPI-N-acetylglucosaminyltransferase (GPI-GnT) complex responsible for the first step of GPI biosynthetic pathway, *SETH1* (At2g34980) and *SETH2* (At3g45100), respectively, might have a role on PT wall deposition or metabolism. *seth1* and *seth2* specifically block male transmission and pollen function in Arabidopsis, likely associated with abnormal callose deposition.

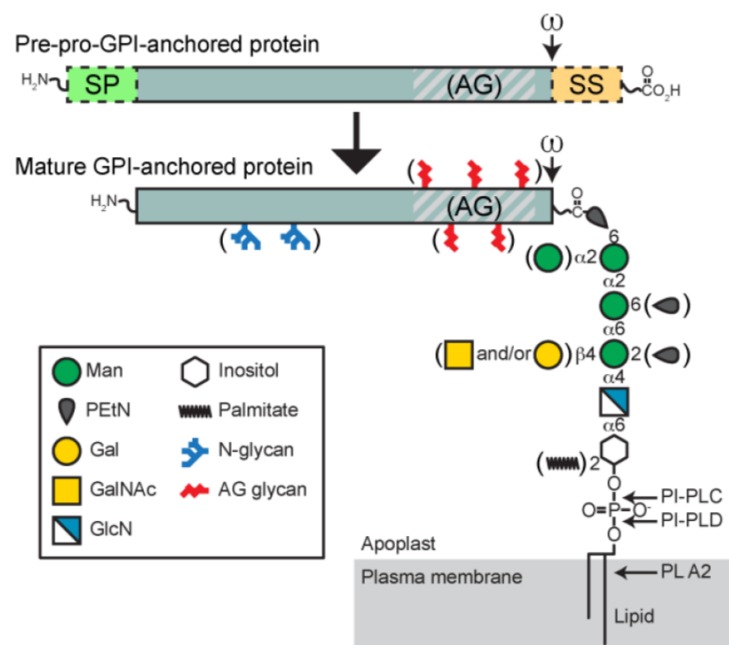


Fig. 8 - GPI-anchored proteins composition.

The pre-pro-GPI-anchored protein includes a N-terminal secretory signal peptide (SP) and a C-terminal GPI-specifying hydrophobic signal sequence (SS) that are cleaved in the endoplasmic reticulum (ER) during maturation. Abbreviations: Man, mannose; PEtN, Phosphoethanolamine; Gal, galactose; GalNAc, N-acetyl galactosamine; GlcN, glucosamine; PI-PLC, phosphatidylinositol-specific phospholipase C; PI-PLD, phosphatidylinositol-specific phospholipase D; PLA2, Phospholipase A2. Retrieved from Yeats et al. (2018).

1.4.1.2. Pro hydroxylation - getting ready

GPI-anchored proteins are then transferred to the Golgi apparatus where they undergo glycosylation (Ellis et al., 2010). Pro hydroxylation is often a prerequisite for O-glycosylation of CWP on hydroxyproline (Hyp) residues (Faye et al., 2005; Duruflé et al., 2017). Selected Pro residues are converted to Hyp residues by prolyl-4-hydroxylases (P4Hs) in the ER (Gaspar et al., 2001; Faye et al., 2005). To date, it is still difficult to

predict with confidence which Pro residues will be hydroxylated. The extended Pro hydroxylation code suggests that Pro residues could be hydroxylated when they are located after Ala, glutamine, Hyp, Pro, Ser, Thr, and valine residues, whereas the first Pro residue following the other amino acids could not be hydroxylated (Duruflé et al., 2017).

1.4.1.3. Glycosylation - becoming sweeter

Glycosylation is one of the major post-translational modifications found in almost every living organism. It is performed by GTs and includes N-glycosylation, O-glycosylation and glypiation (Hurtado-Guerrero and Davies, 2012; Duruflé et al., 2017). N-glycosylation occurs on asparagine (Asn) residues in Asn-X-Ser/Thr specific sequences, where X can be any amino acid except Pro (Faye et al., 2005; Ellis et al., 2010; Duruflé et al., 2017). Classical AGPs do not contain this motif but many chimeric AGPs, including FLAs, contain the consensus sequence (Ellis et al., 2010).

O-glycosylation is the most complex type of glycosylation and it can occur on Ser and Hyp residues. Galactose (Gal) can be linked to Ser and Hyp residues whereas arabinose (Ara) can only be linked to Hyp residues (Duruflé et al., 2017). HRGPs are glycosylated by two types of O-glycosylation on their Hyp residues in the Golgi apparatus: Hyp arabinosylation and Hyp arabinogalactosylation (Kieliszewski, 2001). The Hyp contiguity hypothesis (Kieliszewski and Lampion, 1994) predicts that contiguous Hyp residues are arabinosylated, as occurs on EXTs, whereas non-contiguous Hyp residues are arabinogalactosylated giving rise to the arabinogalactan polysaccharides (AG) polysaccharides of the AGPs (Kieliszewski, 2001; Faye et al., 2005; Nguema-Ona et al., 2012).

Several GPI-anchored proteins were found to be candidates for AG glycosylation (Borner et al., 2002). Therefore, HRGPs are not the only proteins containing AG-polysaccharides.

1.4.1.4. AG polysaccharides - sugar link

The carbohydrate moiety of AGPs represents more than 90% of their total molecular mass (Nguema-Ona et al., 2012; Su and Higashiyama, 2018). AG glycans have a complex and heterogeneous structure that changes throughout development (Geshi, 2014). In AGPs, type II AGs are O-glycosidically linked to Hyp residues (Ellis et al., 2010). Type II AGs consist of β -(1→3)-linked D-Galp substituted at O6 positions with side-chains of β -(1→6)-linked D-Galp which are usually further decorated with Ara and less

frequently with other sugars, such as glucuronic acid (GlcA), rhamnose (Rha) and fucose (Fuc) (Gaspar et al., 2001; Tryfona et al., 2010). The β -1,6-galactan side chains often terminate with β -1,6-GlcA or 4-O-[Me]GlcA (Tryfona et al., 2012; Temple et al., 2019). Other two types of AG polysaccharides have been described in plants. Type I AGs, which can be found in pectins, are composed of β -(1 $\rightarrow$ 4)-linked D-Galp backbone substituted by Ara residues at O3 positions and by β -D-Galp at O6 positions (Clarke et al., 1979; Voragen et al., 2009; Hijazi et al., 2014). Type III AGs are formed by a short linear β -(1 $\rightarrow$ 6)-linked D-Galp chain which contains Gal and Ara residues and large branched Ara chains (Leonard et al., 2005; Hijazi et al., 2014).

1.4.1.5. AGPs in the extracellular space - facing the challenge

Finally, mature O-glycosylated AGPs are then transferred to the extracellular space via post-Golgi vesicles where they remain attached to the extracellular face of the membrane (Ma et al., 2018).

GPI-anchored AGPs can be released from the plasma membrane by specific phospholipases that cleave the GPI moiety (Schultz et al., 1998; Gillmor et al., 2005; Duruflé et al., 2017). Several GPI-anchored proteins such as FLAs, COBRA-like and receptor-like proteins were release by phospholipases in Arabidopsis (Borner et al., 2003). This process may serve to regulate the amount of these glycoproteins in the plasma membrane or to release AGPs into the extracellular matrix where they may act as a soluble signal for neighbouring cells (Schultz et al., 1998).

AGPs may also be processed by proteolytic activities and glycosyl hydrolases or transported by endocytotic multivesicular bodies to the vacuole where they are degraded (Knoch et al., 2014).

1.4.2. AGPs: what do they do and where can they be found?

AGPs are widely expressed throughout the plant kingdom, from bryophytes to angiosperms and are found in distinct developmental stages and cells, being mostly abundant in cell walls, plasma membranes, apoplastic space and secretions (Majewska-Sawka and Nothnagel, 2000; Ellis et al., 2010; Nguema-Ona et al., 2012).

In the past few years, several studies using bioinformatics, AG degradation enzymes monoclonal antibodies like JIM8, JIM13, JIM14, LM2 and CCRC-M7 that recognize carbohydrate epitopes and β -Yariv that specifically binds to the β -(1 $\rightarrow$ 3)- linked D-Galp backbone of AGPs (Kitazawa et al., 2013) have contributed to understand the structure, function and evolution of AGPs (Seifert and Roberts, 2007; Su and Higashiyama, 2018).

AGPs are implicated to function in many growth and development processes, such as cell division and PCD, embryo and postembryonic pattern, somatic embryogenesis, PT growth and guidance, pollen incompatibility, plant microbe interactions, cellular signaling, xylem differentiation, secondary wall deposition, hormone signaling pathways, abiotic stress responses (reviewed in Majewska-Sawka and Nothnagel, 2000; Seifert and Roberts, 2007; Ellis et al., 2010; Pereira et al., 2016).

Several studies revealed that AGPs are recruited into many reproduction processes in *Arabidopsis*: female gametophyte development (*AGP18*, Acosta-Garcia and Vielle-Calzada, 2004; Demesa-Arévalo and Vielle-Calzada, 2013), male gametophyte development (*AGP6*, *AGP11*, Coimbra et al., 2009; *FLA3*, Li et al., 2010) and male-female interactions (*ENODL11-15*, Hou et al. 2016; *JAGGER* or *AGP4*, Pereira et al., 2016b). The mysterious AGPs have several members expressed in distinct stages and specific positions of the reproductive process (Table 1).

AGP18 is expressed during female gametophyte development and is involved in megaspore selection. A RNA interference (RNAi) mutant line showed ovule abortion and consequently reduced seed set (Acosta-García and Vielle-Calzada, 2004; Demesa-Arévalo and Vielle-Calzada, 2013).

Two stamen and pollen-specific genes, *AGP6* and *AGP11*, are thought to be involved in pollen release and PT endosome machinery. Knockout and knockdown mutations on these genes caused pollen abortion (Levitin et al., 2008; Coimbra et al., 2009; Costa et al., 2013). Later, *FLA3* and *FLA14* were identified as being specifically expressed in pollen grains and PTs and involved in microspore development (Li et al., 2010; Miao et al., 2018).

Hou et al. (2016) showed that *ENODLs 11, 12, 13, 14* and *15* are essential for PT reception. *ENODLs 14* and *15* are expressed in the synergid filiform apparatus. The PTs of a quintuple mutant (knockout of *ENODLs 13, 14* and *15* and knockdown of *ENODLs 11* and *12*) fail to rupture and release their sperm cells. Recently, *JAGGER* was described as essential in the signaling pathway responsible for PT attraction blockage. In *jagger* mutants, the persistent synergid does not degenerate after double fertilization and continues to attract more PTs into the embryo sac (polytubey) (Pereira et al., 2016b).

1.5. General objective

Understanding the mechanisms, the protagonists and the interactions involved in a love story that leads to the formation of seeds is an essential question in reproductive biology. This complex biological process, double fertilization, is extremely important to improve

Table 1 - Information about the AGPs expressed in the reproductive tissues and presumed functions during sexual reproduction in *Arabidopsis thaliana*.

AGP	Locus	AGP subfamily	Expression	Mutants	Local of mutation	Mutant Phenotypes	Function(s)	References	
AGP1	At5g64310	Classical AGP	stigma; style; transmitting tract; chalaza; funiculus	-	-	-	PT-stigma interactions (guidance; competence; nutrition)	Pereira et al., 2014, 2016a	
JAGGER (AGP4)	At5g104380		transmitting tract; stigma; ovule integuments (micropyle); synergids; egg cells region	jagger1-1-/- (pst20518)	Exon	polytubey; survival of the persistent synergid after fertilization	persistent synergid death and polytubey block	Pereira et al., 2016b; Lopes et al., 2019	
				jagger1-2-/- (GK-134A10)	3' UTR				
AGP6	At5g14380		stamens; pollen; PTs	CSHL_ET5767	Exon	no apparent morphological phenotype (<i>agp6</i> ; <i>agp11</i> , <i>CSHL_ET3978</i>); collapsed pollen grains (<i>agp6</i> ± <i>agp11</i> ; <i>agp6agp11</i> ±; <i>agp6agp11</i> ; RNAi and amiRNA lines); inhibition of PT growth, hampered pollen release and reduced fertility (<i>CSHL_ET5767</i> ; RNAi lines); reduced seed set and precocious PT germination (<i>agp6agp11agp40</i>)	Pollen release and PT growth	Pereira et al., 2006; Coimbra et al., 2008, 2009; Levitin et al., 2008; Nguema-Ona et al., 2012	
				RNAi line					-
				agp6 (Ds54-4754-1)					-
AGP9	At2g14890		ovules	-	-	-	nutrition	Pereira et al., 2014	
AGP11	At3g01700	stamens; pollen; PTs	RNAi line	-	Exon	same as AGP6	pollen release and PT growth	Pereira et al., 2006; Levitin et al., 2008; Coimbra et al., 2009; Nguema-Ona et al., 2012	
			CSHL_ET3978						
			agp11 (Ds11-4025-1)						
			amiRNA line	-					
AGP12	At3g13520	AG peptide	stigmatic cells; chalaza; funiculus; septum	-	-	-	PT-stigma interactions (nutrition; guidance)	Pereira et al., 2014, 2016a	
AGP15	At5g11740		pistil	-	-	-			

Table 1 - (cont.)

AGP	Locus	AGP subfamily	Expression	Mutants	Local of mutation	Mutant Phenotype	Function(s)	References
AGP18	At4g37450	Lysine-rich AGP	MMC; megaspores; functional megaspore; synergids; antipodals; egg cell; embryo	RNAi line	-	ovule abortion and reduced seed set (RNAi line); no apparent morphological phenotype (<i>SALK_117268</i>)	Megaspore selection	Acosta-García and Vielle-Calzada, 2004; Yang et al., 2007; Demesa-Arévalo and Vielle-Calzada, 2013
				SALK_117268	Exon			
AGP19	At1g68725	AG peptide	flowers (pedicels, sepals, petals and filaments)	<i>agp19</i> (SALK_038728)		shorter inflorescence stems; fewer siliques	PT growth	Yang et al., 2007; Lopes et al., 2019
AGP22	At5g53250		nucellar epidermal cells	-	-	-	-	Tucker and Koltunow, 2014
AGP23	At3g57690		pollen; PTs	-	-	-	PT germination and growth; PT endosome machinery	Pereira et al., 2014, 2016a; Lopes et al., 2019
AGP24	At5g40730		functional megaspore; embryo sac; pollen	-	-	-	-	Yu et al., 2005; Tucker and Koltunow, 2014
AGP40	At3g20865		pollen	<i>agp40</i> (SALK_089247)		Exon	reduced seed set and precocious PT germination (<i>agp6agp11agp40</i>)	PT germination and growth; PT endosome machinery
FLA3	At2g24450	Fasciclin-like AGP	pollen; PTs	RNAi line	-	abnormal pollen grains; shorter siliques	Microspore development	Li et al., 2010
FLA8	At2g45470		chalaza; funiculus; integuments; seed coat; hypocotyl	-	-	-	-	Costa et al., 2019
FLA9	At1g03870		silique	<i>fla9-1</i> (SALK_051688)	3' UTR	seed abortion under drought stress	Seed development	Cagnola et al., 2018
				<i>fla9-2</i> (SALK_088265)	5' UTR			
FLA10	At3g60900		chalaza; endosperm; seed coat	-	-	-	-	Costa et al., 2019
FLA14	At3g12660	pollen; PTs	<i>fla14</i> (CS1014037)	Exon	precocious PT germination in high moisture	Microspore development	Miao et al., 2018	

Table 1 - (cont.)

AGP	Locus	AGP subfamily	Expression	Mutants	Local of mutation	Mutant Phenotype	Function(s)	References
ENODL1	At5g53870	ENODL	flowers; embryo sac	-	-	-	-	Scutt et al., 2003; Yu et al., 2005; Mashiguchi et al., 2009; Hou et al., 2016
ENODL3	At4g28365		flowers	enodl3-1 (SM_3_35947)	Promoter	no apparent morphological phenotype (single mutants; enodl1-1enodl2-2; enodl3-1enodl4-1; enodl5-2enodl6-1)		
ENODL4	At4g32490			enodl4-1 (SM_3_30584)	Exon			
				enod4-2 (GAB_771D10I)				
ENODL5	At3g18590		enodl5-1 (CS811245)	Intron				
			enodl5-2 (SALK_094988)					
			enodl5-3 (SALK_057240)					
ENODL6	At1g48940		enodl6-1 (CS828583)	Promoter				
			enodl6-2 (CS809265)	3' UTR				
ENODL7	At1g79800		flowers; embryo sac	enodl7-1 (CS832808)	Promoter			
ENODL9	At3g20570		flowers	enodl9-1 (SALK_105873)	Exon	Reduction in the number and biomass of the siliques (enodl9-1) No reduction in reproductive potential (enodl9-1; enodl9-2; enodl9-3)		Mashiguchi et al., 2009
				enodl9-2 (SALK_126416)				
				enodl9-3 (GABI_371E08)				
ENODL10	At5g57920		enodl10-1 (SALK_197590)	3' UTR	no apparent morphological phenotype			
			enodl10-2 (SALK_019989)					

Table 1 - (cont.)

AGP	Locus	AGP subfamily	Expression	Mutants	Local of mutation	Mutant Phenotype	Function(s)	References	
ENODL11	At2g23990	ENODL	flowers; embryo sac; developing ovule	<i>enodl11-1</i> (GABI_589D09)	Promoter	no apparent morphological phenotype (single mutants; <i>enodl11-1enodl12-1</i> ; <i>enodl14-2enodl15-1</i> ; <i>enodl13-1enodl14-2</i> ; <i>enodl14-2enodl15-2</i> ; <i>enodl13-1enodl15-2</i> ; <i>enodl13-1enodl14-2enodl15-2</i>)	PT reception	Yu et al., 2005; Khan et al., 2007; Mashiguchi et al., 2009; Wuest et al., 2010; Hou et al., 2016	
				RNAi	-				
ENODL12	At4g30590		flowers; developing ovule; embryo sac; pollen; anthers; synergids	<i>enodl12-1</i> (SALK_023946)	Exon			Yu et al., 2005; Khan et al., 2007; Steffen et al., 2007; Mashiguchi et al., 2009; Wuest et al., 2010; Hou et al., 2016	
ENODL13	At5g25090		flowers; developing ovule	<i>enodl13-1</i> (CS835073)	Intron			Khan et al., 2007; Mashiguchi et al., 2009; Wuest et al., 2010; Hou et al., 2016	
				<i>enodl13-2</i> (CS817616)	Promoter			ovule abortion; reduced seed set; PTs fail to release sperm cells (<i>enodl11enodl12enodl13-1enodl14-2enodl15-2</i>)	
ENODL14	At2g25060		flowers; mature ovule; micropylar region; synergid filiform apparatus	<i>enodl14-1</i> (SM_3_484)	Exon	Mashiguchi et al., 2009; Hou et al., 2016			
			<i>enodl14-2</i> (SALK_019896)	Intron					
ENODL15	At4g31840		flowers; developing ovule; mature ovule; micropylar region; transmitting tissue; funiculus; synergid filiform apparatus; anthers; pollen	<i>enodl15-1</i> (SALK_093027)	Exon	Khan et al., 2007; Mashiguchi et al., 2009; Hou et al., 2016			
			<i>enodl15-2</i> (SALK_056190)						
ENODL16	At3g01070		embryo sac				Yu et al., 2005		
ENODL17	At5g15350		mature ovule				Hou et al., 2016		
ENODL18	At1g08500				-	-	-	-	
ENODL21	At5g14345		flowers						Mashiguchi et al., 2009

agricultural productivity. Hence, exploring candidates targets in the model plant *Arabidopsis thaliana* probably involved in reproduction is the first big step in this long and complicated story.

A. thaliana is a small flowering plant that belongs to the Brassicaceae family. It was the first plant to have its genome sequenced. Due to the small size, short life cycle, small genome and large numbers of seeds, *Arabidopsis* is an ideal plant and one of the model organisms used to study plant biology because it is easily grown under laboratory conditions. Besides, *Arabidopsis* is an excellent model to study seed formation, as it shares a conserved developmental program with major seed-producing crop plants (Kaul et al., 2000). For all these reasons, *A. thaliana* was the chosen plant for this study.

In this thesis, chapter II addresses the role of fasciclin-like arabinogalactan 5 (FLA5), an AGP putative involved in *Arabidopsis* reproduction while chapter III focuses on the role of methyl GlcA ([Me]GlcA) of AGPs that binds and stores calcium (Ca^{2+}) predominantly involved in signaling pathways along the reproductive tissues.

Specifically, this research work aims at:

- Characterize the expression profile of *FLA5*;
- Functional analyze *FLA5*;
- Evaluate the expression of glucuronosyltransferase (GlcAT) genes;
- Assess the subcellular localization of GlcATs;
- Functional characterize GlcATs during sexual reproduction;

Additionally, we expect to provide novel insights into the study of this mysterious class of glycoproteins.

2. GENERAL METHODOLOGY

2.1. Growth *in vitro*

2.1.1. Sterilization of seeds

The equivalent to 100 μL of *Arabidopsis* seeds was sterilized in 1.5 mL microcentrifuge tubes (Eppendorf) with 1 mL of sterilizing solution: 10 mL 80% (v/v) aqueous ethanol (Aga), 100 μL hypochlorite solution (Merck), 100 μL Tween 20 (Sigma-Aldrich); and shaken at 140,000 $\times g$ for 7 min (MiniSpin®, Eppendorf). The sterilizing solution was replaced sequentially and mixed by inversion with 65%, 80% and 100% (v/v) ethanol (Aga) in the flow cabinet. Seeds were transferred to previously sterilized filter paper (Fisher Scientific) and left to dry.

2.1.2. Growth *in vitro*

After sterilization, dry seeds were sown onto 100 mm x 20 mm plates (Corning®) containing solidified half strength Murashige & Skoog (MS) basal medium including vitamins (4.43 g/L; PhytoTechnology Laboratories), 2-(N-Morpholino)ethanesulfonic acid (MES) (1.0 g/L; Sigma-Aldrich), sucrose (10 g/L; Sigma-Aldrich), adjusted to pH 5.8 using KOH (Sigma-Aldrich) and plant agar (9 g/L; PhytoTechnology Laboratories). The plates were kept for 48 h at 4°C in the dark to induce stratification. Then, plates were placed vertically (slightly tilted) for 2 weeks under a long day photoperiod (16 h light at 22°C and 8 h darkness at 18°C) with 50-60% relative humidity and light intensity at 180 $\mu\text{mole m}^{-2} \text{s}^{-1}$.

2.2. Genotyping

To confirm the transfer-DNA (T-DNA) insertions and to identify the homozygous mutants, a polymerase chain reaction (PCR)-based approach was employed. Genotyping primers (Supplementary table 1) were designed using the T-DNA Primer Design tool provided by the Salk Institute Genomic Analysis Laboratory (SIGnAL) database. Phire Plant Direct PCR Kit (without sampling tools) (Thermo Scientific™) was used to genotype mutant lines according to the manufacturer's protocol. The annealing temperatures to the Phire Hot Start II DNA Polymerase were estimated using the T_m Calculator (Thermo Scientific™). The reaction mixtures were placed on a T100™ Thermal Cycler (Bio-Rad). The insertion sites for the mutants were confirmed by DNA sequencing of PCR-amplified products. PCR products were purified by gel extraction with GeneJET Gel Extraction Kit

(Thermo Scientific™) and sequenced. All the sequencing data were analyzed using Geneious Prime (<https://www.geneious.com>).

2.3. Gene expression analysis

2.3.1. RNA extraction and quantification

Total RNA from different tissues (Table 2) was extracted from wt plants in triplicates using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's instructions. For the identification of knockout plants, RNA from 5-week-old inflorescences of each mutant line was prepared in triplicates using the same kit. All RNA was quantified using a NanoDrop™ One UV-Vis Spectrophotometer (Thermo Scientific™) and RNA integrity was evaluated by agarose gel electrophoresis by checking the presence of the bands of 25S and 18S ribosomal RNAs.

Table 2 - Description of the samples used to extract RNA.

Sample	Description
Seedling	Whole seedlings collected at 7 th day after germination. Seedlings grown on Murashige Skoog (MS) medium plates.
Juvenile leaf	Leaves collected from seedlings at 14 th day after germination. Seedlings grown on MS medium plates.
Cauline leaf	Cauline leaves from 5-week-old plants.
Stem	Internodes between rosette leaf and first cauline leaf from 5-week old plants.
Root	Roots collected from seedlings at 14 th day after germination. Seedlings grown on MS medium plates.
Inflorescence	Whole inflorescences from 5-week-old plants.
Silique	Whole mature siliques from 6-week-old plants.

2.3.2. cDNA synthesis

RNA was treated with DNase I, RNase-free (Thermo Scientific™), according to the manufacturer's protocol, to remove co-extracted DNA. cDNA was synthesized using SuperScript™ III Reverse Transcriptase (Invitrogen™) and oligo(dT)18 primers to initiate the reactions, following the manufacturer's instructions. All cDNA was tested using a PCR-based approach using Platinum™ *Taq* DNA Polymerase (Invitrogen™) using the primers to amplify *ACTIN8* (*ACT8*, At1g49240) (Supplementary table 2).

2.3.3. Quantitative polymerase chain reaction (qPCR)

Primers to analyze the expression level of the genes of interest were designed (Supplementary table 2) and tested to verify their specificity. cDNA was amplified as outlined in Burton et al. (2008) with the following modifications. To provide a template for

the standard curve, PCR reaction mixtures were combined for purification by high performance liquid chromatography (HPLC). cDNA fragments were confirmed by DNA sequencing and sequencing data were analyzed using Geneious Prime (<https://www.geneious.com>). The transcript levels of genes encoding *CYCLOPHILIN* (At2g36130), *ACTIN-RELATED PROTEIN 9* (ARP9, At5g43500), *TUBULIN ALPHA-2 CHAIN* (TUA2, At1g50010) and *ELONGATION FACTOR 1B ALPHA-SUBUNIT 1* (EF1B α -subunit 1, At5g12110) were used as controls (Supplementary table 2) (Li et al., 2014). The final concentrations of mRNAs of the genes of interest are expressed as arbitrary units that represent the numbers of copies per microliter of cDNA, normalized against the geometric means of the three control genes that vary the least with respect to each other (Vandesompele et al., 2002). The mean expression level and standard deviation (SD) for each set of three technical replicates for each cDNA were calculated. The HPLC of qPCR products was kindly performed Mr. Chao Ma (School of Food Agriculture and Wine, University of Adelaide, Australia). The qPCR experiments and data normalization were conducted by Dr. Neil Shirley (School of Food Agriculture and Wine, University of Adelaide, Australia).

2.4. Bacteria Screening

The screening for positive colonies of bacteria was performed through colony PCR by testing ten colonies of each transformation using Platinum™ Taq DNA Polymerase (Invitrogen™). The primers used are listed on Supplementary table 3-5. The annealing temperatures to the Platinum™ Taq DNA Polymerase were estimated using the Tm Calculator (Thermo Fisher Scientific). The components to perform a colony PCR are the same to perform a normal PCR but instead of using DNA, it was added a bit of a colony using a toothpick. The colony PCR was performed using the cycling conditions of the manufacturer but with an initial denaturation at 94°C for 5 min.

The positive colonies were inoculated in 5 mL of Luria Broth (LB) Base, Miller (BD Difco™) 2.5% (w/v) with the appropriate antibiotics and incubated with agitation (200 rpm) at 37°C overnight for *Escherichia coli* or at 28°C for two days for *Agrobacterium tumefaciens* (Shaking Incubator, SHEL LAB). Bacterial plasmid DNA was extracted from the liquid cultures using the GeneJET Plasmid Miniprep Kit (Thermo Scientific™), as instructed by the manufacturer. Bacterial DNA was quantified using NanoDrop™ One UV-Vis Spectrophotometer (Thermo Scientific™).

2.5. Arabidopsis transformation

Arabidopsis transformation was performed using 5-week-old plants grown in soil according to Clough and Bent (1998) with minor changes. Three days before the transformation of the plants, 5 mL of LB medium (BD Difco™) 2.5% (w/v) were inoculated with the appropriate antibiotics and with one isolated colony of *A. tumefaciens* with the clone of interest. These cultures were incubated at 28°C with agitation (200 rpm) for two days (Shaking Incubator, SHEL LAB). Then, the cultures were inoculated in 200 mL of LB medium (BD Difco™) 2.5% (w/v) supplemented with the appropriate antibiotics and incubated at 28°C with agitation (200 rpm) overnight (Shaking Incubator, SHEL LAB). For the transformation all siliques and open flowers from the plants were removed. The flowers were dipped for 2 min in the floral dip inoculation medium twice, with gentle agitation, in order to facilitate the entry of the solution in the floral buds. The transformed plants were sheltered with plastic bags for 12-16 h to maintain humidity.

2.6. Developmental analysis

2.6.1. Clearing

Arabidopsis ovules and seeds at different developmental stages were cleared and analyzed as described in Brambilla et al. (2007). Samples were observed using a Zeiss Axio Imager M2 microscope equipped with differential interference contrast (DIC) optics. Images were captured on an Axiocam MRm (Zeiss) camera using the Zen 2011 (Zeiss) software.

2.6.2. PT staining with aniline blue

PT staining with aniline blue was carried out as outlined in Pereira et al. (2016b). The specimens were observed with a Zeiss Axio Imager M2 microscope and the images were captured on an Axiocam MRm (Zeiss) camera using the Zen 2011 (Zeiss) software.

2.6.3. Alexander staining

Alexander staining was performed as described by Peterson et al. (2010) with the following modifications. Anthers were removed from 5-week-old plants, placed on a microscope slide and covered with Alexander's solution for 4 h at room temperature (21°C). The anthers were observed with a Zeiss Axio Imager M2 microscope and the images were captured on an Axiocam MRm camera (Zeiss) using the Zen 2011 (Zeiss) software.

2.6.4. Seed set analysis

Siliques from 6-week-old plants were decolorized in 100% (v/v) ethanol (Aga) at 37°C overnight before dissection to facilitate seed number count and silique length measure. Seeds were divided into four categories: green seeds fully developed, white seeds arrested at globular stage of embryo development, aborted seeds that began to desiccate before the white seeds and aborted ovules.

Siliques were observed under a stereomicroscope (Zeiss Stemi 2000-C) and images were captured using a Zeiss AxioCam ERc 5s camera using Zeiss software (AxioVision SE64 Rel. 4.9.1).

2.7. Statistical Analysis

Data were statistically treated using GraphPad Prism 8 software (www.graphpad.com). For each analysis, phenotypic differences in mutants were compared with the wt using one or two-way ANOVA.

2.8. Bioinformatic analysis

2.8.1. Sequence and phylogenetic analysis

Arabidopsis genes were localized on the chromosomes using the Chromosome Map Tool (<https://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp>) available at The Arabidopsis Information Resource (TAIR).

Arabidopsis protein sequences were scanned through various prediction software. ExPASy (Gasteiger et al., 2003) showed amino acid composition, theoretical isoelectric point (pI) and molecular weight of proteins. TMHMM 2.0 was used to predict the presence of transmembrane domains (TMDs) (Sonnhammer et al., 1998). TargetP 1.1 (Emanuelsson et al., 2000), ARAMEMNON (Schwacke et al., 2003) and SUBA (Hooper et al., 2017) were run to predict the subcellular localization of protein sequences. Pfam (El-Gebali et al., 2019) and Profilescan (Sigrist et al., 2010) were run to predict domains and N-glycosylation sites. The presence of signal peptide and GPI anchor addition sequences were predicted with SignalP 5.0 Server (Almagro Armenteros et al., 2019) and big-PI plant predictor (Eisenhaber et al., 2003), respectively.

The full-length amino acid sequences were aligned using Clustal W (Thompson et al., 1994) to obtain a sequence identity matrix. Phylogenetic trees were obtained with Molecular Evolutionary Genetics Analysis X (MEGA X) software (Kumar et al., 2018), using the Neighbor-Joining method. The bootstrap consensus trees inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed. The

evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

2.8.2. Gene expression data

Gene expression data was obtained from Arabidopsis eFP Browser (Waese et al., 2017) with Klepikova Atlas as data source (Klepikova et al., 2016).

3. REFERENCES

- Acosta-García G, Vielle-Calzada JP** (2004) A classical arabinogalactan protein is essential for the initiation of female gametogenesis in Arabidopsis. *Plant Cell* **16**: 2614–2628 doi: 10.1105/tpc.104.024588
- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H** (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**: 420–423 doi: 10.1038/s41587-019-0036-z
- Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, Orlando R, Scheller H V., Mohnen D** (2011) Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. *Proc Natl Acad Sci U S A* **108**: 20225–20230 doi: 10.1073/pnas.1112816108
- Basu D, Tian L, DeBrosse T, Poirier E, Emch K, Herock H, Travers A, Showalter AM** (2016) Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) Pathway in Arabidopsis. *PLoS One* **11**: e0145092 doi: 10.1371/journal.pone.0145092
- Batista RA, Figueiredo DD, Santos-González J, Köhler C** (2019) Auxin regulates endosperm cellularization in Arabidopsis. *Genes Dev* **33**: 466–476 doi: 10.1101/gad.316554.118
- Beale KM, Johnson MA** (2013) Speed dating, rejection, and finding the perfect mate: Advice from flowering plants. *Curr Opin Plant Biol* **16**: 590–597 doi: 10.1016/j.pbi.2013.08.005
- Bedinger P** (1992) The remarkable biology of pollen. *Plant Cell* **4**: 879–887 doi: 10.1105/tpc.4.8.879
- Berger F, Hamamura Y, Ingouff M, Higashiyama T** (2008) Double fertilization - caught in the act. *Trends Plant Sci* **13**: 437–443 doi: 10.1016/j.tplants.2008.05.011
- Borg M, Brownfield L, Twell D** (2009) Male gametophyte development: A molecular perspective. *J Exp Bot* **60**: 1465–1478 doi: 10.1093/jxb/ern355
- Borg M, Twell D** (2010) Life after meiosis: Patterning the angiosperm male gametophyte. *Biochem Soc Trans* **38**: 577–582 doi: 10.1042/BST0380577
- Borner GH, Sherrier DJ, Stevens TJ, Arkin IT, Dupree P** (2002) Prediction of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A genomic analysis.

- Plant Physiol **129**: 486–499 doi: 10.1104/pp.010884
- Borner GHH, Lilley KS, Stevens TJ, Dupree P** (2003) Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. Plant Physiol **132**: 568–577 doi: 10.1104/pp.103.021170
- 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 doi: 10.1105/tpc.107.051797
- Burton RA, Jobling SA, Harvey AJ, Shirley NJ, Mather DE, Bacic A, Fincher GB** (2008) The genetics and transcriptional profiles of the cellulose synthase-like HvCslF gene family in barley. Plant Physiol **164**: 1821–1833 doi: 10.1104/pp.107.114694
- Cagnola JI, Dumont de Chassart GJ, Ibarra SE, Chimenti C, Ricardi MM, Delzer B, Ghiglione H, Zhu T, Otegui ME, Estevez JM, et al** (2018) Reduced expression of selected FASCICLIN-LIKE ARABINOGLACTAN PROTEIN genes associates with the abortion of kernels in field crops of Zea mays (maize) and of Arabidopsis seeds. Plant Cell Environ **41**: 661–674 doi: 10.1111/pce.13136
- Clarke AE, Anderson RL, Stone BA** (1979) Form and function of arabinogalactans and arabinogalactan-proteins. Phytochemistry **18**: 521–540 doi: 10.1016/S0031-9422(00)84255-7
- Clough SJ, Bent AF** (1998) Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J **16**: 735–743 doi: 10.1046/j.1365-313X.1998.00343.x
- Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG** (2009) Pollen grain development is compromised in Arabidopsis agp6 agp11 null mutants. J Exp Bot **60**: 3133–3142 doi: 10.1093/jxb/erp148
- Coimbra S, Jones B, Pereira LG** (2008) Arabinogalactan proteins (AGPs) related to pollen tube guidance into the embryo sac in Arabidopsis. Plant Signal Behav **3**: 455–456 doi: 10.4161/psb.3.7.5601
- Cosgrove DJ** (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol **6**: 850–861 doi: 10.1038/nrm1746
- Cosgrove DJ, Jarvis MC** (2012) Comparative structure and biomechanics of plant primary and secondary cell walls. Front Plant Sci **3**: 1–6 doi: 10.3389/fpls.2012.00204
- 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 Biol* **13**: 1–19 doi: 10.1186/1471-2229-13-7
- Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S** (2019) In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reprod* **32**: 353–370 doi: 10.1007/s00497-019-00376-7
- Crawford BCW, Yanofsky MF** (2008) The formation and function of the female reproductive tract in flowering plants. *Curr Biol* **18**: R972-978 doi: 10.1016/j.cub.2008.08.010
- Demesa-Arévalo E, Vielle-Calzada JP** (2013) The classical arabinogalactan protein AGP18 mediates megaspore selection in Arabidopsis. *Plant Cell* **25**: 1274–1287 doi: 10.1105/tpc.112.106237
- Dresselhaus T, Franklin-Tong N** (2013) Male-female crosstalk during pollen germination, tube growth and guidance, and double fertilization. *Mol Plant* **6**: 1018–1036 doi: 10.1093/mp/sst061
- Drews GN, Lee D, Christensen CA** (1998) Genetic analysis of female gametophyte development and function. *Plant Cell* **10**: 5–17 doi: 10.1105/tpc.10.1.5
- Drews GN, Yadegari R** (2002) Development and function of the angiosperm female gametophyte. *Annu Rev Genet* **36**: 99–124 doi: 10.1146/annurev.genet.36.040102.131941
- Du H, Clarke AE, Bacic A** (1996) Arabinogalactan proteins: A class of extracellular matrix proteoglycans involved in plant growth and development. *Trends Cell Biol* **6**: 411–414 doi: 10.1016/S0962-8924(96)20036-4
- Durufié H, Hervé V, Balliau T, Zivy M, Dunand C, Jamet E** (2017) Proline hydroxylation in cell wall proteins: Is it yet possible to define rules? *Front Plant Sci* **8**: 1–10 doi: 10.3389/fpls.2017.01802
- Eisenhaber B, Wildpaner M, Schultz CJ, Borner GHH, Dupree P, Eisenhaber F** (2003) Glycosylphosphatidylinositol Lipid Anchoring of Plant Proteins. Sensitive Prediction from Sequence- and Genome-Wide Studies for Arabidopsis and Rice. *Plant Physiol* **133**: 1691–1701 doi: 10.1104/pp.103.023580
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al** (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427–D432 doi: 10.1093/nar/gky995
- Ellis M, Egelund J, Schultz CJ, Bacic A** (2010) Arabinogalactan-proteins: Key

- regulators at the cell surface? *Plant Physiol* **153**: 403–419 doi: 10.1104/pp.110.156000
- Emanuelsson O, Nielsen H, Brunak S, Von Heijne G** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005–1016 doi: 10.1006/jmbi.2000.3903
- Faye L, Boulafloous A, Benchabane M, Gomord V, Michaud D** (2005) Protein modifications in the plant secretory pathway: Current status and practical implications in molecular pharming. *Vaccine* **23**: 1770–1778 doi: 10.1016/j.vaccine.2004.11.003
- Figueiredo DD, Köhler C** (2018) Auxin: A molecular trigger of seed development. *Genes Dev* **32**: 479–490 doi: 10.1101/gad.312546.118
- Foley J** (2014) Feeding 9 Billion. Retrieved August 29, 2019, from <https://www.nationalgeographic.com/foodfeatures/feeding-9-billion/>
- Gaspar Y, Johnson KL, Mckenna JA, Bacic A, Schultz CJ** (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol Biol* **47**: 161–176 doi: 10.1023/A:1010683432529
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A** (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**: 3784–3788 doi: 10.1093/nar/gkg563
- Geshi N** (2014) Arabinogalactan glycosyltransferases: enzyme assay, protein-protein interaction, subcellular localization, and perspectives for application. *Adv Bot* **2014**: 1–7 doi: 10.1155/2014/434979
- Gillmor CS, Lukowitz W, Brininstool G, Sedbrook JC, Hamann T, Poindexter P, Somerville C** (2005) Glycosylphosphatidylinositol-anchored proteins are required for cell wall synthesis and morphogenesis in arabidopsis. *Plant Cell* **17**: 1128–1140 doi: 10.1105/tpc.105.031815
- Hijazi M, Velasquez SM, Jamet E, Estevez JM, Albenne C** (2014) An update on post-translational modifications of hydroxyproline-rich glycoproteins: Toward a model highlighting their contribution to plant cell wall architecture. *Front Plant Sci* **5**: 1–10 doi: 10.3389/fpls.2014.00395
- Höfte H, Voxeur A** (2017) Plant cell walls. *Curr Biol* **27**: R865–R870 doi: 10.1016/j.cub.2017.05.025
- Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH** (2017) SUBA4: The interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res* **45**: D1064–D1074 doi: 10.1093/nar/gkw1041

- Hou Y, Guo X, Cyprys P, Zhang Y, Bleckmann A, Cai L, Huang Q, Luo Y, Gu H, Dresselhaus T, et al** (2016) Maternal ENODLs Are Required for Pollen Tube Reception in Arabidopsis. *Curr Biol* **26**: 2343–2350 doi: 10.1016/j.cub.2016.06.053
- Hurtado-Guerrero R, Davies GJ** (2012) Recent structural and mechanistic insights into post-translational enzymatic glycosylation. *Curr Opin Chem Biol* **16**: 479–487 doi: 10.1016/j.cbpa.2012.10.013
- Jamet E, Albenne C, Boudart G, Irshad M, Canut H, Pont-Lezica R** (2008) Recent advances in plant cell wall proteomics. *Proteomics* **8**: 893–908 doi: 10.1002/pmic.200700938
- Johnson KL, Cassin AM, Lonsdale A, Wong GKS, Soltis DE, Miles NW, Melkonian M, Melkonian B, Deyholos MK, Leebens-Mack J, et al** (2017) Insights into the evolution of hydroxyproline-rich glycoproteins from 1000 plant transcriptomes. *Plant Physiol* **174**: 904–921 doi: 10.1104/pp.17.00295
- Johnson KL, Jones BJ, Bacic A, Schultz CJ** (2003) The fasciclin-Like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiol* **133**: 1911–1925 doi: 10.1104/pp.103.031237
- Kasahara RD, Portereiko MF, Sandaklie-Nikolova L, Rabiger DS, Drews GN** (2005) MYB98 is required for pollen tube guidance and synergid cell differentiation in Arabidopsis. *Plant Cell* **17**: 2981–2992 doi: 10.1105/tpc.105.034603
- Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, et al** (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**: 796–815 doi: 10.1038/35048692
- Keegstra K** (2010) Plant Cell Walls. *Plant Physiol* **154**: 483–486 doi: 10.1104/pp.110.161240
- Khan JA, Wang Q, Sjölund RD, Schulz A, Thompson GA** (2007) An early nodulin-like protein accumulates in the sieve element plasma membrane of arabidopsis. *Plant Physiol* **143**: 1576–1589 doi: 10.1104/pp.106.092296
- Kieliszewski MJ** (2001) The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**: 319–323 doi: 10.1016/S0031-9422(01)00029-2
- Kieliszewski MJ, Lamport DTA** (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J* **5**: 157–172 doi: 10.1046/j.1365-313X.1994.05020157.x
- Kitazawa K, Tryfona T, Yoshimi Y, Hayashi Y, Kawauchi S, Antonov L, Tanaka H, Takahashi T, Kaneko S, Dupree P, et al** (2013) β -galactosyl yariv reagent binds to the β -1,3-galactan of arabinogalactan proteins. *Plant Physiol* **161**: 1117–1126

doi: 10.1104/pp.112.211722

- Klemm D, Heublein B, Fink HP, Bohn A** (2005) Cellulose: Fascinating biopolymer and sustainable raw material. *Angew Chemie - Int Ed* **44**: 3358–3393 doi: 10.1002/anie.200460587
- Klepikova A V., Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA** (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J* **88**: 1058–1070 doi: 10.1111/tpj.13312
- Knoch E, Dilokpimol A, Geshi N** (2014) Arabinogalactan proteins: Focus on carbohydrate active enzymes. *Front Plant Sci* **5**: 1–9 doi: 10.3389/fpls.2014.00198
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K** (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549 doi: 10.1093/molbev/msy096
- Lalanne E, Honys D, Johnson A, Borner GHH, 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 doi: 10.1105/tpc.014407
- Leonard R, Petersen BO, Himly M, Kaar W, Wopfner N, Kolarich D, Van Ree R, Ebner C, Duus J, Ferreira F, et al** (2005) Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. *J Biol Chem* **280**: 7932–7940 doi: 10.1074/jbc.M410407200
- Lerouxel O, Cavalier DM, Liepman AH, Keegstra K** (2006) Biosynthesis of plant cell wall polysaccharides - a complex process. *Curr Opin Plant Biol* **9**: 621–630 doi: 10.1016/j.pbi.2006.09.009
- Levitin B, Richter D, Markovich I, Zik M** (2008) Arabinogalactan proteins 6 and 11 are required for stamen and pollen function in *Arabidopsis*. *Plant J* **56**: 351–363 doi: 10.1111/j.1365-313X.2008.03607.x
- Li J, Yu M, Geng LL, Zhao J** (2010) The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of *Arabidopsis*. *Plant J* **64**: 482–497 doi: 10.1111/j.1365-313X.2010.04344.x
- Li M, Lopato S, Hrmova M, Pickering M, Shirley N, Koltunow AM, Langridge P** (2014) Expression patterns and protein structure of a lipid transfer protein END1 from *Arabidopsis*. *Planta* **240**: 1319–1334 doi: 10.1007/s00425-014-2155-6
- Lopes AL, Moreira D, Ferreira MJ, Pereira AM, Coimbra S** (2019) Insights into secrets along the pollen tube pathway in need to be discovered. *J Exp Bot* **70**: 2979–2992

doi: 10.1093/jxb/erz087

- Ma Y, Zeng W, Bacic A, Johnson K** (2018) AGPs Through Time and Space. *Annu Plant Rev* **1**: 1–38 doi: 10.1002/9781119312994.apr0608
- Majewska-Sawka A, Nothnagel EA** (2000) The multiple roles of arabinogalactan proteins in plant development. *Plant Physiol* **122**: 3–10 doi: 10.1104/pp.122.1.3
- Maruyama D, Völz R, Takeuchi H, Mori T, Igawa T, Kurihara D, Kawashima T, Ueda M, Ito M, Umeda M, et al** (2015) Rapid elimination of the persistent synergid through a cell fusion mechanism. *Cell* **161**: 907–918 doi: 10.1016/j.cell.2015.03.018
- Mashiguchi K, Asami T, Suzuki Y** (2009) Genome-wide identification, structure and expression studies, and mutant collection of 22 early nodulin-like protein genes in arabidopsis. *Biosci Biotechnol Biochem* **73**: 2452–2459 doi: 10.1271/bbb.90407
- McCormick S** (2004) Control of male gametophyte development. *Plant Cell* **16**: S142–S153 doi: 10.1105/tpc.016659
- McCormick S** (1993) Male Gametophyte Development. *Plant Cell* **5**: 1265–1275 doi: 10.1007/978-3-642-02301-9
- Miao Y, Lin S, Cao J** (2018) The fasciclin-like arabinogalactan protein gene, FLA14, is involved in microspore development of Arabidopsis. Poster presented at 25th International Congress on Sexual Plant Reproduction; 2018 Jun 11-16; Gifu, Japan
- Mohnen D** (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* **11**: 266–277 doi: 10.1016/j.pbi.2008.03.006
- Nguema-Ona E, Coimbra S, Vitré-Gibouin M, Mollet JC, Driouich A** (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. *Ann Bot* **110**: 383–404 doi: 10.1093/aob/mcs143
- O'Sullivan AC** (1997) Cellulose: The structure slowly unravels. *Cellulose* **4**: 173–207 doi: 10.1023/A:1018431705579
- Owen HA, Makaroff CA** (1995) Ultrastructure of microsporogenesis and microgametogenesis in *Arabidopsis thaliana* (L.) Heynh. ecotype Wassilewskija (Brassicaceae). *Protoplasma* **185**: 7–21 doi: 10.1007/BF01272749
- 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 doi: 10.1002/wdev.6
- 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**: 1–15 doi: 10.3389/fpls.2016.01895
- Pereira AM, Masiero S, Nobre MS, Costa ML, Solís 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 doi: 10.1093/jxb/eru300
- 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 doi: 10.1016/j.molp.2016.01.002
- Pereira AM, Pereira LG, Coimbra S** (2015) Arabinogalactan proteins: rising attention from plant biologists. *Plant Reprod* **28**: 1–15 doi: 10.1007/s00497-015-0254-6
- Pereira LG, Coimbra S, Oliveira H, Monteiro L, Sottomayor M** (2006) Expression of arabinogalactan protein genes in pollen tubes of *Arabidopsis thaliana*. *Planta* **223**: 374–380 doi: 10.1007/s00425-005-0137-4
- Peterson R, Slovin JP, Chen C** (2010) A simplified method for differential staining of aborted and non-aborted pollen grains. *Int J Plant Biol* **1**: 66–69 doi: 10.4081/pb.2010.e13
- Pinto SC, Mendes MA, Coimbra S, Tucker MR** (2019) Revisiting the Female Germline and Its Expanding Toolbox. *Trends Plant Sci* **24**: 455–467 doi: 10.1016/j.tplants.2019.02.003
- Schatz G, Dobberstein B** (1996) Common principles of protein translocation across membranes. *Science* **271**: 1519–1526 doi: 10.1126/science.271.5255.1519
- Scheller HV, Ulvskov P** (2010) Hemicelluloses. *Annu Rev Plant Biol* **61**: 263–289 doi: 10.1146/annurev-arplant-042809-112315
- Schultz C, Gilson P, Oxley D, Youl J, Bacic A** (1998) GPI-anchors on arabinogalactan-proteins: Implications for signalling in plants. *Trends Plant Sci* **3**: 426–431 doi: 10.1016/S1360-1385(98)01328-4
- Schultz CJ, Johnson KL, Currie G, Bacic A** (2000) The classical arabinogalactan protein gene family of *Arabidopsis*. *Plant Cell* **12**: 1751–1768 doi: 10.1105/tpc.12.9.1751
- Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A** (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* **129**: 1448–1463 doi: 10.1104/pp.003459
- Schwacke R, Schneider A, Van Der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R** (2003) ARAMEMNON, a novel database for *Arabidopsis* integral membrane proteins. *Plant Physiol* **131**: 16–26 doi: 10.1104/pp.011577

- Scott RJ, Spielman M, Dickinson HG** (2004) Stamen structure and function. *Plant Cell* **16**: S46–S60 doi: 10.1105/tpc.017012
- Scutt CP, Vinauger-Douard M, Fourquin C, Ailhas J, Kuno N, Uchida K, Gaude T, Furuya M, Dumas C** (2003) The identification of candidate genes for a reverse genetic analysis of development and function in the *Arabidopsis* gynoecium. *Plant Physiol* **132**: 653–665 doi: 10.1104/pp.102.017798
- Seifert GJ, Roberts K** (2007) The biology of arabinogalactan proteins. *Annu Rev Plant Biol* **58**: 137–161 doi: 10.1146/annurev.arplant.58.032806.103801
- Showalter AM** (1993) Structure and function of plant cell wall proteins. *Plant Cell* **5**: 9–23 doi: 10.2307/3869424
- Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR** (2010) A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol* **153**: 485–513 doi: 10.1104/pp.110.156554
- Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N** (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**: D161–6 doi: 10.1093/nar/gkp885
- Soltis D, Soltis P, Endress P, Chase M, Manchester S, Majure L, Mavrodiev E** (2018) *Phylogeny and Evolution of the Angiosperms, Revised and Updated Edition*. University of Chicago Press
- Somerville C, Bauer S, Brininstool G, Facette M, Hamann T, Milne J, Osborne E, Paredes A, Persson S, Raab T, et al** (2004) Toward a systems approach to understanding plant cell walls. *Science* **306**: 2206–2211 doi: 10.1126/science.1102765
- Sonnhammer EL, von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175–182
- Steffen JG, Kang IH, Macfarlane J, Drews GN** (2007) Identification of genes expressed in the *Arabidopsis* female gametophyte. *Plant J* **51**: 281–292 doi: 10.1111/j.1365-313X.2007.03137.x
- Su S, Higashiyama T** (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. *Plant Reprod* **31**: 67–75 doi: 10.1007/s00497-018-0329-2
- Takeuchi H, Higashiyama T** (2016) Tip-localized receptors control pollen tube growth and LURE sensing in *Arabidopsis*. *Nature* **531**: 245–248 doi: 10.1038/nature17413

- Temple H, Mortimer JC, Tryfona T, Yu X, Lopez-Hernandez F, Sorieul M, Anders N, Dupree P** (2019) Two members of the DUF579 family are responsible for arabinogalactan methylation in *Arabidopsis*. *Plant Direct* **3**: 1–4 doi: 10.1002/pld3.117
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680 doi: 10.1093/nar/22.22.4673
- Tryfona T, Liang HC, Kotake T, Kaneko S, Marsh J, Ichinose H, Lovegrove A, Tsumuraya Y, Shewry PR, Stephens E, et al** (2010) Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohydr Res* **345**: 2648–2656 doi: 10.1016/j.carres.2010.09.018
- Tryfona T, Liang HC, Kotake T, Tsumuraya Y, Stephens E, Dupree P** (2012) Structural characterization of *Arabidopsis* leaf arabinogalactan polysaccharides. *Plant Physiol* **160**: 653–666 doi: 10.1104/pp.112.202309
- Tucker MR, Koltunow AMG** (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 doi: 10.1016/j.pbi.2013.11.015
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F** (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**: 1–12 doi: 10.1186/gb-2002-3-7-research0034
- Varner JE, Lin LS** (1989) Plant cell wall architecture. *Cell* **56**: 231–239 doi: 10.1016/0092-8674(89)90896-9
- Vizcay-Barrena G, Wilson ZA** (2006) Altered tapetal PCD and pollen wall development in the *Arabidopsis* ms1 mutant. *J Exp Bot* **57**: 2709–2717 doi: 10.1093/jxb/erl032
- Voragen AGJ, Coenen GJ, Verhoef RP, Schols HA** (2009) Pectin, a versatile polysaccharide present in plant cell walls. *Struct Chem* **20**: 263–275 doi: 10.1007/s11224-009-9442-z
- Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, et al** (2017) ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* **29**: 1806–1821 doi: 10.1105/tpc.17.00073
- Walter P, Johnson AE** (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* **10**: 87–119 doi:

10.1146/annurev.cb.10.110194.000511

- Willats WGT, McCartney L, Mackie W, Knox JP** (2001) Pectin: Cell biology and prospects for functional analysis. *Plant Mol Biol* **47**: 9–27 doi: 10.1023/A:1010662911148
- Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenführer J, von Mering C, Grossniklaus U** (2010) Arabidopsis Female Gametophyte Gene Expression Map Reveals Similarities between Plant and Animal Gametes. *Curr Biol* **20**: 506–512 doi: 10.1016/j.cub.2010.01.051
- Yadegari R, Drews GN** (2004) Female Gametophyte Development. *Plant Cell* **16**: S133–S141 doi: 10.1105/tpc.018192
- Yang J, Sardar HS, McGovern KR, Zhang Y, Showalter AM** (2007) A lysine-rich arabinogalactan protein in Arabidopsis is essential for plant growth and development, including cell division and expansion. *Plant J* **49**: 629–640 doi: 10.1111/j.1365-313X.2006.02985.x
- Yeats TH, Bacic A, Johnson KL** (2018) Plant glycosylphosphatidylinositol anchored proteins at the plasma membrane-cell wall nexus. *J Integr Plant Biol* **60**: 649–669 doi: 10.1111/jipb.12659
- Yu HJ, Hogan P, Sundaresan V** (2005) Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. *Plant Physiol* **139**: 1853–1869 doi: 10.1104/pp.105.067314
- Zhou X, Li W, Mabon R, Broadbelt LJ** (2017) A critical review on hemicellulose pyrolysis. *Energy Technol* **5**: 52–79 doi: 10.1002/ente.201600327

CHAPTER 2

UNRAVELLING THE ROLE OF
FASCICLIN-LIKE ARABINOGALACTAN
PROTEIN 5 IN ARABIDOPSIS
REPRODUCTION

ABSTRACT

Since seeds are essential units for plant propagation and the first step in their development is the formation of ovules, it is very important to study the genes involved in the reproductive process. A key factor controlling seed development is SEEDSTICK (STK), a MADS-box transcription factor that determines ovule identity, seed development and abscission. Recently, STK was shown to directly regulate many AGP genes. Chromatin immunoprecipitation sequencing data revealed that FLA5 is a putative target of STK.

FLAs belong to the AGP family and have 21 members in Arabidopsis. FLAs are a heterogeneous group characterized by having one or two FAS1 domains and one or two AGP-like glycosylated regions in their polypeptide core. FLAs are widely distributed in various cells and tissues. Besides, these glycoproteins are implicated in plant growth and development processes. However, only few FLA genes in Arabidopsis have been identified and functionally characterized.

In the past few years, two FLAs phylogenetic related to FLA5, FLA3 and FLA14, were shown to be involved in microspore development. As some redundancy is expected among AGPs, FLA5 might have a similar role in reproduction.

Therefore, to understand the role of FLA5 in this network, a characterization of *FLA5* was performed. The objective of the present work was to discover the *FLA5* expression pattern by qPCR and β -glucuronidase (GUS) promoter analysis and to characterize *fla5* mutants during sexual reproduction in *A. thaliana*.

FLA5 is highly expressed in pollen and PTs indicating that FLA5 might be involved in pollen development. However, T-DNA insertion mutants of *FLA5* showed no visible phenotype probably due to the redundancy related to AGPs. Interestingly, in *fla5* mutants, *FLA3* and *STK* were upregulated while *FLA14* was downregulated.

These results support and open doors for the study of new pathways in specific AGPs.

1. INTRODUCTION

1.1. Fasciclin-like arabinogalactan proteins – cell adhesion proteins

FLAs are a subclass of AGPs that have, in addition to predicted AGP regions, one or two FAS1 domains in their polypeptide core thought to be involved in protein-protein interactions (Johnson et al., 2003). Besides occurring in the FLAs family in higher plants, the FAS1 domain is also found in bacteria, fungi, algae, insects and animals (MacMillan et al., 2010). FAS1 domains contain 110 to 150 amino acids residues with low sequence similarity but with two highly conserved regions, H1 and H2 (approximately 10 amino acids each), at either end of the domain (Johnson et al., 2003). FAS1 domains are extracellular polypeptide modules that, at least in animal and algal cells, have experimentally confirmed roles in cell adhesion (Costa et al., 2019a).

Several FLAs have been identified from algae to angiosperms, including *Chondrus crispus* (1), *Chlamydomonas reinhardtii* (11), *Chara braunii* (24), *Marchantia polymorpha* (14), *Physcomitrella patens* (12), *Selaginella moellendorffii* (9), *Picea abies* (24) (He et al., 2019), *Pinus taeda* (1) (Yang et al., 2005), *Amborella trichopoda* (12), *Brachypodium distachyon* (23) (He et al., 2019), *Triticum aestivum* (34) (Faik et al., 2006), *Oryza sativa* (27) (Ma and Zhao, 2010), *Zea mays* (22) (Costa et al., 2019a), *A. thaliana* (21) (Schultz et al., 2002; Johnson et al., 2003), *Eucalyptus grandis* (18) (MacMillan et al., 2010; MacMillan et al., 2015), *Populus trichocarpa* (53) (Lafarguette et al., 2004; Zang et al., 2015a; Showalter et al., 2016; He et al., 2019), *Zinnia elegans* (1) (Dahiya et al., 2006), *Gossypium hirsutum* (19) (Huang et al., 2008), *Gossypium barbadense* (1) (Liu et al., 2013), *Brassica rapa* (33) (Jun and Xiaoming, 2012), *Quercus suber* (33) (Costa et al., 2019a) and *Cannabis sativa* (23) (Guerriero et al., 2017).

In *Arabidopsis*, the 21 FLAs can be divided into four groups (A-D) based on the number and position of FAS1 and AGP domains and the presence/absence of GPI anchor addition sequence (Fig. 1). Briefly, group A (FLAs 6, 7, 9, 11, 12 and 13) is characterized by FLAs with one FAS1 domain flanked by two AGP regions and a GPI-anchor signal at the C-terminus. Group B (FLAs 15, 16, 17 and 18) contains FLAs with two FAS1 domains flanking an AGP region. The FLAs from group C (FLAs 1, 2, 3, 5, 8, 10 and 14) have a variable domain structure with one or two FAS1 domains, one or two AGP regions and a C-terminal GPI anchor signal. The final group D is composed by 4 members (FLA 4, 19, 20 and 21) which share low similarities among them or with other FLAs (Johnson et al., 2003; Zang et al., 2015b).

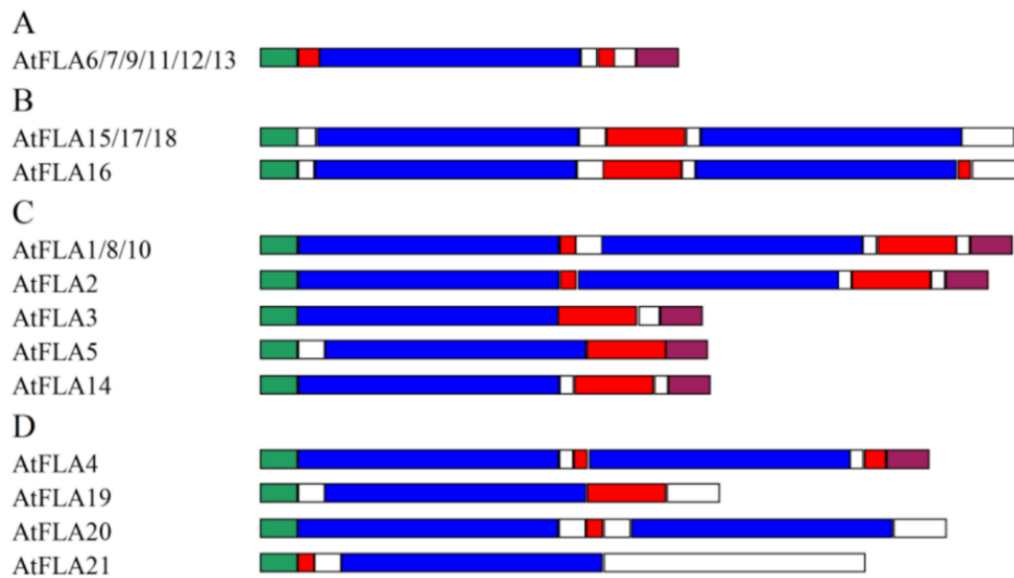


Fig. 1 - Schematic representation of Arabidopsis FLAs.

The colored regions indicate the N-terminal secretion signal (green), FAS1 domain (blue), AGP-like region (red) and GPI addition signal (purple). Taken from Zang et al. (2015b).

At the present, only the function of some Arabidopsis FLAs members have been characterized: FLA1 has a role in shoot development and formation of lateral roots (Johnson et al., 2011); FLA3 and FLA14 are involved in microspore development (Li et al., 2010; Miao et al., 2018); FLA4, also known as SALT OVERLY SENSITIVE 5 (SOS5) is implicated in wall remodeling and signaling, and is required for normal root growth and salt tolerance (Shi et al., 2003; Harpaz-Saad et al., 2011; Griffiths et al., 2014; Seifert et al., 2014; Basu et al., 2016; Xue et al., 2017; Turupcu et al., 2018); FLA11 and FLA12 contribute to plant stem strength (Ito et al., 2005; MacMillan et al., 2010); FLA9 is related to seed development (Cagnola et al., 2018). The role of other FLAs remains to be elucidated. Each new study supports that FLAs play important roles in plant growth and development. However, there is still much to discover about the mechanisms involved. Interestingly to this work, three FLAs participate in the sexual reproduction process: FLA9, FLA3 and FLA14. Drought stress reduced *FLA9* expression in siliques. The loss-of-function *fla9-1* mutant (SALK_051688C) with reduced *FLA9* expression showed increased abortion rates in siliques under non-stressful conditions and abortion rates similar to wild-type (wt) in stressful conditions. The gain-of-function *fla9-2* mutant (SALK_088265) with enhanced *FLA9* expression showed reduced abortion rate under drought conditions and increased abortion under control conditions. Seeds from both mutants revealed abnormal embryo development. This negative correlation between *FLA9* expression and abortion rate indicated a role for *FLA9* in seed development

(Cagnola et al., 2018). In addition, *FLA3* and *FLA14* were found to be specifically expressed in pollen grains and PTs (Li et al., 2010; Miao et al., 2018). *FLA3* localized in the plasma membrane as it has a GPI-anchoring signal sequence. A T-DNA insertion line (SALK_016582) showed no reduction in *FLA3* expression and no phenotype compared to wt. *FLA3*-RNAi and overexpression lines displayed shorter mature siliques with low seed set compared with wt plants. RNAi transgenic plants had 50% of non-viable pollen and pollen abortion occurs during the transition from uninucleate microspores to bicellular pollens. Besides, the intine layer of these plants also appeared to have some abnormalities. Reciprocal crosses with RNAi pollen as male donor led to some sterility on seeds as a low number of PTs grew in wt pistils (Li et al., 2010). In the same way, *FLA14*-overexpressing plants also had approximately 40% of abnormal pollen grains leading to reduced fertility with shorter mature siliques and less seeds. Pollen abortion occurred at the uninucleate microspore stage. Moreover, a T-DNA insertion mutant for *FLA14* (CS1014037) caused precocious PT germination inside the mature anthers under high moisture conditions (Miao et al., 2018). Taken together these results, the authors suggested that both FLAs are involved in microspore development.

1.2. Why FLA5? – the chosen one

Recently, Chromatin immunoprecipitation sequencing data for SEEDSTICK (STK; At4g09960) in siliques from 1 to 6 days after fertilization was obtained in Prof. Lucia Colombo laboratory (University of Milan, Italy). This experiment revealed that FLA5 is one of the direct AGP targets of STK during seed development (unpublished data).

STK is a MADS-box transcription factor known to act as the master regulator of seed production. STK is involved in the regulation of ovule development and attraction of the PTs to the embryo sac. Mature seeds of *stk* mutant do not detach from the mother plant siliques, the funiculus is larger than wt and seeds are smaller, indicating that STK also plays a role during seed development (Pinyopich et al., 2003; Mizzotti et al., 2012). STK is expressed in the ovules integuments and funiculus (Fig. 2) (Mizzotti et al., 2014).

Evidences point out that FLA5 may have a role during sexual reproduction as it is a target of STK (unpublished data).

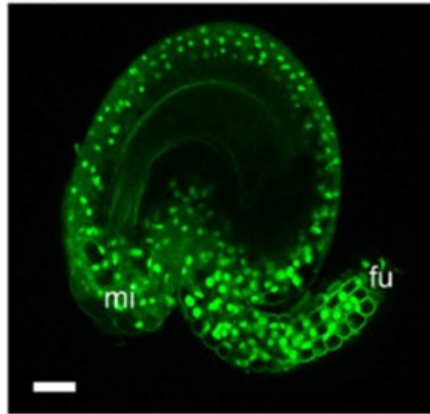


Fig. 2 - Expression pattern of *pSTK::STK-GFP* after fertilization.

STK, the ovule identity factor, is expressed in the integument tissues and funiculus of developing seeds. Retrieved from Mizzotti et al. (2014).

1.3. Objectives

Double fertilization is a characteristic process of angiosperms. AGPs are involved in this phenomenon from pollen and ovule development to seed formation. Hence, discovering new candidate targets in *Arabidopsis* is a key step to understand better sexual reproduction. The aim of this chapter was to gain further insights into the localization and function of a specific AGP, FLA5, along several reproductive tissues.

2. MATERIAL AND METHODS

2.1. Plant material and growth conditions

Arabidopsis thaliana (L.) Heynh. seeds, ecotype Columbia-0 (Col-0) were obtained from the Nottingham Arabidopsis Stock Centre (NASC), UK. T-DNA insertion lines for *FLA5* were selected using the SIGnAL database and were obtained from the NASC (Table 1). Wt and mutant seeds were sown directly on soil (COMPO SANA® Universal), kept for two days at 4°C in the dark to induce stratification and afterwards they were grown in an indoor growth facility with 50-60% relative humidity and a day/night cycle of 16 h light at 22°C and 8 h darkness at 18°C 16 h light (light intensity = 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$). For phosphinothricin acetyltransferase selection, the seedlings were sprayed with 200 mg/L of glufosinate ammonium (BASTA®; Bayer Crop Science), three times in nine days.

Table 1 - List of *A. thaliana* T-DNA insertion lines for *FLA5* used.

Accession number	Locus	Gene name	Mutant name	T-DNA insertion	NASC Stock ID
NM_119286	At4g31370	<i>FLA5</i>	<i>fla5-1</i>	SALK_139382C	N654924
			<i>fla5-2</i>	SALK_024772C	N682124

2.2. Plant marker lines generation

2.2.1. DNA amplification and purification

Arabidopsis genomic DNA (gDNA) was extracted according to Edwards et al. (1991). A DNA fragment of about 2000-2200 bp positioned upstream of the start codon of *FLA5* was amplified using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™) with genome specific primers with adaptors for the Gateway® Gene Cloning (Supplementary table 3). The respective annealing temperatures were estimated using the Tm Calculator (Thermo Fisher Scientific).

DNA analysis was made in 0.8% (w/v) agarose (Grisp) gel in 1X sodium boric acid (SB) buffer [10 mM NaOH (Panreac), pH 8.5 with H₃BO₃ (Merck)] with 0.5 $\mu\text{g/mL}$ Ethidium Bromide (Amresco). DNA Gel Loading Dye (6X) (Thermo Scientific™) was added to each sample prior to loading in gel. The electrophoretic separation was conducted in a horizontal electrophoresis system (Bio-Rad) at 200-240V for 10-20 min using GeneRuler DNA Ladder Mix (Thermo Scientific™) as a molecular weight marker.

The amplified DNA fragments were purified from the agarose gel. To prevent DNA damage caused by exposure to UV rays, all the desired bands were quickly excised from the agarose gel. The DNA was recovered using GeneJET Gel Extraction Kit (Thermo

Scientific™) according to the manufacturer's instructions. The quantification of DNA was performed in a NanoDrop™ One UV-Vis Spectrophotometer (Thermo Scientific™).

2.2.2 Cloning

Invitrogen™ Gateway Recombination Cloning Technology (Fig. 3) was performed to obtain the *FLA5_{prom}:GUS* construct. The *FLA5* promoter fragment was cloned into the entry vector pDONR207™ (Invitrogen™) through a BP recombination reaction [Gateway™ BP Clonase™ II Enzyme mix (Invitrogen™)] and then subcloned into the destination vector pBGWFS7 (Karimi et al., 2002) vector by an LR recombination reaction [Gateway™ LR Clonase™ II Enzyme mix (Invitrogen™)]. All constructs were confirmed by DNA sequencing and the sequencing data were analyzed using Geneious Prime (<https://www.geneious.com>).

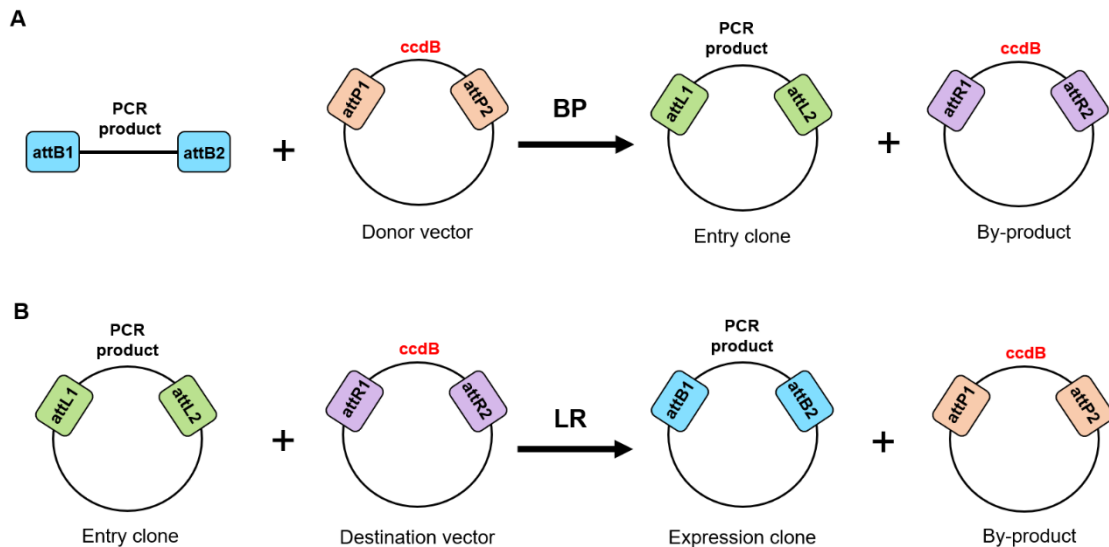


Fig. 3 - Schematic representation of Invitrogen™ Gateway Recombination Cloning Technology.

(A) The BP reaction facilitates recombination of an *attB*-flanked (blue) PCR product with an *attP*-containing (orange) donor vector to create an *attL*-containing (green) entry clone. (B) The LR Reaction facilitates recombination of an *attL* substrate (entry clone) with an *attR* (violet) substrate (destination vector) to create an *attB*-containing expression clone. The *ccdB* gene (red) allows negative selection of the donor and destination vectors in *E. coli* following recombination and transformation. Adapted from Gateway® Technology (2003).

2.2.3. Bacteria transformation

2.2.3.1. Preparation of electrocompetent cells

A small aliquot of bacteria (*E. coli* DH5α or *A. tumefaciens* GV3101) was inoculated into 5 mL and 100 mL of liquid LB media (BD Difco™) 2.5% (w/v), respectively. The *E. coli* culture was incubated overnight at 37°C shaking and the *A. tumefaciens* was incubated

with the appropriate antibiotics (gentamicin 30 µg/mL and rifampicin 20 µg/mL) for two days at 28°C shaking. Then, 5 mL of the starter culture was inoculated in 250 mL of LB media (BD Difco™) 2.5% (w/v). The cultures were incubated at 37°C shaking until OD₆₀₀ reached 0.3-0.4 (*E. coli*) or at 28°C shaking until OD₆₀₀ reached 0.5-1.0 (*A. tumefaciens*) and immediately placed on ice. The cultures were split into 50 mL ice cold falcon tubes (VWR) and the cells were harvested at 3,000 x g for 10 min at 4°C. The supernatant was decanted and each pellet was resuspend in ice cold sterile water. The centrifugation step was repeated as done before twice more for a total of three washes. In the last centrifugation, the pellet was resuspended in ice cold 10% (v/v) sterilized glycerol (Sigma-Aldrich), the cells were aliquoted into ice cold sterile 1.5 mL centrifuge tubes (Eppendorf) that were snap freeze with liquid nitrogen and stored at -80°C.

2.2.3.2. Electroporation

2 µL of BP or LR reaction was gently added to an *E. coli* DH5α or *A. tumefaciens* GV3101 shot containing 50 µL of electrocompetent cells, previously thawed on ice for 10 min and then transferred to an electroporation cuvette. The cuvette was placed in the MicroPulser™ electroporator (Bio-Rad) and an electrical shock was delivered to the cells. 350 µL of liquid LB medium (BD Difco™) 2.5% (w/v) was immediately added for cell recovery and followed by incubation (incubator INCU-Line, VWR) at 37°C for 45 min for *E. coli* or 28°C for 2 h for *A. tumefaciens*. 300 µL of cells were plated in LB agar, Miller (BD Difco™) 4% (w/v) with the corresponding antibiotics (Table 2) and left at 37°C overnight for *E. coli* or 28°C for 2 days for *A. tumefaciens* in the incubator (Shaking Incubator, SHEL LAB).

Table 2 - List of antibiotics used in the constructs generation in *E. coli* and *A. tumefaciens*.

Bacteria	Construct	Antibiotics
<i>E. coli</i>	<i>pDONR207:FLA5_{prom}</i>	Gentamicin (15 µg/mL)
	<i>pBGWFS7:FLA5_{prom}</i>	Spectinomycin (100 µg/mL)
<i>A. tumefaciens</i>	<i>pBGWFS7:FLA5_{prom}</i>	Gentamicin (30 µg/mL)
		Rifampicin (20 µg/mL)
		Spectinomycin (100 µg/mL)

2.2.4. Restriction digestion analysis

The presence and the orientation of *FLA5* promoter in the entry clone was confirmed with a restriction digestion analysis. EcoRV (Thermo Scientific™) was used to cut *pDONR207:FLA5_{prom}* in line with the manufacturer's procedure. The digestion fragments

(2026 bp and 3557 bp) were separated by agarose gel electrophoresis as mentioned in the section 2.2.1. The DNA fragments were visualized in an UV light transilluminator (302-365 nm), images were acquired using a Kodak camera (Electrophoresis Documentation and Analysis System 120) and analyzed using the software Image Analysis Kodak 1D software.

The plasmid DNA of positive colonies carrying the entry and expression clones were sequenced and the sequencing data were analyzed using Geneious Prime (<https://www.geneious.com>). The bacterial colonies of interest were stored in 30% (v/v) glycerol (Sigma-Aldrich) at -80°C.

2.3. *In vitro* pollen germination assay

Flowers from plants carrying the *FLA5_{prom}:GUS* construct were collected to germinate pollen *in vitro*. Pollen was germinated as described in Costa et al. (2019b).

2.4. Detection of GUS activity

The T2 generation *FLA5_{prom}:GUS* transformants were used for GUS analysis. GUS assays were performed on inflorescences and germinated pollen as described in Pereira et al. (2016). A Zeiss Axio Imager M2 microscope equipped with differential DIC optics was used. Images were captured with a Zeiss Axiocam 105 color camera using Zen 2011 (Zeiss) software.

2.5. Reciprocal crosses

For reciprocal crosses, plants were genotyped before crossing. Ten closed flower buds at stages 11-12, according to Smyth et al. (1990) were emasculated 24 h before hand pollination with fresh pollen collected from flowers at stage 13. After 24 h, pistils were collected and analyzed as described in Chapter I in 2.5.2. After 10 days, siliques were collected and analyzed as defined in Chapter I in 2.5.4.

2.6. RNA extraction, cDNA synthesis and qPCR

For the identification of knockout plants for *FLA5*, RNA extraction, cDNA synthesis and qPCR was performed according to Pereira et al. (2016) with the following modifications. RNA was extracted from 5-week-old inflorescences. cDNA was amplified in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using the primers listed in (Supplementary table 2). Real-time PCRs were run in triplicate. Data was analyzed with CFX Maestro™ Software (Bio-Rad).

FLA3, *FLA14* and *STK* expression on *fla5* mutants was revealed as described above using the primers listed in Supplementary table 2.

2.7. Co-expression data

Based on RNA-sequencing (RNA-seq) data (Klepikova et al., 2016), a tool was developed at professor Matthew Tucker laboratory. This tool allows to find genes putatively co-expressed with the gene of interest, based on the similarity of the expression profile.

2.8. Protein and genomic sequences

TAIR was used to identify the AGI codes and the accession numbers of Arabidopsis FLAs genes. Genomic and protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) (At5g55730, NP_200384; At4g12730, NP_193009; At2g24450, NP_180021; At3g46550, NP_190239; At4g31370, NP_194865; At2g20520, NP_565475; At2g04780, NP_849935; At2g45470, NP_566043; At1g03870, NP_563692; At3g60900, NP_191649; At5g03170, NP_195937; At5g60490, NP_200857; At5g44130, NP_199226; At3g12660, NP_187872; At3g52370, NP_190805; At2g35860, NP_850253; At5g06390, NP_196257; At3g11700, NP_566398; At1g15190, NP_172971; At5g40940, NP_198910; At5g06920, NP_196309).

Arabidopsis FLAs homologue protein sequences of the species *Physcomitrella patens* (Pp3c16_8540; Pp3c25_5850; Pp3c26_5590; Pp3c4_16840; Pp3c7_430), *Selaginella moellendorffii* (SMO114G0121; SMO137G0289; SMO143G0148; SMO174G0041; SMO203G0171; SMO234G0076; SMO236G0071; SMO236G0088; SMO356G0520), *Hordeum vulgare* (HvU0036G0897; HvU0036G1574; HvU0036G2673; HvU0037G0321; HvU0037G0639; HvU0037G0869; HvU0037G1715; HvU0037G2289; HvU0037G2665; HvU0037G3263; HvU0037G3407; HvU0037G3448; HvU0038G1550; HvU0038G1615; HvU0040G0134; HvU0041G0911; HvU0041G1413; HvU0041G3187; HvU0042G1219; HvU0045G0845; HvU0798G0161; HvU1208G0085) and *Populus trichocarpa* (Potri.001G037800; Potri.001G306800; Potri.001G320800; Potri.001G367900; Potri.002G223300; Potri.004G210600; Potri.006G129200; Potri.006G174900; Potri.006G200300; Potri.006G276200; Potri.008G012400; Potri.008G127500; Potri.008G128200; Potri.009G012100; Potri.009G012200; Potri.010G244900; Potri.011G093500; Potri.012G006200; Potri.012G015000; Potri.012G127900;

Potri.013G014200; Potri.013G120600; Potri.013G151300; Potri.013G151366;
 Potri.013G151432; Potri.013G151500; Potri.013G152200; Potri.014G071700;
 Potri.014G162900; Potri.014G168100; Potri.015G013300; Potri.015G129400;
 Potri.016G066500; Potri.016G088700; Potri.017G111600; Potri.018G005100;
 Potri.018G097000; Potri.019G002300; Potri.019G008400; Potri.019G093300;
 Potri.019G120800; Potri.019G120900; Potri.019G121100; Potri.019G121200;
 Potri.019G121300; Potri.019G122600; Potri.019G122800; Potri.019G123000;
 Potri.019G123100; Potri.019G123200; Potri.019G125200; Potri.T130300;
 Potri.T172033) were retrieved from PLAZA (Monocots PLAZA 4.5) from homologous
 gene family HOM04M000154, HOM04M001821, HOM04M009341, HOM04M010567,
 HOM04M013112, using FLA5, FLA15, FLA19, FLA20 and FLA21 as baits, respectively
 (Van Bel et al., 2018).

2.9. 3D protein structure

Homology modeling of AtFLA5 was performed on MODELLER (Webb and Sali, 2016). The
 structure alignment was achieved using MPI Bioinformatics Toolkit – HHpred
 (Zimmermann et al., 2018). PyMOL (Schrödinger, 2015) was used to visualize the 3D
 structure.

3. RESULTS

3.1. Phylogenetic analysis of FLAs family

The Arabidopsis genome has 21 FLAs encoding genes divided into four groups (Johnson et al., 2003; Zang et al., 2015b). Seven FLAs have been identified and characterized (FLA1, Johnson et al., 2011; FLA3, Li et al., 2010; SOS5, Shi et al., 2003; Basu et al., 2016; Turupcu et al., 2018; FLA9, Cagnola et al., 2018; FLA11, Ito et al., 2005; MacMillan et al., 2010; FLA12, MacMillan et al., 2010; FLA14, Miao et al., 2018).

The AGI codes of FLA genes were identified in TAIR and the corresponding nucleotide and amino acid sequences were retrieved from the NCBI. Full length of genomic DNA sequences of FLAs varied between 768 bp (*FLA14*) and 3292 bp (*FLA17*) (Table 3). The open reading frames (ORF) diverged between 741 bp (*FLA11*) to 1389 bp (*FLA18*) and, consequently, the proteins encoded have 246 to 462 amino acids. Furthermore, the majority of FLAs had only one transcription variant except *FLA1*, *FLA7* and *FLA15* that have two.

Table 3 - FLAs coding region information in Arabidopsis.

Group A is highlighted in orange, group B in green, group C in blue and group D in pink.

Gene	Locus	Genomic DNA (bp)	Coding region (bp)	Amino acids	Splice variants	Group
<i>FLA1</i>	At5g55730	2498	1275	424	AT5G55730.1 AT5G55730.2	C
<i>FLA2</i>	At4g12730	1854	1212	403	AT4G12730.1	C
<i>FLA3</i>	At2g24450	1190	843	280	AT2G24450.1	C
<i>FLA4/SOS5</i>	At3g46550	1914	1263	420	AT3G46550.1	D
<i>FLA5</i>	At4g31370	1031	837	278	AT4G31370.1	C
<i>FLA6</i>	At2g20520	1009	744	247	AT2G20520.1	A
<i>FLA7</i>	At2g04780	1739	765	254	AT2G04780.1 AT2G04780.2	A
<i>FLA8</i>	At2g45470	1846	1263	420	AT2G45470.1	C
<i>FLA9</i>	At1g03870	1164	744	247	AT1G03870.1	A
<i>FLA10</i>	At3g60900	1948	1269	422	AT3G60900.1	C
<i>FLA11</i>	At5g03170	908	741	246	AT5G03170.1	A
<i>FLA12</i>	At5g60490	1371	750	249	AT5G60490.1	A
<i>FLA13</i>	At5g44130	1225	744	247	AT5G44130.1	A
<i>FLA14</i>	At3g12660	768	768	255	AT3G12660.1	C
<i>FLA15</i>	At3g52370	2487	1311	436	AT3G52370.1 AT3G52370.2	B
<i>FLA16</i>	At2g35860	2360	1338	445	AT2G35860.1	B
<i>FLA17</i>	At5g06390	3292	1377	458	AT5G06390.1	B
<i>FLA18</i>	At3g11700	2857	1389	462	AT3G11700.1	B
<i>FLA19</i>	At1g15190	1221	747	248	AT1G15190.1	D
<i>FLA20</i>	At5g40940	1300	1275	424	AT5G40940.1	D
<i>FLA21</i>	At5g06920	1062	1062	353	AT5G06920.1	D

The protein sequences of Arabidopsis FLAs were aligned with Clustal W (Thompson et al., 1994) to obtain a sequence identity matrix (Fig. 4). The three pairs of proteins FLA8/FLA10, FLA15/FLA16 and FLA17/FLA18 shared the highest percentage of sequence identity (approximately 85%, 86% and 84%, respectively). Then, FLA9/FLA13, FLA15/FLA18, FLA15/FLA19, FLA16/FLA17 and FLA16/FLA18, shared almost 80% of sequence identity between each pair. Interestingly, FLA5 shared approximately 56% and 31% of amino acid identity with FLA3 and FLA14, respectively.

Protein	FLA1	FLA2	FLA3	SOS5	FLA5	FLA6	FLA7	FLA8	FLA9	FLA10	FLA11	FLA12	FLA13	FLA14	FLA15	FLA16	FLA17	FLA18	FLA19	FLA20	FLA21
FLA1																					
FLA2	49,5																				
FLA3	28,74	29,92																			
SOS5	28,35	27,37	22,05																		
FLA5	27,78	28,17	56,12	19,54																	
FLA6	25,33	27,03	12,04	27,85	9,43																
FLA7	27,63	25,45	10,91	30	13,89	35,84															
FLA8	40,79	39,39	31,64	31,33	29,13	28,25	29,02														
FLA9	26,84	28,25	10,19	27,27	11,32	56,07	33,77	28													
FLA10	38,74	40,31	33,46	32,63	29,41	25,56	28,7	84,78	26,22												
FLA11	23,89	29,22	12,5	27,15	12,75	44,84	33,05	30,45	45,61	28,77											
FLA12	24,34	28,64	16,98	28,83	15,38	42,54	35,9	28,83	40,61	28,18	53,19										
FLA13	24,24	26,46	8,57	28,18	9,71	56,07	32,31	28	79,1	27,11	46,02	39,57									
FLA14	30,45	35,66	36	23,46	30,52	14,44	13,04	39,27	14,29	39,52	19,54	12,36	12,5								
FLA15	17,9	20,25	17	18,05	15,54	20,83	18,65	18,6	22,16	17,99	21,88	22,28	20,73	17,3							
FLA16	17,63	19,58	17,12	17,15	15,29	21,24	20,1	17,72	23,08	18,02	23,32	22,68	24,23	17,43	85,78						
FLA17	16,96	20,65	17,23	17,05	14,72	20,53	19,9	17,11	22,92	17,06	23,68	21,99	22,51	16,8	77,14	76,24					
FLA18	17,01	19,82	15,15	17,38	12,6	19,79	17,95	15,98	22,16	16,81	22,4	21,76	22,28	17,81	77,19	75,85	83,92				
FLA19	18,66	20,97	11,96	23,12	10	22,7	21,13	22,12	18,95	24,29	23,28	20,97	19,37	12,05	16,02	17,58	16,2	16,02			
FLA20	17,09	19,82	12,82	19,2	11,21	21,35	21,62	21,43	22,16	22,65	25,56	22,16	21,97	13,91	17,45	17,21	16,74	18,06	23,53		
FLA21	16,24	15,98	12,26	19,21	13,46	21,03	15	17,91	23,2	17,86	20,71	16,59	21,35	15,05	15,38	16,48	17,78	17,13	22,6	24,05	

Fig. 4 - Percent of identity matrix between aligned proteins sequences of Arabidopsis FLAs.

The percent identity was calculated using Clustal W (Thompson et al., 1994). The values in blue represent high sequence identity whereas the values in red represent low sequence identity.

FLAs aligned protein sequences were used to generate a neighbor-joining tree with MEGA X (Fig. 5) (Kumar et al., 2018). The 21 sequences clustered into four different groups (A, B, C and D). Many pairs of paralog genes emerged accordingly to the previous sequence identity matrix, such as FLA15/FLA16, FLA17/FLA18, FLA11/FLA12, FLA9/FLA13, FLA1/FLA2 and FLA8/FLA10. FLA5 was closely related to FLA3 and FLA14.

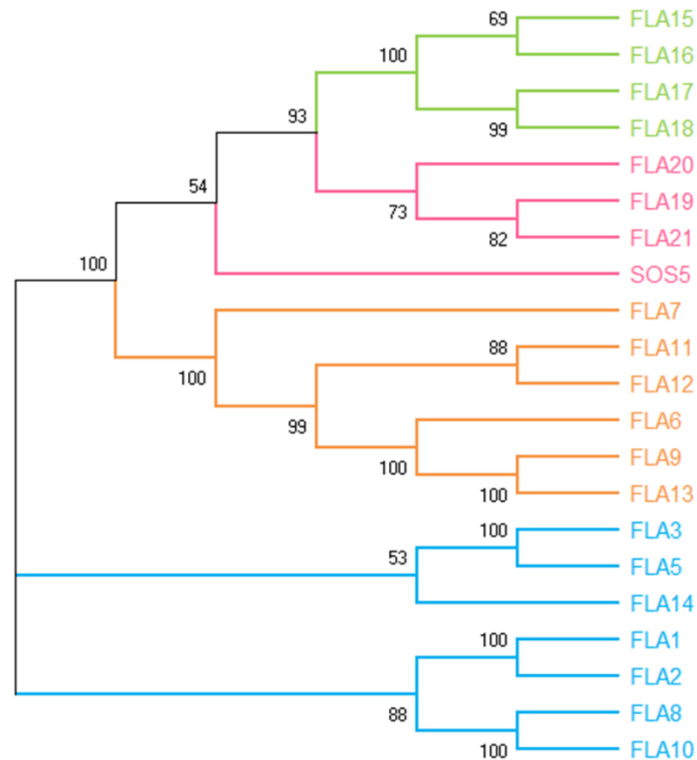


Fig. 5 - Phylogenetic tree of Arabidopsis FLAs.

The 21 FLAs cluster into four different groups: A (orange), B (green), C (blue) and D (pink). The phylogenetic tree was generated with MEGA X (Kumar et al., 2018) using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. This analysis involved 21 amino acid sequences. After deletion of ambiguous positions, there were a total of 510 positions in the final dataset.

FLAs from Arabidopsis were distributed along five different chromosomes, with two genes on chromosome 1 (*FLA9* and *FLA19*) and 4 (*FLA2* and *FLA5*), five genes on chromosome 2 (*FLA3*, *FLA6*, *FLA7*, *FLA8* and *FLA16*) and 3 (*SOS5*, *FLA10*, *FLA14*, *FLA15* and *FLA18*), and 7 genes on chromosome 5 (*FLA1*, *FLA11*, *FLA12*, *FLA13*, *FLA17*, *FLA20* and *FLA21*) (Fig. 6). Besides, there was no evidence of clustering. Taking FLAs group C has an example: *FLA1* was located on chromosome 5, *FLA2* on chromosome 4, *FLA3* on chromosome 2, *FLA5* on chromosome 4, *FLA8* on chromosome 2, *FLA10* on chromosome 3 and finally *FLA14* on chromosome 3. FLAs seemed to be randomly distributed across the different Arabidopsis chromosomes.

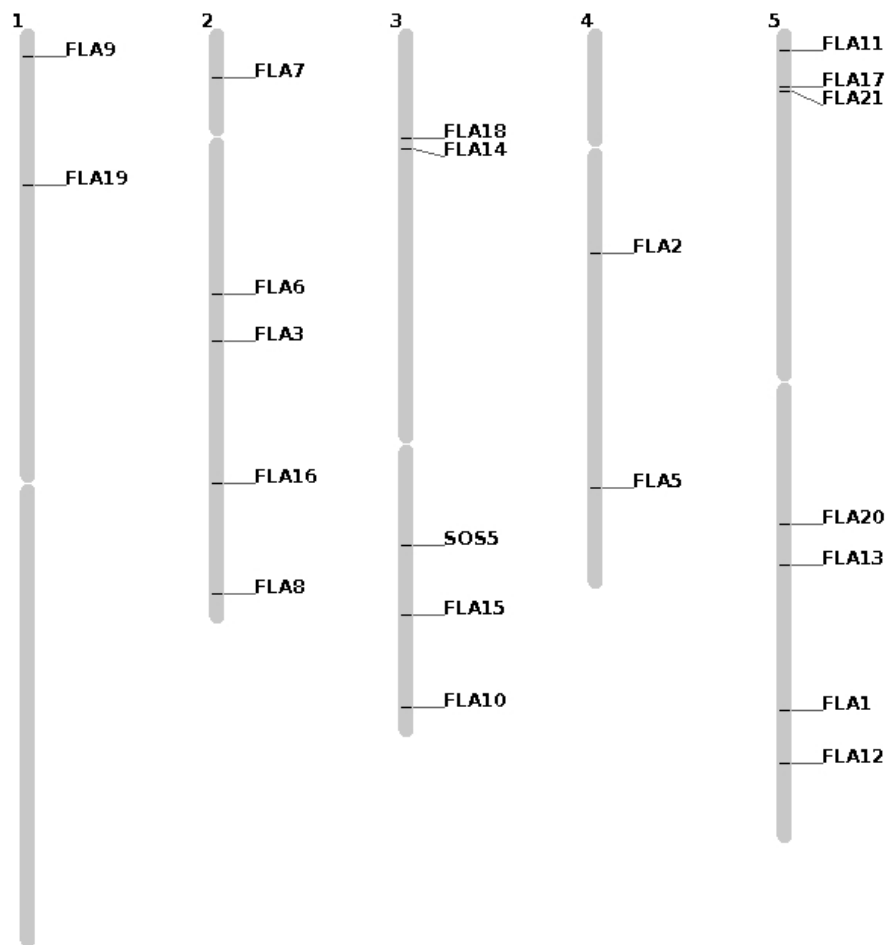
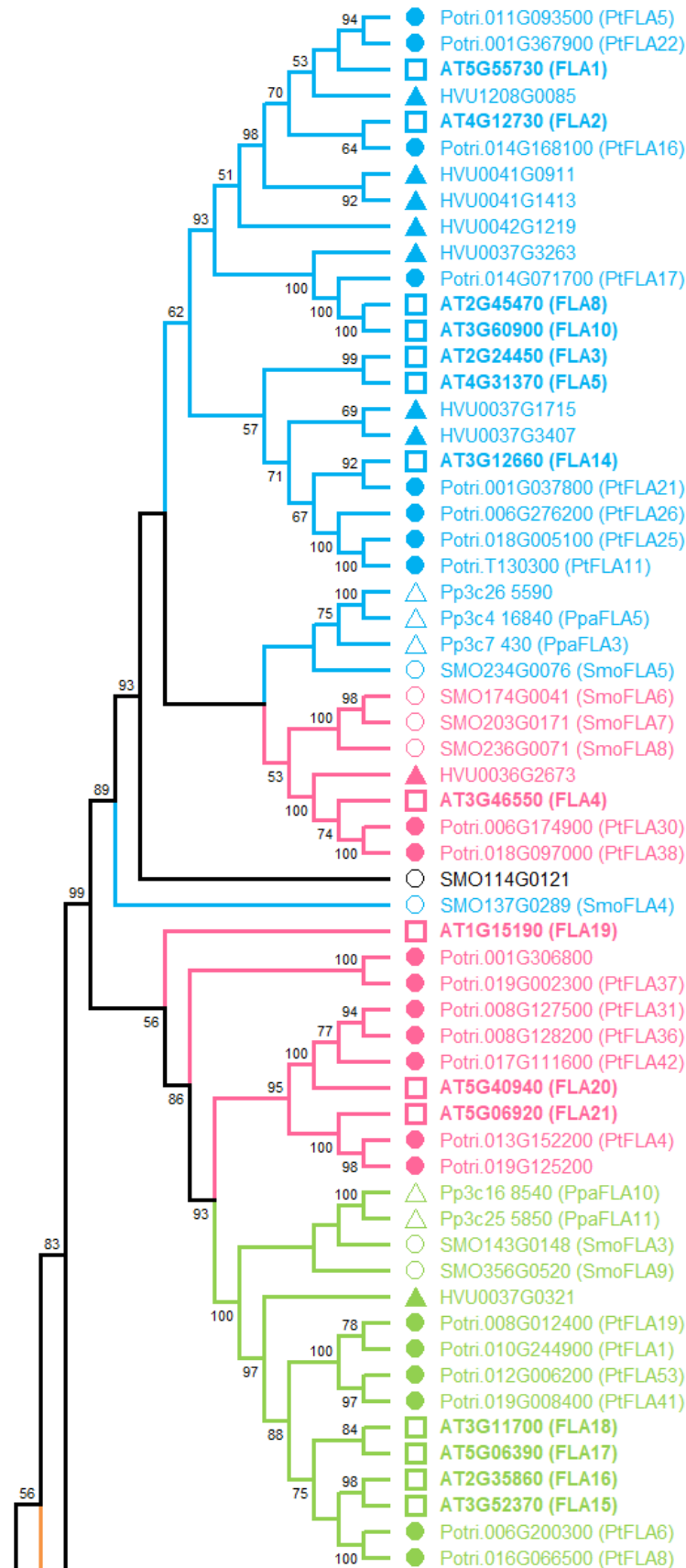


Fig. 6 - FLAs gene positions on *A. thaliana* chromosomes.

The localization of the chromosomes in Arabidopsis was obtained from the Chromosome Map Tool available at TAIR.

Moreover, to understand the evolutionary relationship between FLAs, a phylogenetic tree of FLAs from five plant species (*P. patens*, *S. moellendorffii*, *H. vulgare*, *P. trichocarpa* and *A. thaliana*) was generated with MEGA X (Kumar et al., 2018) using the Neighbor-Joining method with the bootstrap test (1000 replicates) (Fig. 7). The homologue sequences of Arabidopsis FLAs were retrieved from Monocots PLAZA 4.0 (Van Bel et al., 2018). However, three proteins (Potri.T172033, Hv004060134 and SMO23660088) were not included in this analysis as their sequence did not aligned with majority. The nomenclature and classification of FLA genes for *P. patens*, *S. moellendorffii* and *P. trichocarpa* was performed according to Showalter et al. (2016) and He et al. (2019).

The number of putative FLA genes varies across the different plant species. *P. patens* has 5 FLA genes while *S. moellendorffii* has 9 FLA genes (Fig. 7). Spermatophyte plants have a highest number of FLA genes. *A. thaliana*, *H. vulgare* and *P. trichocarpa* have



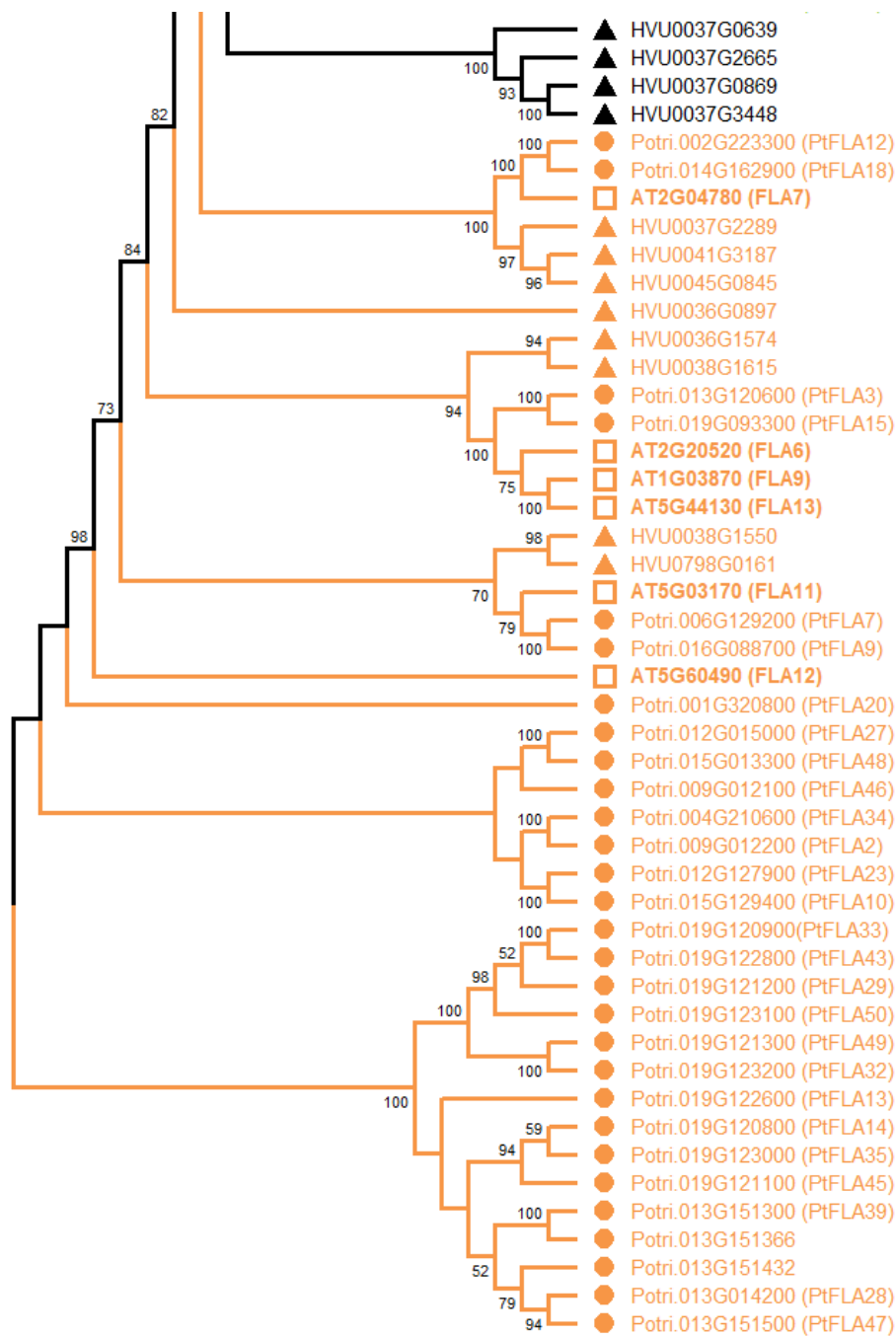


Fig. 7 - Phylogenetic tree of Arabidopsis FLAs and putative homologues in *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa*.

The phylogenetic tree was generated with MEGA X (Kumar et al., 2018) using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. This analysis involved 107 amino acid sequences. After deletion of ambiguous positions, there were a total of 737 positions in the final dataset. The tree is divided into four main clades: group A (orange), group B (green), group C (blue) and group D (pink). Each plant species is identified with a different shape (Δ, *P. patens*; ○, *S. moellendorffii*; ▲, *H. vulgare*; ●, *P. trichocarpa*; □, *A. thaliana*). FLAs from *A. thaliana* are highlighted in bold.

21, 22 and 53 FLA genes, respectively. Overall, the number of FLA genes increased from bryophytes to flowering plants.

Previous studies indicate that the 21 FLAs can be divided into four groups (Johnson et al., 2003). In this analysis, FLAs from *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa* were also classified into four clades (A-D) (Fig. 7). FLAs from group C formed three distinct subclades (Fig. 7). One comprised FLAs from Arabidopsis with only one FAS1 domain (AtFLA3, AtFLA5 and AtFLA14), the other with FLAs with two FAS1 domains (AtFLA1, AtFLA2, AtFLA8 and AtFLA10) and lastly, one subclade with three FLAs from *P. patens* (Pp3c26_5590, PpFLA5 and PpFLA3) and one FLA from *S. moellendorffii* (SmoFLA5). SmoFLA4 did not follow this trend. The FLA group D was also divided into two subclades composed by the putative orthologs of AtFLA4 and AtFLA19-21 (Fig. 7). FLAs from group B formed a unique tight cluster (Fig. 7). The group A formed three clusters with putative orthologs of AtFLA7, AtFLA6/9/13 and AtFLA11, and two clusters of FLAs from *P. trichocarpa*. SMO114G0121, HVU0037G0639, HVU0037G265, HVU0037G0869 and HVU0037G3448 were not identified or classified previously. Therefore, it is still necessary to attribute a group to these putative genes.

According to the phylogenetic tree, it can be inferred that AtFLA14 may be quadruplicated in *P. trichocarpa* (PtFLA21, PtFLA26, PtFLA25 and PtFLA11), AtFLA4 may be triplicated in *S. moellendorffii* (SmFLA6, SmFLA7 and SmFLA8), AtFLA20 may be triplicated in *P. trichocarpa* (PtFLA31, PtFLA36 and PtFLA42) and AtFLA7 may be triplicated in *H. vulgare* (HVU0037G2289, HVU0041G3187 and HVU0045G30845). Several FLAs may be duplicated in *P. trichocarpa* [AtFLA1 (PtFLA2 and PtFLA22); AtFLA4 (PtFLA30 and PtFLA38); AtFLA6 (PtFLA3 and PtFLA15); AtFLA7 (PtFLA12 and PtFLA18); AtFLA11 (PtFLA7 and PtFLA9); AtFLA19 (PtFLA37 and Potri.001G306800); AtFLA21 (PtFLA4 and Potri.019G125200)] and in *H. vulgare* [AtFLA2 (HVU0041G0911 and HVU0041G1413); AtFLA6 (HVU0036G1574 and HVU0038G1615); AtFLA11 (HVU0038G1550 and HVU0798G0161); AtFLA14 (HVU0037G1715 and HVU0037G3407)].

3.2. Bioinformatic analysis of FLA5

Focusing only on *FLA5*, the gene comprised 1031 bp with only one exon with an 837-bp long ORF encoding a 278-amino-acid protein (Table 3). Examining the protein composition on ExPASy (Gasteiger et al., 2003), FLA5 was rich in Ala (10.8%), Ser (10.4%), leucine (10.1%), Pro (9.0%) and Thr (6.8%). Moreover, FLA5 theoretical pI was 5.93 and its molecular weight was 29.9 kDa. Further, several tools were run to predicted

FLA5 domains and specific recognition sites. SignalP 5.0 (Almagro Armenteros et al., 2019) predicted a N-terminal signal peptide cleavage site between position 24 and 25 while Big-PI plant predictor (Eisenhaber et al., 2003) identified a potential C-terminal GPI-modification site on Ser255. Profilescan (Sigrist et al., 2010) recognized a FAS1 domain from position 24 to 141 and a Pro-rich domain between position 180 to 237. Besides, 4 potential N-glycosylation sites were also predicted: position 26, 74, 126 and 159. Therefore, according to the analysis, FLA5 was predicted to have an N-terminal signal peptide, a FAS1 domain, a Pro-rich domain and a GPI-anchor addition site to attach a GPI anchor after proteolytic cleavage of a C-terminal propeptide (Fig. 8).

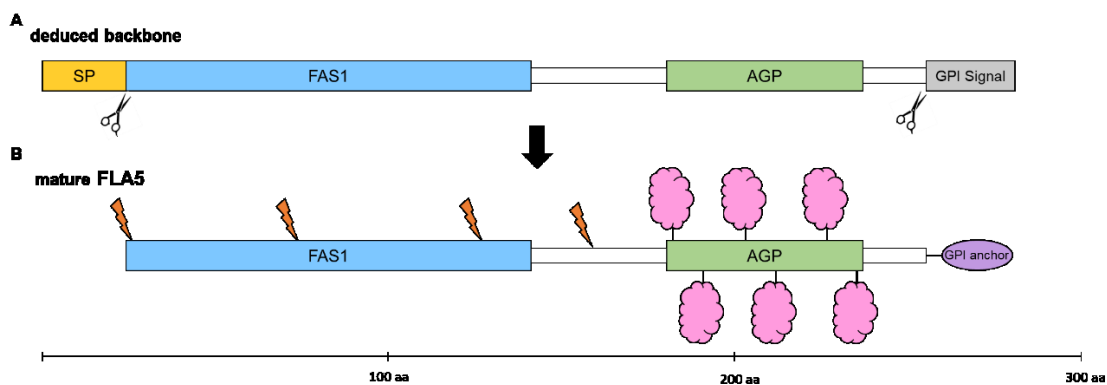


Fig. 8 - Schematic representation of FLA5 before and after post-translational modifications.

The colored regions indicate the N-terminal secretion signal (yellow), fasciclin-like domain (blue), AGP glycosylated region (green), C-GPI anchor signal (grey), N-glycosylation sites (orange), arabinogalactan polysaccharides (pink) and GPI anchor (purple).

Besides, to predict if FLA5 contained TMDs, the protein sequence was run in TMHMM 2.0 (Sonnhammer et al., 1998). A TMD was predicted for FLA5 between amino acid number 7 to 29.

In addition, to deduce the subcellular localization of FLA5, three platforms were used: TargetP 1.1 (Emanuelsson et al., 2000), SUBA (Hooper et al., 2017) and Aramemnon (Schwacke et al., 2003). TargetP 1.1 and Aramemnon identified a signal peptide sequence and located FLA5 in the secretory pathway. SUBA localized FLA5 in the plasma membrane.

As there is not a FLA5 protein structure in the Protein Data Bank (PDB), the MPI Bioinformatics Toolkit – HHpred (Zimmermann et al., 2018) was used to align the FLA5 protein sequence and to detect a protein remote homology. From the list of HHpred alignments, a fasciclin I structure from *Drosophila melanogaster* (1O70_A) with 99.26% of probability was chosen. Then, using the MODELLER program (Webb and Sali, 2016), a FLA5 model was simulated based on the model of the fasciclin I domain of *D.*

melanogaster. Finally, the model was inserted in the software PyMOL (Schrödinger, 2015) which allows the visualization of the tridimensional protein structures (Fig. 9).

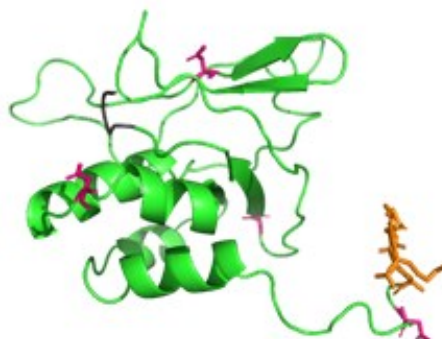


Fig. 9 - Representation of FLA5 tridimensional structure predicted by the software PyMOL (Schrödinger, 2015). The orange amino acids correspond to the signal peptide, the fasciclin-like domain is marked with green and the four pink amino acids correspond to the N-glycosylation sites.

3.3. Expression pattern of *FLA5*

To determine the expression pattern of *FLA5*, a qPCR analysis was carried out in several tissues of *Arabidopsis* (Supplementary fig. 1). The qPCR fragment of *FLA5* was purified and quantified by HPLC (Supplementary fig. 2). The size in base pairs and identity of qPCR product was confirmed by sequencing (Supplementary fig. 3). In Fig. 10A, the qPCR analysis was expressed as a ratio of maximal levels observed. *FLA5* was transcribed at its maximal rate in the siliques. Lower levels of expression were found in the seedlings, roots, stems, inflorescences, juvenile and cauline leaves.

Available RNA-seq data on *FLA5* expression was also studied (Klepikova et al., 2016). According to the RNA-seq data, siliques were the tissues where *FLA5* was most transcribed (Fig. 10B). RNA-seq barely detected *FLA5* transcripts in leaf blades young and intermediate 2. Low levels were detected in roots, mature leaves and internodes. However, high levels of *FLA5* transcripts were detected in flower 1.

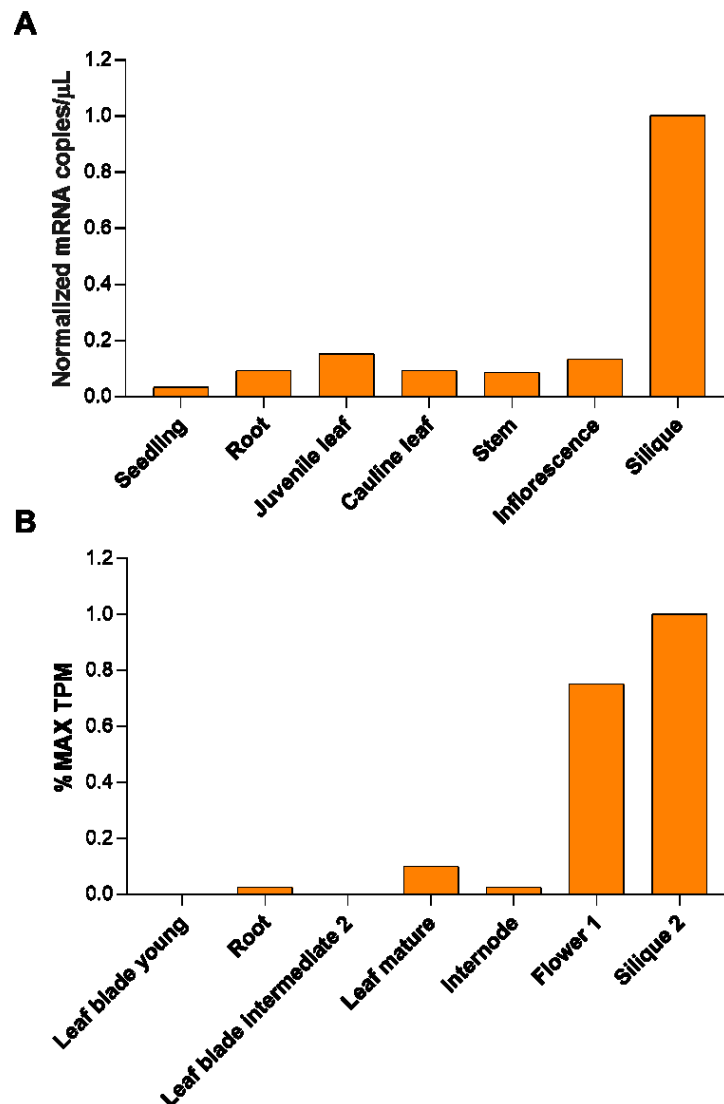


Fig. 10 - Expression pattern of *FLA5*.

(A) Normalized transcript levels of *FLA5* in a range of tissues. Means of three biological replicates are expressed as a ratio of maximal levels observed. Values of 1.0 therefore indicate the tissue in which *FLA5* is transcribed at its maximal rate. (B) RNA-seq data on *FLA5* expression (Klepikova et al., 2016). Number of transcripts per million (TPM) of *FLA5* expressed as a ration of maximal levels observed.

Additionally, the *GUS* reporter gene was placed under the control of *FLA5* promotor. A 2213 bp fragment positioned upstream of the start codon of *FLA5* was cloned into the entry vector pDONR207™ (Invitrogen™) and then subcloned into the destination vector pBGWFS7 containing the *GUS* gene (Karimi et al., 2002). To verify the presence and the orientation of *FLA5* promoter in *pDONR207:FLA5prom*, a restriction digestion analyses using EcoRV enzyme was performed (Fig. 11). The digestion fragments (2026 bp and 3557 bp) had the expected sizes.

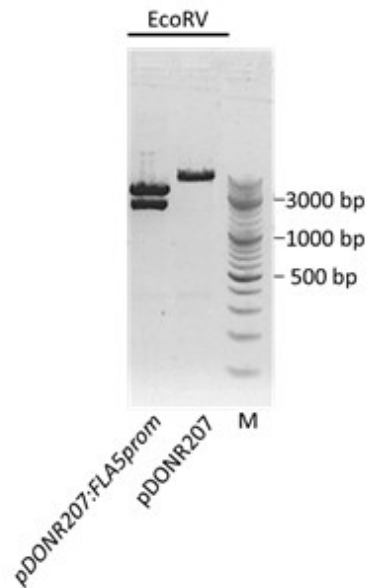


Fig. 11 - Restriction digestion analysis of *pDONR207:FLA5prom* with *EcoRV* enzyme. *pDONR207:FLA5prom* was cut with *EcoRV*. The resulting DNA fragments had the expected sizes (2026 bp and 3557 bp). *pDONR207* was also cut with *EcoRV* as control resulting in one DNA fragment of 5585 bp. Abbreviations: M, DNA ladder; *pDONR207*, vector without insert.

The *FLA5prom:GUS* construct was confirmed by DNA sequencing (Supplementary fig. 3) and transformed into *Arabidopsis* plants by the *Agrobacterium*-mediated floral dip. The T2 transgenic *Arabidopsis* plants containing the transgene were subjected to histochemical staining for GUS activity. The blue color observed in the tissues indicates the active expression of *FLA5*.

During flower development, GUS activity driven by the *FLA5* promoter was detected in the anthers from stage 10, when microgametogenesis begin, to stage 15, when stigma extends above the long anthers (Fig. 12). *GUS* expression was detected in the pollen in mature (Fig. 13C) and young anthers (Fig. 13E). During stages 14 and 15, after anthesis, *GUS* expression was detected in the pollen at the stigma during germination and initiation of PT growth phases (Fig. 13A). No GUS activity was spotted in any reproductive tissue during early flower development. In the petals and sepals from stage 10 to stage 15, GUS activity was detected in the vascular tissues (Fig. 13B,D).

Pollen from *FLA5prom:GUS* plants was germinated *in vitro* and histochemical stained for GUS activity. GUS activity was detected in the germinating pollen (Fig. 13F) and in growing PTs (Fig. 13G) *in vitro*.

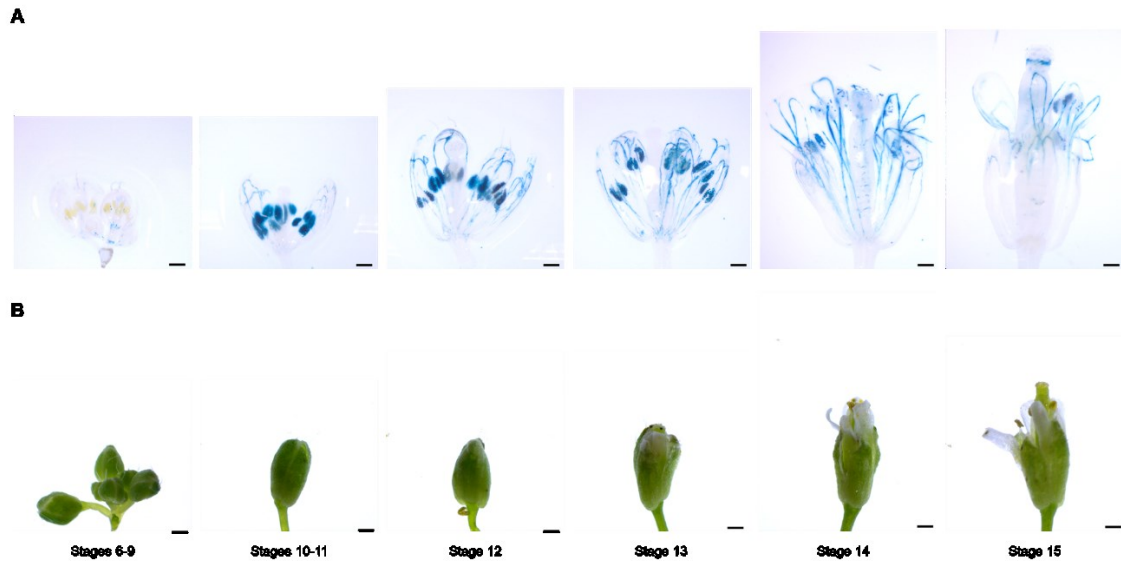


Fig. 12 - GUS expression analysis under *FLA5* promoter in during Arabidopsis flower development.

(A) Histochemical localization of GUS activity in transgenic Arabidopsis flowers expressing the *FLA5prom:GUS* fusion gene. (B) Wild-type Arabidopsis flowers in different stages of development (stages 6 to 15 according to Smyth et al. 1990)

All images were obtained with a stereo microscope. Scale bars: 0.2 mm.

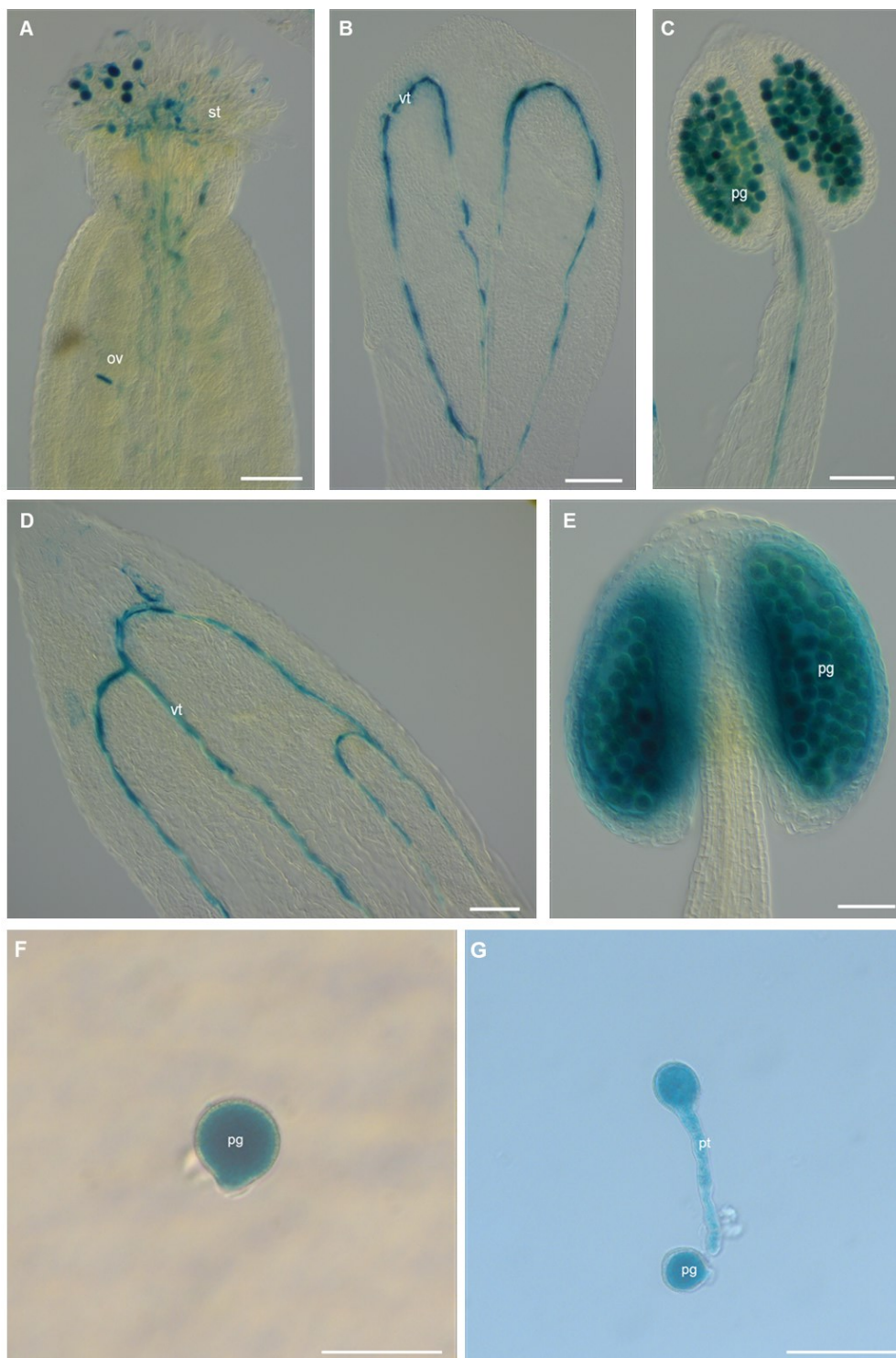


Fig. 13 - Histochemical localization of GUS activity in transgenic *Arabidopsis* tissues expressing the *FLA5*prom:GUS fusion gene.

(A) GUS activity driven by the *FLA5* promoter is detected in the pollen germinating in the stigma. GUS activity detected in the vascular tissues of sepals (B) and petals (D). Strong GUS activity detected in the pollen grains in mature (C) and young anthers (E). Detail of GUS activity detected in the germinating pollen grains (F) and in growing PTs (G) *in vitro*. All images were obtained with a DIC microscope. Abbreviations: ov, ovule; pg, pollen grain; pt, pollen tube; st, stigma; vt, vascular tissue. Scale bars: 100 µm (A-D), 50 µm (E) and 500 µm (F-G).

Taking another look in the RNA-seq data about *FLA5* expression (Klepikova et al., 2016), *FLA5* seems to be expressed in the flowers 4 to 14 (Fig. 14). These flowers are in order of position on an inflorescence from the oldest flower (flower 1) to the primordia buds (flowers +19). Therefore, *FLA5* is expressed from flower development stage 6 to stage 12 according to Smyth et al. (1990). Lower expressions were found in the male tissues (Fig. 14). All together these results suggest that *FLA5* may have a role in pollen development.

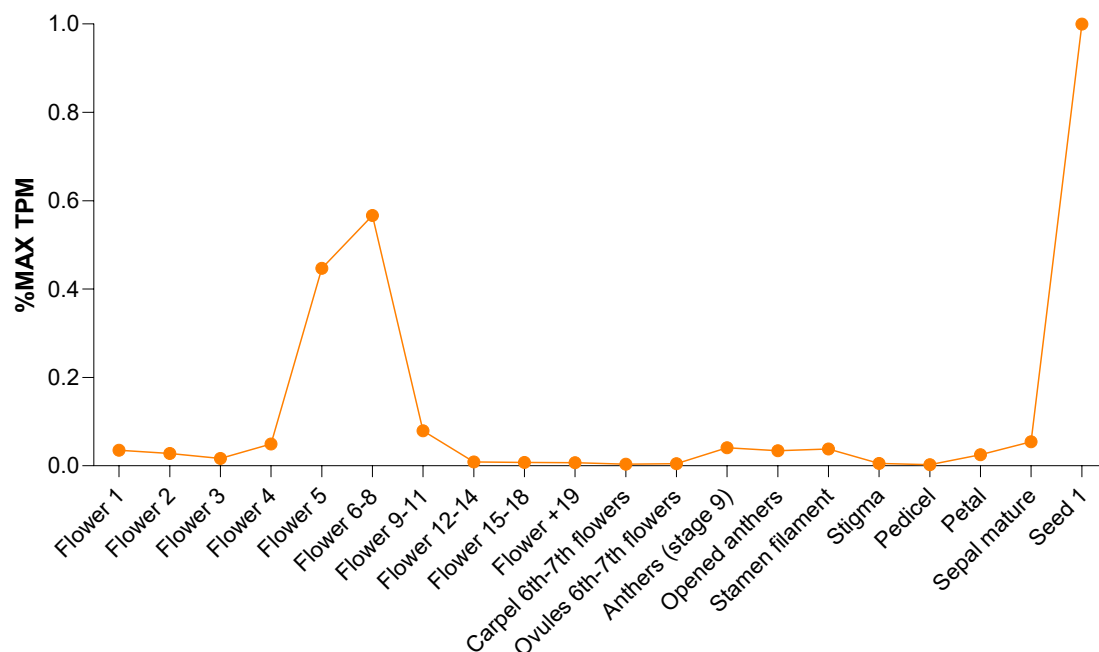


Fig. 14 - RNA-seq data on *FLA5* expression (Klepikova et al., 2016).

Number of TPM of *FLA5* expressed as a ration of maximal levels observed. The flower numbers are in order of position on an inflorescence from the oldest flower (flower 1) to the primordia buds (flowers +19).

Then, the *FLA5* co-expressed genes were evaluated to provide further insights into *FLA5* putative function (Table 4). A tool developed at professor Matthew Tucker laboratory allows to find genes putatively co-expressed with the gene of interest, based on the similarity of the expression profile of RNA-seq data (Klepikova et al., 2016). The data was analysed considering only the floral tissues. The function or information about the listed genes was briefly investigating using ePlant (Waese et al., 2017). All listed genes have unknown function (Table 4). Genes such as *MICROSPORE-SPECIFIC PROMOTER 2 (MSP2)*, *B-GALACTOSIDASE 7 (BGAL7)*, *ATP-BINDING CASSETTE G17 (ABCG17)* and *WRKY42* have a similar expression profile to *FLA5*.

Table 4 - List of the 24 genes with the most similar expression pattern to *FLA5*, considering only floral tissues. For each gene the locus, gene description and known functions are presented. The list is organized in an ascendant order according to the deltasum. Deltasum measures the similarity of expression pattern, relative to *FLA5* expression pattern. Most similar expression patterns have a deltasum value closer to zero.

Deltasum	Locus	Gene name / Description	Function	Reference
2,71	At5g48550	F-box associated ubiquitination effector family protein	Unknown	-
3,23	At5g43695	-	Unknown	-
3,41	At5g46795	<i>MSP2 (Microspore-specific promoter 2)</i>	Unknown	Honys et al., 2006
3,52	At5g20710	<i>BGAL7 (β-galactosidase 7)</i>	Unknown	Ahn et al., 2007
4,12	At1g23590	Domain of unknown function DUF220	Unknown	-
4,56	At2g46770	<i>Arabidopsis NAC domain containing protein 43 (ANAC043)</i> . NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	Unknown	Mitsuda et al., 2005
4,94	At3g48208	Plant thionin family protein	Unknown	-
5,05	At5g09480	hydroxyproline-rich glycoprotein family protein	Unknown	-
5,13	At5g07571	Oleosin family protein	Unknown	-
5,21	At1g30350	Pectin lyase-like superfamily protein	Unknown	-
5,50	At1g23610	-	Unknown	-
5,52	At5g07570	glycine/proline-rich protein	Unknown	-
5,53	At1g23240	Caleosin-related family protein	Unknown	-
5,54	At2g27670	Domain of unknown function DUF220	Unknown	-
5,64	At1g05290	CCT motif family protein	Unknown	-
5,75	At2g21490	dehydrin LEA	Unknown	-
5,78	At3g55100	<i>ATP-binding cassette G17 (ABCG17)</i> . ABC-2 type transporter family protein	Unknown	Sánchez-Fernández et al., 2001
5,86	At5g17500	Glycosyl hydrolase superfamily protein	Unknown	-
6,00	At4g27570	<i>UGT79B3. UDP-Glycosyltransferase superfamily protein</i>	Unknown	Li et al., 2017
6,08	At5g17470	EF hand calcium-binding protein family	Unknown	-
6,12	At4g04450	<i>WRKY42</i> . WRKY family transcription factor	Unknown	Eulgem et al., 2000
6,13	At3g03540	<i>non-specific phospholipase C5 (NPC5)</i>	Unknown	Nakamura et al., 2005
6,60	At1g23600	Domain of unknown function DUF220	Unknown	-
6,62	At1g23690	Domain of unknown function DUF220	Unknown	-

3.4. Isolation and characterization of the *FLA5* mutant lines

To elucidate the *in vivo* functions of *FLA5* in *Arabidopsis*, a reverse genetic approach was adopted. Two T-DNA insertion lines were identified for *FLA5*: SALK_139382C and SALK_024772C. The homozygous lines were isolated and named as *fla5-1* and *fla5-2*, respectively. Both lines harbored a T-DNA insertion in the *FLA5*-coding sequence (Fig. 15A). Homozygous lines were identified by PCR analysis of gDNA (Fig. 15B-C) and T-DNA insertion sites were confirmed by sequencing (Supplementary fig. 5). A lower amount of *FLA5* transcripts were found in *fla5-1* when comparing to wt (Fig. 15D). *fla5-2*

showed no reduction in *FLA5* expression in homozygous lines and no obvious morphological differences compared to wt plants (data not shown). The analysis for *fla5-2* did not proceed.

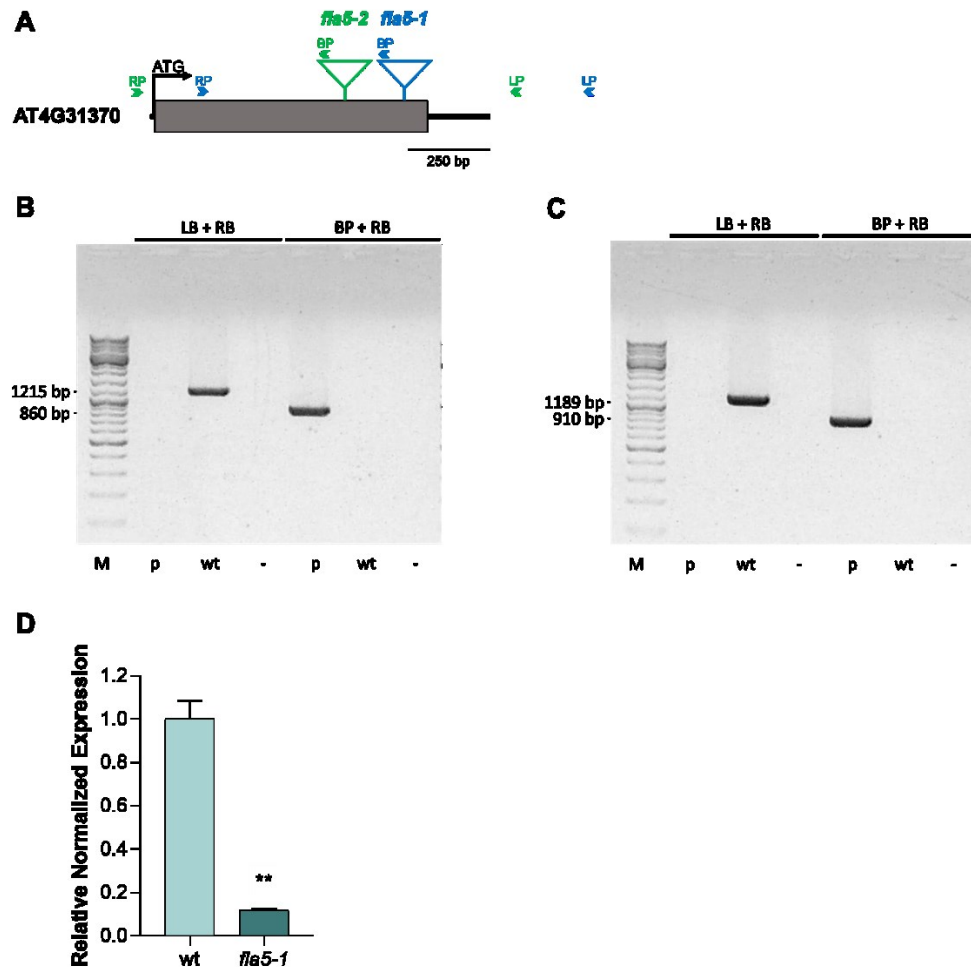


Fig. 15 - Analysis of *FLA5* T-DNA insertion lines.

(A) Schematic gene model structure of *FLA5* (At4g31370) and T-DNA insertion site in *fla5-1* (SALK_139382C) and *fla5-2* (SALK_024772C). The grey box represents the exon and the black line represents the untranslated regions. *fla5-1* and *fla5-2* insertions are indicated by blue and green inverted triangles, respectively. The arrows indicate the genotyping primers (LP - left genomic primer; RP - right genomic primer; BP - T-DNA border primer). (B-C) Genotyping electrophoresis gel of wt, *fla5-1* (B) and *fla5-2* (C) plants, respectively. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (D) qPCR analysis to detect the transcript abundance of *FLA5* in the inflorescences of wt and *fla5-1* plants. The relative gene expression was measured using stably expressed reference genes (*RUB1* and *ACT8*) in three biological samples. The data correspond to the ratio of the expression compared with the wt and are the mean $\pm$ SD of three replicates of a biological sample. Asterisks indicate value significantly different from wt expression according to a Student's t test (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

For *fla5-1* mutant line, the 10th oldest siliques in the stems of each mutant were collected and fixed in ethanol. The number of viable seeds, aborted ovules and seeds inside the

siliques were analyzed, revealing no significant difference from wt siliques grown in the same conditions (Fig. 16A). The silique length of *fla5-1* mutant was also analyzed and showed no significant difference from wt siliques (Fig. 16B).

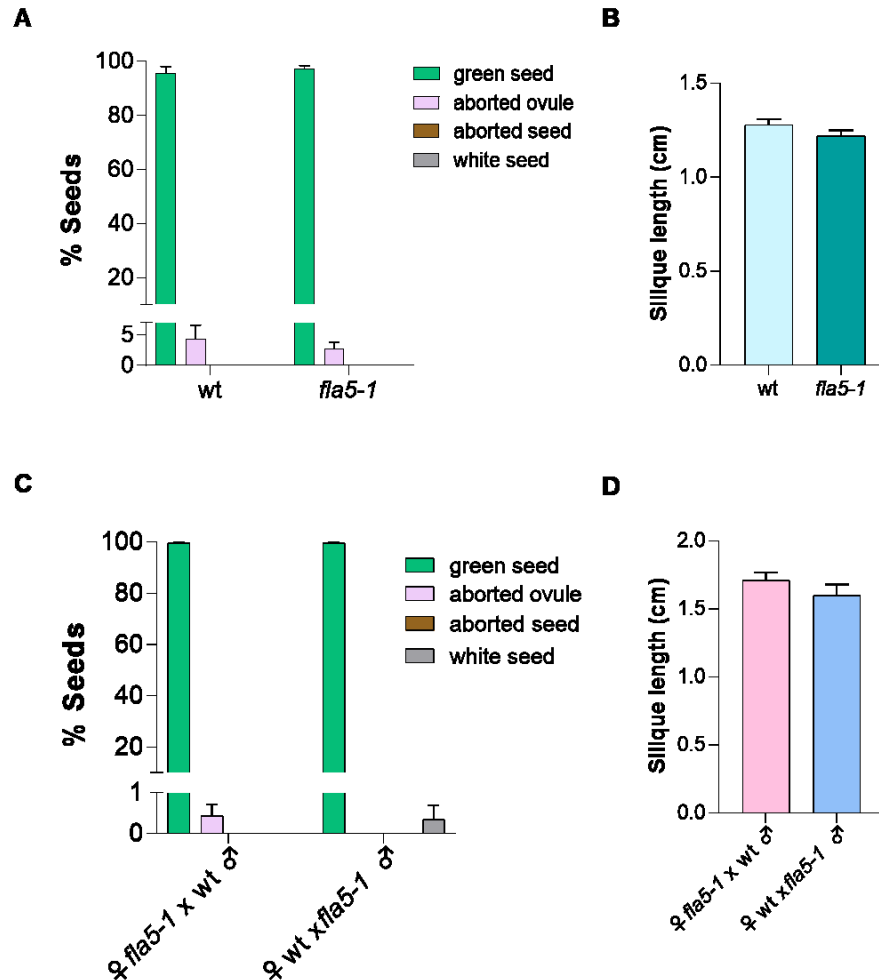


Fig. 16 - Seed set and silique length analysis of Arabidopsis wt and *fla5-1* mutant siliques.

Percentage of seeds per silique (A) and silique length (B) of six-week old Arabidopsis wt and *fla5-1* plants. Percentage of seeds per silique (C) and length (D) of siliques from reciprocal crosses of *fla5-1*. Means of at least three biological replicates are shown. Error bars show SEM. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

To determine whether the male or the female gametophytes were affected in *fla5-1* mutant, reciprocal crosses were conducted between wt and the homozygous mutants. The seed set and silique length showed no significant differences between both crosses (Fig. 16C-D).

Hand-pollination of emasculated wt pistils with *fla5-1* pollen grains, followed by aniline blue staining (Fig. 17A), showed that the *fla5-1* PTs successfully grow through the transmitting tissue and reach the embryo sac of the wt ovules accomplishing double

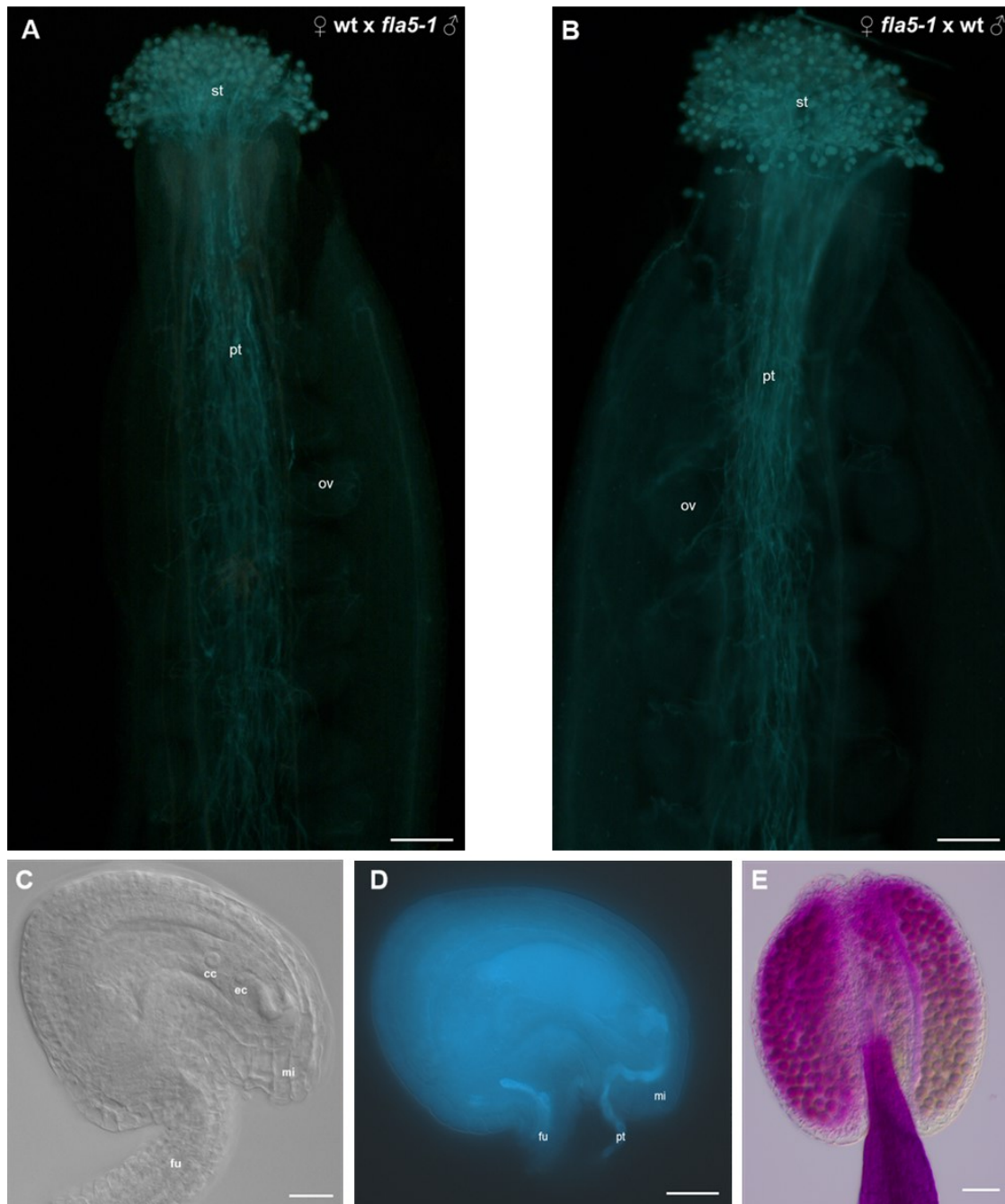


Fig. 17 - *FLA5* mutant phenotypic characterization.

Aniline blue staining of reciprocal crosses between female *wt* and *fla5-1* pollen (A) and female *fla5-1* and *wt* pollen (B). (C) Cleared ovule of *fla5-1* mutant. (D) Aniline blue staining of *fla5-1* ovule. (E) Anther from *fla5-1* mutant stained with Alexander's staining. Pink-red stained pollen grains are viable. Abbreviations: cc, central cell; ec, egg cell; fu, funiculus; mi, micropyle; ov, ovule; pt, pollen tube; st, stigma. Size bar, 100 μ m (A-B), 20 μ m (C-D) and 50 μ m (E).

fertilization, completing their journeys by reaching the bottom of the ovary 24 h after pollination. In the reciprocal crosses, *wt* PTs also grow normally along the transmitting tissue reaching the micropyle and entering the embryo sac of *fla5-1* pistils (Fig. 17B).

fla5-1 flowers were cleared to observe the development of ovules. The synergids, the egg cell and the central cell develop normally (Fig. 17C). In addition, aniline blue staining was performed in *fla5-1* open flowers to verify if the ovules were being fertilized by a PT carrying the sperm cells (Fig. 17D). The PTs have in their walls a substance called callose that forms wall thickenings at irregular distances named callose plugs. Callose, a cell wall polysaccharide composed generally of (1→3)-β-glucans, can be localized by the decolorized aniline blue staining when illuminated by blue or ultraviolet light (Kho and Baër, 1968; Dumas and Knox, 1983).

To determine if *fla5-1* pollen was viable, *fla5-1* anthers were stained with Alexander's staining (Fig. 17E). Alexander's staining allows to determine if a pollen grain is viable or non-viable based on the color of the pollen stained. The malachite green dye stains the pollen walls while acid fuchsin stains the cytoplasm. As the aborted pollen only has walls, it will stain in green-blue whereas the normal pollen will appear with a red-magenta color (Alexander, 1969). Pollen from *fla5-1* plants stained red-magenta and so, it was viable. Therefore, no phenotype was found for *fla5-1* during reproduction.

Then, putative interactors of *FLA5* were tested by qPCR to determine whether their expression was altered in *fla5-1* inflorescences comparatively to wt inflorescences (Fig. 18). In the *fla5-1* mutant, the relative expression of *FLA3* and *STK* was upregulated comparing to the expression in wt. However, *FLA14* was downregulated in *fla5-1* mutant.

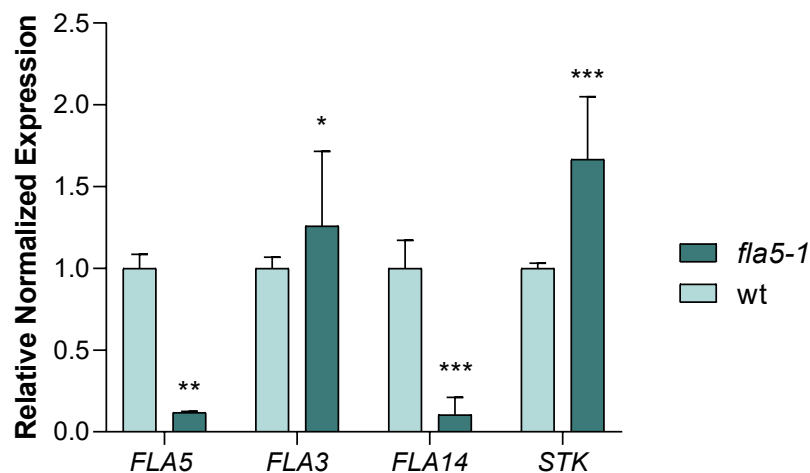


Fig. 18 - Relative expression of *FLA5*, *FLA3*, *FLA14* and *STK* in wt and *fla5-1* inflorescences.

The relative gene expression was measured using stably expressed reference genes (*RUB1* and *ACT8*) in three biological samples. The data correspond to the ratio of the expression compared with the wt and are the mean ± SD of three replicates of a biological sample. Asterisks indicate value significantly different from wt expression according to a Student's t test (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

4. DISCUSSION

Angiosperms, the flowering plants, are the most successful group of plants due to their system of sexual reproduction, known as double fertilization. During this process, two male sperm cells are delivered by a PT to the female gametophyte where one fuses with the egg cell and the other fuses with the central cell, giving rise to the embryo and the endosperm, respectively.

AGPs are expressed along the reproductive tissues, playing different roles and contributing to a successful fertilization process and consequent seed formation.

4.1. Phylogenetic analysis of FLAs family

FLAs are a subclass of AGPs that contain FAS1 domains in addition to AGP regions. The 21 FLAs protein sequences in Arabidopsis clustered into four different groups (A, B, C and D) following the classification proposed by Johnson et al. (2003). Analyzing the phylogenetic relationships among FLAs family, three closely related paralogs are evident: FLA3, FLA5 and FLA14. FLA3 and FLA14 are involved in microspore development and are highly expressed in pollen grains and in PTs (Li et al., 2010; Miao et al., 2018). FLA3 and FLA14 are expressed in the same tissues and share functions but only share 36% sequence identity. This suggest that even proteins that share low sequence identity might play the same biological role. FLA5 shares 31% of amino acid identity with FLA14 and 56% with FLA3. Therefore, these results might implicate a similar role for FLA5.

In this study, five, nine, 22 and 53 putative FLA proteins were identified in *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa*, respectively. Recently, the evolutionary history of this gene family in plants was discovered. FLA genes from 13 plants were identified (He et al., 2019). Despite the evolutionary divergences, in this analysis, FLAs from *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa* were also classified into four clades (A-D) as described previously (Costa et al., 2019a; He et al., 2019).

However, the number of putative FLAs found in this study was different from previous studies. In *P. patens*, from the 12 FLAs identified in He et al. (2019), only PpFLA3, PpFLA5, PpFLA10 and PpFLA11 were identified. A new putative FLA gene, Pp3c26_5590 was also identified. In the same way, in *S. moellendorffii*, two FLA genes, SmFLA1 and SmFLA2 were not identified while SMO114G0121 was identified as a new putative FLA gene. Similarly, in *P. trichocarpa*, PtFLA24, PtFLA40, PtFLA44, PtFLA50, PtFLA51 and PtFLA52 were missing. However, Potri.001G306800, Potri.019G125000,

Potri.013G151366 and Potri.013G151432 were new putative FLA genes identified in this study. No previous study regarding the identification of FLAs in *H. vulgare* was found. Therefore, 22 new putative FLA genes were identified in this study. No structural characteristics were analyzed from the identified putative FLAs. This type of analysis was only carried out for AtFLA5, the key protein in this study.

FLAs are present across several plant species since mosses to flowering plants. Some functions have been implicated for FLAs over the years in plant development (Zang et al., 2015b). These results show the importance of the presence of these glycoproteins for plant survival.

4.2. Bioinformatic analysis of FLA5

After using several programs to predict FLA5 domains and specific sites, FLA5 deduced back bone is probably composed by: a N-terminal signal peptide (positions 1-24), a FAS1 domain (positions 24-141), a Pro-rich domain (positions 180-237) and GPI-anchor signal on the C-terminal (positions 256-278). This is in accordance with what is expected for the structure of a FLA belonging to the group C. Group C comprises FLAs with either one or two Pro-rich domains, one or two FAS1 domains and a C-terminal GPI anchor (Johnson et al., 2003). As many AGPs, FLA5 also undergoes PTMs. During maturation, FLA5 loses the signal peptide (position 24) and the GPI signal (Ser255) which is replaced by a GPI-anchor. FLA5 was also predicted to have four N-glycosylation sites (positions 26, 74, 126 and 159). Therefore, the mature protein is constituted by a FAS1 domain N-glycosylated, a Pro-rich domain with O-linked polysaccharide chains and a GPI-anchor.

4.3. Expression pattern of *FLA5*

qPCR results showed that *FLA5* was highly expressed in the siliques, which was in accordance to the RNA-seq data (Klepikova et al., 2016). Interestingly, the RNA-seq data pointed out high levels of *FLA5* transcripts in flower 1. qPCR analysis only found low levels of expression in inflorescence probably due to divergences between the tissues used in both analyses. While for qPCR all flower stages were used, for RNA-seq only the first flower at stage 15 according to Smyth et al. (1990) was used.

FLA5 promoter region fused with the GUS reporter gene was used to determine the spatiotemporal expression pattern in the reproductive tissues. GUS expression driven by the *FLA5* promoter in the flower localized in the vascular tissues of sepals and petals. Between stages 10 to 15, GUS was found to be expressed in pollen grains in the anthers. During stages 14 to 15, the signal was found in the germinating pollen in the stigma. *In*

vitro pollen germination was performed to allow detailed gene expression analyses during PT growth. GUS expression was detected in pollen grains and PTs.

Relative to the male gametophyte development, *FLA5* seemed to have a similar expression pattern as *FLA3* and *FLA14* being all expressed in pollen grains in the anthers from stage 12 to stage 15 and in PTs (Li et al., 2010; Miao et al., 2018). This small phylogenetic cluster seems to have an important role during pollen development in Arabidopsis. Other phylogenetic group constituted by four AGPs (AGP6, AGP11, AGP23 and AGP40) were also specifically expressed in pollen grains and PTs and the four AGPs were involved in microspore development, pollen germination and PT growth (Nguema-Ona et al., 2012).

4.4. Isolation and characterization of the *FLA5* mutant lines

AGPs are proposed to be a big family of glycoproteins where some gene redundancy is expected. Therefore, the study of single T-DNA insertion mutant is most of the time ineffective. To overcome this problem, one solution is to obtain high order mutants of related AGPs (double, triple and even quadruple). However, this resolution is extremely time-consuming. Moreover, some FLAs phenotypes and respective function have been discovered by growing the mutants in different abiotic conditions.

A reverse genetic approach was undertaken to characterize the biological significance of *FLA5*. Only one T-DNA insertion mutant was analyzed and no macroscopic phenotypic alterations were observed comparing to wt. This finding corroborates the hypothesis that AGPs have redundant functions.

Phenotypic analysis of *fla5-1* revealed no impact of the mutation on female or male gametophyte development under normal growth conditions. All cell types developed correctly and fertilization occurred normally giving rise to seeds with normal embryo development. Similarly, *fla14* mutants showed no plant phenotype under normal growth conditions. However, under high humidity conditions, *fla14* pollen germinated precociously inside the anthers (Miao et al., 2018). Besides, a T-DNA line for *FLA3* showed no reduction in *FLA3* expression and consequently, no phenotype. *FLA3* biological function was revealed generating overexpressing and RNAi transgenic plants (Li et al., 2010). The same singularity happened for *agp6*, *agp11* and *agp40* single mutants. Nevertheless, the triple mutant (*agp6agp11agp40*) had several early germinating PTs inside the anthers (Nguema-Ona et al., 2012). As *FLA3*, *FLA5* and *FLA14* were closely related to each other and expressed in the pollen, a possible redundant function is expected. As single mutants revealed no visible phenotype, double

(*fla3fla5* and *fla5fla14*) or triple mutants (*fla3fla5fla14*) are needed to access the precise function of *FLA5*.

Since *FLA5* shared 56% and 31% of amino acid sequence identity with *FLA3* and *FLA14*, respectively, the levels of transcripts of *FLA3* and *FLA14* were measured in *fla5-1* plants. Besides, as *FLA5* is a putative target of *STK*, the *STK* transcript levels were also measured. Interestingly, *FLA14* was downregulated whereas *FLA3* and *STK* were upregulated in *fla5-1* mutant. It is hypothesized that *STK* positively regulates *FLA5* and *FLA5* positively regulates *FLA14*. In addition, *FLA5* transcript levels showed no decrease in *fla3* RNAi plants (Li et al., 2010). Therefore, it seems that *FLA5* is negatively regulating *FLA3*. This result is a small step to understand the network of genes regulating pollen and PT development.

5. CONCLUSIONS

Since double fertilization drives formation and development of seeds, the most important source of food in the planet, significant advances have been achieved regarding the understanding of the mechanisms involved in sexual reproduction. FLA5 is a putative target of STK, a MADS-box transcription factor, deeply involved in the formation of seeds. Thereby, the objective of this chapter was to study this protein and verify its involvement during sexual reproduction in *A. thaliana*.

The bioinformatic analysis suggested an involvement of FLA5 in pollen development. FLA5 had two closely related paralogs (FLA3 and FLA14), both involved in microspore development (Li et al., 2010; Miao et al., 2018). By qPCR it was demonstrated that *FLA5* was highly expressed in the siliques. *FLA5prom:GUS* localized in the vascular tissues and in the pollen from a young stage until the formation of a PT. Besides, *FLA3* and *STK* were upregulated while *FLA14* was downregulated in *fla5-1* mutant. The data collected supports the hypothesis that FLA5 might be involved in pollen development.

There is still much to be done in order to make more progresses in the study of this specific AGP. The next steps will include:

- growth of *fla5* mutants under high moisture conditions;
- determination of subcellular localization of FLA5;
- generation and characterization *FLA5* over-expressing plants;
- production and examination *FLA5* CRISPR/Cas9 lines;
- electron microscopy observation throughout pollen development.

It will be also essential to carry on with further analyses of double or even triple mutants with genes that might be acting redundantly with *FLA5* such as *FLA3* and *FLA14*.

These results greatly contribute to expand the knowledge regarding FLAs present in the reproductive tissues and their transcriptional regulation by the STK. It can be predicted that FLA5 might act on pollen development and PT growth.

6. REFERENCES

- Ahn YO, Zheng M, Bevan DR, Esen A, Shiu SH, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, et al** (2007) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. *Phytochemistry* **68**: 1510–1520 doi: 10.1016/j.phytochem.2007.03.021
- Alexander MP** (1969) Differential staining of aborted and nonaborted pollen. *Biotech Histochem* **44**: 117–122 doi: 10.3109/10520296909063335
- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H** (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**: 420–423 doi: 10.1038/s41587-019-0036-z
- Basu D, Tian L, DeBrosse T, Poirier E, Emch K, Herock H, Travers A, Showalter AM** (2016) Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) Pathway in *Arabidopsis*. *PLoS One* **11**: e0145092 doi: 10.1371/journal.pone.0145092
- Cagnola JI, Dumont de Chassart GJ, Ibarra SE, Chimenti C, Ricardi MM, Delzer B, Ghiglione H, Zhu T, Otegui ME, Estevez JM, et al** (2018) Reduced expression of selected FASCICLIN-LIKE ARABINO GALACTAN PROTEIN genes associates with the abortion of kernels in field crops of *Zea mays* (maize) and of *Arabidopsis* seeds. *Plant Cell Environ* **41**: 661–674 doi: 10.1111/pce.13136
- Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S** (2019a) In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reprod* **32**: 353–370 doi: 10.1007/s00497-019-00376-7
- Costa M, Silva J, Coimbra S** (2019b) *Arabidopsis thaliana* pollen tube culture for multi-omics studies. In A Geitmann, ed, *The Pollen Tube: Methods and Protocols*. Springer. Manuscript submitted for publication.
- Dahiya P, Findlay K, Roberts K, McCann MC** (2006) A fasciclin-domain containing gene, ZeFLA11, is expressed exclusively in xylem elements that have reticulate wall thickenings in the stem vascular system of *Zinnia elegans* cv Envy. *Planta* **223**: 1281–1291 doi: 10.1007/s00425-005-0177-9
- Dumas C, Knox RB** (1983) Callose and determination of pistil viability and incompatibility. *Theor Appl Genet* **67**: 1–10 doi: 10.1007/BF00303914

- Edwards K, Johnstone C, Thompson C** (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* **19**: 1349 doi: 10.1093/nar/19.6.1349
- Eisenhaber B, Wildpaner M, Schultz CJ, Borner GHH, Dupree P, Eisenhaber F** (2003) Glycosylphosphatidylinositol Lipid Anchoring of Plant Proteins. Sensitive Prediction from Sequence- and Genome-Wide Studies for Arabidopsis and Rice. *Plant Physiol* **133**: 1691–1701 doi: 10.1104/pp.103.023580
- Emanuelsson O, Nielsen H, Brunak S, Von Heijne G** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005–1016 doi: 10.1006/jmbi.2000.3903
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE** (2000) The WRKY superfamily of plant transcription factors. *Trends Plant Sci* **5**: 199–206 doi: 10.1016/S1360-1385(00)01600-9
- Faik A, Abouzouhair J, Sarhan F** (2006) Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): Identification and bioinformatic analyses. *Mol Genet Genomics* **276**: 478–494 doi: 10.1007/s00438-006-0159-z
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A** (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**: 3784–3788 doi: 10.1093/nar/gkg563
- Griffiths JS, Tsai AY-L, Xue H, Voiniciuc C, Sola K, Seifert GJ, Mansfield SD, Haughn GW** (2014) SALT-OVERLY SENSITIVE5 Mediates Arabidopsis Seed Coat Mucilage Adherence and Organization through Pectins. *Plant Physiol* **165**: 991–1004 doi: 10.1104/pp.114.239400
- Guerriero G, Mangeot-Peter L, Legay S, Behr M, Lutts S, Siddiqui KS, Hausman JF** (2017) Identification of fasciclin-like arabinogalactan proteins in textile hemp (*Cannabis sativa* L.): In silico analyses and gene expression patterns in different tissues. *BMC Genomics* **18**: doi: 10.1186/s12864-017-3970-5
- Harpaz-Saad S, McFarlane HE, Xu S, Divi UK, Forward B, Western TL, Kieber JJ** (2011) Cellulose synthesis via the FEI2 RLK/SOS5 pathway and cellulose synthase 5 is required for the structure of seed coat mucilage in Arabidopsis. *Plant J* **68**: 941–53 doi: 10.1111/j.1365-313X.2011.04760.x
- He J, Zhao H, Cheng Z, Ke Y, Liu J, Ma H** (2019) Evolution analysis of the fasciclin-like arabinogalactan proteins in plants shows variable fasciclin-AGP domain constitutions. *Int J Mol Sci* **20**: 1–17 doi: 10.3390/ijms20081945

- Honys D, Oh S-A, Renák D, Donders M, Solcová B, Johnson JA, Boudová R, Twell D** (2006) Identification of microspore-active promoters that allow targeted manipulation of gene expression at early stages of microgametogenesis in *Arabidopsis*. *BMC Plant Biol* **6**: 31 doi: 10.1186/1471-2229-6-31
- Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH** (2017) SUBA4: The interactive data analysis centre for *Arabidopsis* subcellular protein locations. *Nucleic Acids Res* **45**: D1064–D1074 doi: 10.1093/nar/gkw1041
- Huang GQ, Xu WL, Gong SY, Li B, Wang XL, Xu D, Li XB** (2008) Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. *Physiol Plant* **134**: 348–359 doi: 10.1111/j.1399-3054.2008.01139.x
- Invitrogen** (2003) Gateway® Technology - A universal technology to clone DNA sequences for functional analysis and expression in multiple systems. Retrieved from <http://tools.thermofisher.com/content/sfs/manuals/gatewayman.pdf>
- Ito S, Suzuki Y, Miyamoto K, Ueda J, Yamaguchi I** (2005) AtFLA11, a fasciclin-like arabinogalactan-protein, specifically localized in sclerenchyma cells. *Biosci Biotechnol Biochem* **69**: 1963–1969 doi: 10.1271/bbb.69.1963
- Johnson KL, Jones BJ, Bacic A, Schultz CJ** (2003) The fasciclin-Like arabinogalactan proteins of *Arabidopsis*. A multigene family of putative cell adhesion molecules. *Plant Physiol* **133**: 1911–1925 doi: 10.1104/pp.103.031237
- Johnson KL, Kibble NAJ, Bacic A, Schultz CJ** (2011) A fasciclin-like Arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, *fla1*, shows defects in shoot regeneration. *PLoS One* **6**: 1–11 doi: 10.1371/journal.pone.0025154
- Jun L, Xiaoming W** (2012) Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Mol Biol Rep* **39**: 10541–10555 doi: 10.1007/s11033-012-1940-1
- Karimi M, Inzé D, Depicker A** (2002) GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* **7**: 193–195 doi: 10.1016/S1360-1385(02)02251-3
- Kho YO, Baër J** (1968) Observing pollen tubes by means of fluorescence. *Euphytica* **17**: 298–302 doi: 10.1007/BF00021224
- Klepikova A V., Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA** (2016) A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J* **88**: 1058–1070 doi: 10.1111/tpj.13312

- Kumar S, Stecher G, Li M, Knyaz C, Tamura K** (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549 doi: 10.1093/molbev/msy096
- Lafarguette F, Leplé JC, Déjardin A, Laurans F, Costa G, Lesage-Descauses MC, Pilate G** (2004) Poplar genes encoding fasciclin-like arabinogalactan proteins are highly expressed in tension wood. *New Phytol* **164**: 107–121 doi: 10.1111/j.1469-8137.2004.01175.x
- Li J, Yu M, Geng LL, Zhao J** (2010) The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of Arabidopsis. *Plant J* **64**: 482–497 doi: 10.1111/j.1365-313X.2010.04344.x
- Li P, Li Y-J, Zhang F-J, Zhang G-Z, Jiang X-Y, Yu H-M, Hou B-K** (2017) The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. *Plant J* **89**: 85–103 doi: 10.1111/tpj.13324
- Liu H, Shi R, Wang X, Pan Y, Li Z, Yang X, Zhang G, Ma Z** (2013) Characterization and Expression Analysis of a Fiber Differentially Expressed Fasciclin-like Arabinogalactan Protein Gene in Sea Island Cotton Fibers. *PLoS One* **8**: doi: 10.1371/journal.pone.0070185
- Ma H, Zhao J** (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). *J Exp Bot* **61**: 2647–2668 doi: 10.1093/jxb/erq104
- MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG** (2010) Fasciclin-like arabinogalactan proteins: Specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. *Plant J* **62**: 689–703 doi: 10.1111/j.1365-313X.2010.04181.x
- MacMillan CP, Taylor L, Bi Y, Southerton SG, Evans R, Spokevicius A** (2015) The fasciclin-like arabinogalactan protein family of *Eucalyptus grandis* contains members that impact wood biology and biomechanics. *New Phytol* **206**: 1314–27 doi: 10.1111/nph.13320
- Miao Y, Lin S, Cao J** (2018) The fasciclin-like arabinogalactan protein gene, FLA14, is involved in microspore development of Arabidopsis. Poster presented at 25th International Congress on Sexual Plant Reproduction; 2018 Jun 11–16; Gifu, Japan
- Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M** (2005) The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* **17**: 2993–3006 doi:

10.1105/tpc.105.036004

- Mizzotti C, Ezquer I, Paolo D, Rueda-Romero P, Guerra RF, Battaglia R, Rogachev I, Aharoni A, Kater MM, Caporali E, et al** (2014) SEEDSTICK is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat. *PLoS Genet* **10**: e1004856 doi: 10.1371/journal.pgen.1004856
- Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L** (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *Plant J* **70**: 409–420 doi: 10.1111/j.1365-313X.2011.04878.x
- Nakamura Y, Awai K, Masuda T, Yoshioka Y, Takamiya K, Ohta H** (2005) A novel phosphatidylcholine-hydrolyzing phospholipase C induced by phosphate starvation in Arabidopsis. *J Biol Chem* **280**: 7469–76 doi: 10.1074/jbc.M408799200
- Nguema-Ona E, Coimbra S, Vicré-Gibouin M, Mollet JC, Driouich A** (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. *Ann Bot* **110**: 383–404 doi: 10.1093/aob/mcs143
- Pereira AM, Nobre MS, Pinto SC, Lopes AL, Costa ML, Masiero S, Coimbra S** (2016) “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 doi: 10.1016/j.molp.2016.01.002
- 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 doi: 10.1038/nature01741
- Sánchez-Fernández R, Davies TG, Coleman JO, Rea PA** (2001) The Arabidopsis thaliana ABC protein superfamily, a complete inventory. *J Biol Chem* **276**: 30231–44 doi: 10.1074/jbc.M103104200
- Schrödinger L** (2015) The PyMol Molecular Graphics System, Version 2.0.
- Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A** (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* **129**: 1448–1463 doi: 10.1104/pp.003459
- Schwacke R, Schneider A, Van Der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R** (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiol* **131**: 16–26 doi: 10.1104/pp.011577
- Seifert GJ, Xue H, Acet T** (2014) The Arabidopsis thaliana FASCICLIN LIKE ARABINOGLACTAN PROTEIN 4 gene acts synergistically with abscisic acid

- signalling to control root growth. *Ann Bot* **114**: 1125–33 doi: 10.1093/aob/mcu010
- Shi H, Kim YS, Guo Y, Stevenson B, Zhu JK** (2003) The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. *Plant Cell* **15**: 19–32 doi: 10.1105/tpc.007872
- Showalter AM, Keppler BD, Liu X, Lichtenberg J, Welch LR** (2016) Bioinformatic Identification and Analysis of Hydroxyproline-Rich Glycoproteins in *Populus trichocarpa*. *BMC Plant Biol* **16**: 1–34 doi: 10.1186/s12870-016-0912-3
- Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N** (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**: D161–6 doi: 10.1093/nar/gkp885
- Smyth DR, Bowman JL, Meyerowitz EM** (1990) Early flower development in *Arabidopsis*. *Plant Cell* **2**: 755–767 doi: 10.1105/tpc.2.8.755
- Sonnhammer EL, von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175–182
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680 doi: 10.1093/nar/22.22.4673
- Turupcu A, Almohamed W, Oostenbrink C, Seifert GJ** (2018) A speculation on the tandem fasciclin 1 repeat of FLA4 proteins in angiosperms. *Plant Signal Behav* **13**: e1507403 doi: 10.1080/15592324.2018.1507403
- Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van De Peer Y, Coppens F, Vandepoele K** (2018) PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res* **46**: D1190–D1196 doi: 10.1093/nar/gkx1002
- Waese J, Fan J, Pasha A, Yu H, Fucile G, Shi R, Cumming M, Kelley LA, Sternberg MJ, Krishnakumar V, et al** (2017) ePlant: Visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* **29**: 1806–1821 doi: 10.1105/tpc.17.00073
- Webb B, Sali A** (2016) Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinforma* **15**: 1–47 doi: 10.1002/cpbi.3
- Xue H, Veit C, Abas L, Tryfona T, Maresch D, Ricardi MM, Estevez JM, Strasser R, Seifert GJ** (2017) *Arabidopsis thaliana* FLA4 functions as a glycan-stabilized

soluble factor via its carboxy-proximal Fasciclin 1 domain. *Plant J* **91**: 613–630 doi: 10.1111/tpj.13591

Yang SH, Wang H, Sathyan P, Stasolla C, Loopstra CA (2005) Real-time RT-PCR analysis of loblolly pine (*Pinus taeda*) arabinogalactan-protein and arabinogalactan-protein-like genes. *Physiol Plant* **124**: 91–106 doi: 10.1111/j.1399-3054.2005.00479.x

Zang L, Zheng T, Chu Y, Ding C, Zhang W, Huang Q, Su X (2015a) Genome-wide analysis of the fasciclin-like arabinogalactan protein gene family reveals differential expression patterns, localization, and salt stress response in populus. *Front Plant Sci* **6**: 1–16 doi: 10.3389/fpls.2015.01140

Zang L, Zheng T, Su X (2015b) Advances in research of fasciclin-like arabinogalactan proteins (FLAs) in plants. *Plant Omics* **8**: 190–194

Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**: 2237–2243 doi: 10.1016/j.jmb.2017.12.007

CHAPTER 3

IDENTIFICATION OF THE ROLE OF METHYL GLUCURONIC ACID OF ARABINOGALACTAN PROTEINS DURING ARABIDOPSIS SEXUAL REPRODUCTION

ABSTRACT

AGPs constitute a highly diverse class of heavily glycosylated proteins that consist of a core-protein backbone surrounded by carbohydrates. The carbohydrate moiety of AGPs represents more than 90% of their total molecular mass and is composed of Ara and Gal and minor sugars such as glucuronic acid (GlcA), fucose and rhamnose. AGP glycosylation is catalyzed by a large number of GTs in the secretory pathway. The biosynthesis of AGP glycans requires at least ten functionally distinct transferases. In recent years, 22 transferases responsible for AGP glycosylation have been identified in Arabidopsis. However, several enzymes remain to be identified. Identifying the genes involved in AGPs biosynthesis is crucial to understand their function and structure.

Several studies revealed that AGPs are recruited into sexual reproduction processes. Recently, AGPs were proposed to bind and store Ca^{2+} through GlcA residues at the plasma membrane. Since Ca^{2+} oscillations are present in many reproductive processes, some AGPs expressed in the reproductive tissues may be surface Ca^{2+} capacitors.

The Carbohydrate Active Enzymes (CAZy) GT14 family in Arabidopsis contains β -GlcATs such as GlcAT14A, GlcAT14B, GlcAT14C, GlcAT14D and GlcAT14E (Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez, 2018) that add GlcA to AG polysaccharides of AGPs. The mutants *glcat14a*, *glcat14ab*, *glcat14abe* and *glcat14abd* showed a reduction in the content of [Me]GlcA (Knoch et al., 2013; Lopez-Hernandez, 2018) and AGP extracts from *glcat14abd-1* held less Ca^{2+} than wt plants *in vitro*. Lopez-Hernandez (2018) reported some Ca^{2+} -dependent growth phenotypes in the mutants of these enzymes. However, the involvement of these enzymes in sexual reproduction was not deeply analyze.

In this thesis, the expression pattern of GT14 genes was evaluated by qPCR analysis that revealed overlapping but distinct expression patterns. *GlcAT14A*, *GlcAT14B*, *GlcAT14D* and *GlcAT14E* were the genes most expressed in the flowers. Promotor analysis using multiple YFP fused to a NLS was accomplished to confirm the GlcATs expression pattern. *GlcAT14B* was expressed in the stamens, in the transmitting tissue and in the chalazal and funiculus tissues of the ovule. *GlcAT14E* was present along the transmitting tract and the anthers. *GlcAT14E* was also expressed in leaves, sepals, petals, funiculus and chalaza of ovules. *At4g27480* was present in stamens and petals. In addition, GlcAT14B, GlcAT14E, At4g27480, At1g71070, At1g53100 and At1g03520 were localized to Golgi vesicles in *N. benthamiana* leaves.

To explore the biological role of [Me]GlcA of AGPs in plant growth, a functional characterization of the mutants with reduced content of [Me]GlcA from AGPs was performed. As one of the roles of GlcA is its capacity of binding to Ca^{2+} , the mutants were grown in two different concentrations of Ca^{2+} : 0.2 mM and 2 mM. The triple mutants *glcat14abd-1* and *glcat14abd-2* grown in low Ca^{2+} were dwarfed plants with smaller siliques containing a significant number of aborted ovules and seeds relative to wt. Although these mutants presented a normal developed embryo sac, some pollen grains were non-viable and this was reflected in a huge percentage of non-fertilized ovules. The phenotypic studies led us to hypothesize the involvement of GT14 enzymes in pollen development and in pollen tube attraction to the ovule.

This project allows a deeper understanding of the functional importance of [Me]GlcA of AGPs during sexual reproduction.

1. INTRODUCTION

1.1. Glycosyltransferases - meet the sugar workers

Glycoproteins have a fascinating complex structure and therefore plants require the action of numerous GTs families to assist their biosynthesis (Zhou et al., 2009). AGP glycosylation is catalyzed by a large number of GTs (EC 2.4.x.y) in the secretory pathway (Geshi, 2014). GTs are ubiquitous enzymes that catalyze glycosidic bond formation between a sugar moiety and a specific acceptor molecule (sugars, proteins, lipids or small molecules), creating a diverse collection of oligosaccharides and glycoconjugates in nature (Hansen et al., 2012; Gloster, 2014). GTs have been classified into 108 families in the Carbohydrate Active Enzymes (CAZy) database (<http://www.cazy.org>; Coutinho et al., 2003; Lombard et al., 2014). GTs are very specific and thus each different linkage may require a distinct GT (Henrissat and Davies, 2000). Almost five hundred putative GTs sequences have been identified in the Arabidopsis genome and classified into 42 different CAZy families (Coutinho et al., 2003). GTs can be type I membrane proteins located in the ER, type II membrane proteins located in the Golgi or integral membrane proteins such as cellulose synthases (Hansen et al., 2009). GTs with a type II membrane protein topology have a short N-terminal fragment most likely protruding into the cytosol, one helical transmembrane domain and a C-terminal catalytic domain containing a DXD motif localized in the Golgi lumen (Hansen et al., 2009; Chou et al., 2012).

However, little is known about the enzymes involved in AGP glycosylation in plants. The biosynthesis of AGP glycans requires at least ten functionally distinct enzymes (Knoch et al., 2014), e.g. galactosyltransferases (GALTs), arabinosyltransferases, fucosyltransferases (FUTs), rhamnosyltransferases, GlcATs and glucuronic acid methyltransferases. To date, 22 transferases responsible for AGP glycosylation have been identified in Arabidopsis (Fig. 1; Table 1): eight hydroxyproline-O- β -GALTs (Hyp-O-GALTs) (GALT2, GALT3, GALT4, GALT5 and GALT6, Basu et al., 2013, 2015a, 2015b, 2016; HPTG1, HPTG2 and HPTG3, Ogawa-Ohnishi and Matsubayashi, 2015), two β -1,3-GALTs (At1g77810, Qu et al., 2008; KSN4, Suzuki et al., 2017), two β -1,6-GALTs (GALT31A, Geshi et al., 2013; GALT29A; Dilokpimol et al., 2014), five β -1,6-GlcATs (GlcAT14A, GlcAT14B, GlcAT14C, GlcAT14D and GlcAT14E, Knoch et al., 2013; Dilokpimol and Geshi, 2014; Hernández, 2018), two α -1,2-FUTs (FUT4 and FUT6, Wu et al., 2010; Liang et al., 2013; Tryfona et al., 2014), a putative AGP β -arabinosyltransferase (RAY1, Gille et al., 2013) and two glucuronic acid methyltransferases (AGM1 and AGM2, Temple et al., 2019). However, several enzymes

Table 1 - Information about the characterized transferases involved in the biosynthesis of AGPs glycans in *A. thaliana*.

CAZy GT family	Enzyme	Locus	Enzyme Activity	Subcellular localization	Mutants	T-DNA insertion	Mutants Phenotypes	References
GT14	GlcAT14A	At5g39990	β -1,6-glucuronosyltransferase	Golgi	<i>glcat14a-1</i> (SALK_064313)	Exon	enhanced cell elongation in seedlings; reduction of GlcA substitution on Gal- β -1,6-Gal and β -1,3-Gal, increase of Gal and reduction of Ara in their AGPs	Knoch et al., 2013; Lopez-Hernandez, 2018
					<i>glcat14a-2</i> (SALK_043905)	Exon		
	GlcAT14B	At5g15050		-	<i>glcat14b-1</i> (SALK_080923)	Exon	reduction in [Me]GlcA content of AGPs (<i>glcat14ab</i> , <i>glca14abd-1</i> , <i>glcat14abd-2</i> , <i>glcat14abe</i>); reduction in Ca ²⁺ binding capacity of AGPs (<i>glcat14abd-1</i>); shorter inflorescence stems (<i>glcat14abd-1</i>); shorter siliques (<i>glcat14ab</i> , <i>glcat14abd-1</i>); reduction in trichome branching (<i>glcat14ab</i> , <i>glcat14abd-1</i> , <i>glcat14abd-2</i>); third leaf with ellipse shape (<i>glcat14ab</i> , <i>glcat14abd-1</i>); root skewness (<i>glcat14ab</i> , <i>glcat14abe</i> , <i>glcat14abd-1</i>) and hair branching (<i>glcat14abe</i>); smaller etiolated hypocotyls (<i>glcat14abe</i>)	Dilokpimol and Geshi, 2014; Lopez-Hernandez, 2018
	GlcAT14D	At3g24040		-	<i>glcat14d-1</i> (GK363F05) <i>glcat14d-2</i> (GK508D01)	Intron Intron		
	GlcAT14E	At3g15350		-	<i>glcat14e-1</i> (SALK_022820)	Intron		Lopez-Hernandez, 2018
	GlcAT14C	At2g37585		-	-	-	-	
GT29	GALT29A	At1g08280	β -1,6-galactosyltransferase	Golgi	-	-	-	Dilokpimol et al., 2014

Table 1 - (cont.).

CAZy GT family	Enzyme	Locus	Enzyme Activity	Subcellular localization	Mutants	T-DNA insertion	Mutants Phenotypes	References
GT31	GALT2	At4g21060	hydroxyproline-O- β -galactosyltransferase	ER and Golgi	<i>galt2-1</i> (SALK_117233)	Exon	lower GALT activities and reduced β -Yariv-precipitable AGPs; root and pollen tubes sensitivity reduction to β -Yariv reagent; reduced growth and root tip swelling in response to salt and sucrose; reduced root hair length and density (<i>galt2</i> , <i>galt3</i> , <i>galt5</i> , <i>galt2galt5</i>), seed production (<i>galt4</i> , <i>galt6</i>) and seed coat mucilage (<i>galt3</i> , <i>galt6</i> , <i>galt2galt5</i>); premature senescence (<i>galt6</i>); large number of rosette leaves, delayed flowering and shorter siliques (<i>galt2galt5</i>)	Basu et al., 2013, 2015a, 2015b
					<i>galt2-2</i> (SALK_141126)	Exon		
	GALT3	At3g06440		Golgi	<i>galt3-1</i> (SALK_085633)	Promoter		
					<i>galt3-2</i> (SALK_005178)	Promoter		
	GALT4	At1g27120		Golgi	<i>galt4-1</i> (SALK_136251)	Exon		
					<i>galt4-2</i> (SALK_131723)	Exon		
	GALT5	At1g74800		Golgi	<i>galt5-1</i> (SALK_105404)	5' UTR		
					<i>galt5-2</i> (SALK_115741)	Exon		
	GALT6	At5g62620		Golgi	<i>galt6-1</i> (SAIL_59_D08)	Exon		
					<i>galt6-2</i> (SAIL_70_B02)	Exon		
	HPGT1	At5g53340		Golgi	<i>hpgt1-1</i> (SALK_007547)	Intron	longer lateral roots, longer and denser root hairs, shorter inflorescence stems, shorter siliques (<i>hpgt2-1</i> , <i>hpgt3-1</i> , <i>hpgt1hpgt2hpgt3</i>); thicker roots, smaller rosette leaves, shorter petioles and reduced fertility (<i>hpgt1hpgt2hpgt3</i>)	Ogawa-Ohnishi and Matsubayashi, 2015
	HPGT2	At4g32120		Golgi	<i>hpgt2-1</i> (SALK_070368)	Exon		
	HPGT3	At2g25300		Golgi	<i>hpgt3-1</i> (SALK_009405)	Exon		
	At1g77810	At1g77810	β -1,3-galactosyltransferase	Golgi	-	-	-	Qu et al., 2008
	KNS4 (GALT14)	At1g33430		-	-	-	-	Suzuki et al., 2017
	GALT31A	At1g32930	β -1,6-galactosyltransferase	Golgi	FLAG_379B06	Exon	embryo development arrested at the globular stage	Geshi et al., 2013

Table 1 - (cont.).

CAZy GT family	Enzyme	Locus	Enzyme Activity	Subcellular localization	Mutants	T-DNA insertion	Mutants Phenotypes	References
GT37	FUT4	At2g15390	α-1,2-fucosyltransferase	-	<i>fut4</i> (SAIL_284_B05)	Exon	<i>fut4</i> , <i>fut6</i> , and <i>fut4fut6</i> mutants show reduced root growth under salt stress; <i>fut4fut6</i> double mutant lacks fucose in leaf and root AGPs	Wu et al., 2010; Liang et al., 2013; Tryfona et al., 2014
					<i>fut4-2</i> (SALK_125310)	Exon		
	FUT6	At1g14080		Golgi	<i>fut6</i> (SALK_099500)	Exon		
					<i>fut6-2</i> (SALK_078257)	Exon		
GT77	RAY1	At1g70630	β-arabinofuranosyltransferase	Golgi	<i>ray1-1</i> (SALK_053158)	Exon	reduced level of arabinose in their AGPs in etiolated seedlings, roots and rosette leaves; reduced root growth; reduced rosette and inflorescence size	Gille et al., 2013
					<i>ray1-2</i> (GK_001C09)	Exon		
-	AGM1	At1g27930	GlcA O-methyltransferase	Golgi	<i>agm1-1</i> (SALK_000253)	Exon	reduction on GlcA methylation (<i>agm1</i>); absence of GlcA methylation (<i>agm1agm2</i>)	Temple et al., 2019
	AGM2	At1g67330		-	<i>agm2-1</i> (GABI_054A04)	Exon		
					<i>agm2-2</i> (SALK_057182)	5' UTR		

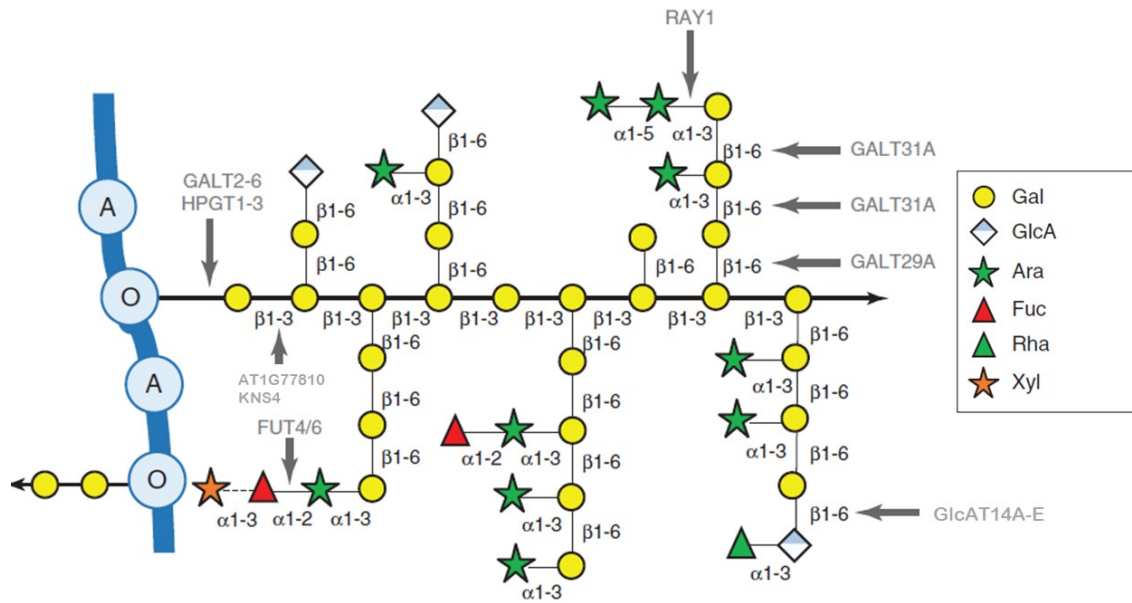


Fig. 1 - AGP type II glycan structure and sites of action of known glycosyltransferases acting on AGPs of Arabidopsis. Adapted from Ma et al. (2018).

remain to be identified, including α -arabinosyltransferases, α -rhamnosyltransferases and other β -GALTs (Showalter and Basu, 2016a).

1.1.1. Galactosyltransferases - the first ones to arrive

AGP glycosylation is initiated by the action of Hyp-O-GALTs, that add the first Gal onto Hyp residues (Showalter and Basu, 2016b). The eight GALTs specific for Hyp identified in Arabidopsis belong to the CAZy GT31 family (Coutinho et al., 2003; Showalter and Basu, 2016b). GALT2, GALT3, GALT4, GALT5 and GALT6 encode a GALT domain as well as a GALECTIN domain, whereas HPGT1, HPGT2 and HPGT3 lack a GALECTIN domain (Showalter and Basu, 2016b). Their activity was demonstrated by heterologous expression in *Pichia pastoris* (GALT2, GALT5) (Basu et al., 2013, 2015b), in *Nicotiana tabacum* leaf epidermal cells (GALT3, GALT4, GALT6) (Basu et al., 2015a, 2015b) and BY-2 cells (HPGT1, HPGT2 and HPGT3) (Ogawa-Ohnishi and Matsubayashi, 2015). *galt2*, *galt3*, *galt4*, *galt5*, *galt6*, *galt2galt5*, *hpgt1hpgt2hpgt3* mutants demonstrated lower GALT activities and reduced β -Yariv-precipitable AGPs compared to wt plants (Basu et al., 2013, 2015a, 2015b; Showalter and Basu, 2016a).

Some physiological phenotypes were revealed under normal growth conditions in GALT mutants: *galt2*, *galt3*, *galt5* and *galt2galt5* displayed reduced root hair length and density, *galt4* and *galt6* showed reduced seed production, *galt3*, *galt6* and *galt2galt5* presented reduced seed coat mucilage and *galt6* revealed accelerated leaf senescence. Besides, all GALT mutants (*galt2*, *galt3*, *galt4*, *galt5*, *galt6* and *galt2galt5*) roots and pollen tubes

exhibited less sensitivity to β -Yariv reagent, and root growth and root tip swelling was impaired under salt stress and in elevated levels of sucrose. *galt2galt5* displayed a large number of rosette leaves, delayed flowering and reduce silique length (Basu et al., 2013, 2015a, 2015b; Showalter and Basu, 2016a, 2016b). In addition, HPTG mutants exhibited several pleiotropic phenotypes such as longer lateral roots, increased root hair length and density, shorter inflorescence stems, shorter siliques (*hpgt2-1*, *hpgt3-1*, *hpgt1hpgt2hpgt3*), thicker roots, smaller rosette leaves, shorter petioles and reduced fertility (*hpgt1hpgt2hpgt3*) (Ogawa-Ohnishi and Matsubayashi, 2015; Showalter and Basu, 2016a, 2016b).

At1g77810 encodes a β -1,3-GALT that belongs to the GT31 family and likely functions in β -1,3-galactan backbone synthesis. This Golgi-located membrane enzyme was demonstrated to add Gal to a synthetic β -1,3-Gal disaccharide using heterologous expression in COS cells (Qu et al., 2008; Showalter and Basu, 2016b).

KSN4 is also a member of GT31 family. Heterologous expression in *N. benthamiana* showed β -1,3-GALT activity in AG glycans from AGPs and pectins. *ksn4* mutants present an abnormality in the exine of developing microspores. Immunolabelling showed that *ksn4* have reduced AGP content in the primexine of developing microspores. Furthermore, *ksn4* exhibit pollen aggregation and reduced fertility (shorter fruit lengths and lower seed set compared to wt) (Suzuki et al., 2017).

GALT31A is an Arabidopsis galactosyltransferase of CAZy family GT31 involved in elongation of β -1,6-galactan. Its activity was demonstrated by heterologous expression in *E. coli* and *N. benthamiana*. A mutation in this gene showed an abnormal asymmetric cell division in the hypophysis causing the arrest of embryo development at the globular stage. This phenotype reveals the importance of AG glycans in the normal development of the embryo (Geshi et al., 2013).

Dilokpimol et al. (2014) showed that GALT31 is co-expressed with GALT29A in the Golgi. GALT29A resides in the GT29 family and this enzyme recombinantly expressed in *N. benthamiana* possesses β -1,6-GALT activity. GALT29A and GALT31A interact as indicated by Förster resonance energy transfer (FRET), forming an enzyme complex that exhibited an increased level of its β -1,6-GALT activities when compared to GALT29 alone.

1.1.2. Glucuronosyltransferases - adding the special sugar

The CAZy GT14 family comprises numerous proteins with a variety of activities such as β -1,3-galactosyl-O-glycosyl-glycoprotein β -1,6-N-acetylglucosaminyltransferase, N-acetyllactosaminide β -1,6-N-acetylglucosaminyltransferase, protein O- β -xylosyltransferase and UDP-GlcA:arabinogalactan β -glucuronosyltransferase (Lombard et al., 2014). Several animal GT14 proteins (β -1,6-N-acetylglucosaminyltransferases and O- β -xylosyltransferases) have been elucidated through various studies (Bierhuizen et al., 1993; Kuhn et al., 2001; Yu et al., 2001; Wilson, 2002).

GT14 plant enzymes are phylogenetically related to the protein family containing Domain of Unknown Function 266 (DUF266) (Ye et al., 2011; Knoch et al., 2013). A mutation in brittle culm 10 (BC10), a DUF266-containing glycosyltransferase in rice, led to a significant decrease in mechanical strength and an alteration of the AG quantity and structure (Zhou et al., 2009).

Few years ago, Ye et al. (2011), based on a phylogenetic analysis on a shared protein domain, the Branch domain (PF02485), in *Arabidopsis*, *Oryza sativa*, *P. trichocarpa*, *Sorghum bicolor* and *Vitis vinifera*, proposed that the original GT14 genes and GT14-like (DUF266) genes belong to one family: GT14/GT14-like family. GT14-like subfamily has been suggested to be plant specific, however some bacteria genes were also found (Zhou et al., 2009; Ye et al., 2011).

The GT14 subfamily is composed of 11 homologous members in *Arabidopsis* and until now only six have been characterized (Pagnussat et al., 2005; Boavida et al., 2009; Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez, 2018). GlcAT14A, GlcAT14B, GlcAT14C, GlcAT14D and GlcAT14E add glucuronic acid to AGPs as confirmed by *in vitro* enzyme assays (GlcAT14B and GlcAT14C) (Dilokpimol and Geshi, 2014), recombinant expression in *P. pastoris* (GlcAT14A) (Knoch et al., 2013) and enzymatic hydrolysis and PACE analysis (GlcAT14A-E) (Lopez-Hernandez, 2018). GlcAT14A was transiently expressed in *N. benthamiana* and localized into the Golgi apparatus. *glcat14a* knockout mutants showed enhanced cell elongation during seedling growth and a reduction of GlcA substitution on Gal- β -1,6-Gal and β -1,3-Gal in their AGPs when compared to wt, indicating an *in vivo* role of GlcAT14A in the synthesis of those structures. Besides, an increase of Gal and a reduction of Ara was detected in AG extracts. The increase of galactosylation may result from the increase of O6 acceptor sites which are shared by GlcAT14A and galactosyltransferases and thus addition of GlcA may terminate Gal chain extension (Knoch et al., 2013). As several sugar levels were altered in these mutants, the observed phenotype is not only a consequence of the

reduced levels of GlcA but may be due to dynamic conformational changes of the glycans in AGPs (Knoch et al., 2014). *GlcAT14A* is co-expressed with *GALT31A* and *GALT29A* in the Golgi apparatus (Knoch et al., 2013; Dilokpimol et al., 2014).

Lopez-Hernandez (2018) showed that AGPs of mutants of *GlcAT14A*, *GlcAT14B*, *GlcAT14D* and *GlcAT14E* have reduced content of [Me]GlcA and Ca^{2+} -sensitive phenotypes. Some conditional growth phenotypes appear when mutants were grown in different concentrations of Ca^{2+} : *glcat14abd-1* presented reduction in stem length, *glcat14ab* and *glcat14abd-1* displayed smaller siliques, and *glcat14ab*, *glcat14abd-1* and *glcat14abd-2* showed reduced branching point in leaves trichomes. In addition, *glcat14ab* and *glcat14abd-1* presented ellipse shape in the third leaves, and *glcat14abe* exposed smaller etiolated hypocotyls.

Additionally, two large-scale insertional mutagenesis screens were performed and two mutants were found for *At3g03690* (Pagnussat et al., 2005; Boavida et al., 2009). *At3g03690* was named *UNFERTILIZED EMBRYO SAC 7 (UNE7)* as the mutant *une7* displayed unfertilized ovules but normal pollen tube attraction (Pagnussat et al., 2005). Posteriorly, Boavida et al. (2009) found another mutant named *pollen tube growth defective 12 (tgd12)* which had shorter PTs. However, the β -GlcAT activity for *UNE7* has not been confirmed yet.

1.1.3. Glucuronic acid methyltransferases - changing GlcA

In type II AGs from Arabidopsis, most of the GlcA substituted is 4-O-methylglucuronosyl (4-O-Me-GlcA) (Tryfona et al., 2012). Recently, two GlcA-specific methyltransferases have been identified. ArabinoGalactan Methyltransferases 1 (AGM1) and 2 (AGM2), two family members of DUF579 family, have GlcA-O-methylation activity on AGPs. AGM1 localized in the Golgi apparatus when transiently expressed in tobacco leaves. *agm1* mutant showed reduce methylation of GlcA on AGPs and in the *agm1agm2* double mutant there was no AG GlcA methylation (Temple et al., 2019).

1.1.4. Fucosyltransferases - continuing the hard work

FUT4 and FUT6, two Arabidopsis α -1,2-FUT of AGPs, belong to the plant GT37 CAZy family. Their enzymatic activities were demonstrated by heterologous expression in *N. tabacum* BY2 cells (Wu et al., 2010; Liang et al., 2013). Tobacco BY2 cells contain nonfucosylated AGPs but heterologous expression of these genes resulted in fucosylated AGPs (Wu et al., 2010; Knoch et al., 2014). Knock-out mutants (*fut4*, *fut6* and *fut4fut6*) grown under salt stress show reduced root growth (Liang et al., 2013;

Tryfona et al., 2014). Fuc was absent in *fut4* leaf AGPs whereas *fut4fut6* double mutants lack fucose both in leaf and root AGPs (Liang et al., 2013). It is likely that FUT4 and FUT6 are both responsible for AGP fucosylation in roots while FUT4 is also responsible in leaves (Wu et al., 2010). This conditional phenotype reveals the importance of this sugar in root growth and salt sensitivity (Showalter and Basu, 2016b). FUT6-GFP was transiently expressed in tobacco leaves and localized into the Golgi complex (Wu et al., 2010).

1.1.5. Rhamnosyltransferases - the ghost workers

Finally, *ray1* mutants presented reduced level of Ara in their AGPs in etiolated seedlings, roots and rosette leaves of Arabidopsis, leading to its name *REDUCED ARABINOSE YARIV1*. Knockout mutants also exhibit reduced root growth, reduced rosette and inflorescence size. RAY1 belongs to the GT77 family and its heterologous expression in *N. benthamiana* demonstrated β -arabinosyltransferase activity. However, β -linked Ara has not been reported in AGPs so it remains unclear whether RAY1 functions in the biosynthesis of AGPs glycans (Gille et al., 2013; Knoch et al., 2014; Showalter and Basu, 2016b).

1.2. Glycoside hydrolases - cutting all the sugar

Glycoside hydrolases (GHs; EC 3.2.1.x) hydrolyze the glycosidic bond between two molecules and therefore are important for the metabolism of AG sugar chains (Henrissat and Davies, 1997; Su and Higashiyama, 2018). The hydrolysis of AGP glycans requires several GHs, such as β -galactosidases, β -galactanases, α -arabinofuranosidases, β -arabinopyranosidases, β -glucuronidases, α -fucosidases and α -rhamnosidases (Knoch et al., 2014). These enzymes can be a useful tool to study the function of the sugars decorating AGPs (Su and Higashiyama, 2018). GHs are also important for the metabolism of AGPs (Knoch et al., 2014). AGPs on the transmitting tract of tobacco pistils are degraded releasing sugars that are used as resource for PT growth (Cheung et al., 1995; Knoch et al., 2014).

AtGUS2 (At5g07830) is a β -glucuronidase that belongs to the GH79 family. Knockout mutants showed increased levels of GlcA while plants over-expressing AtGUS2 showed the opposite phenotype, no GlcA in seedlings AGP fractions and increased cell elongation in seedlings, a similar phenotype of *atglcat14a* (Eudes et al., 2008; Knoch et al., 2014).

1.3. Calcium in sexual reproduction – anywhere, anytime...

Ca^{2+} is an important ubiquitous second messenger involved in many plant processes such as abiotic stress responses, stomatal aperture, self-incompatibility during fertilization, interactions with microorganisms and development of tip-growing structures such as PTs and root hairs (Dodd et al., 2010; Vanneste and Friml, 2013; Edel et al., 2017). Calcium signals correspond to transient increases in cytosolic free Ca^{2+} that arise from the flux of Ca^{2+} from the extracellular fluid or intracellular compartments such as the ER or vacuole (Dodd et al., 2010; Edel et al., 2017).

Plant sexual reproduction also relies on changes in Ca^{2+} homeostasis (Tekleyohans et al., 2017). It was demonstrated that when β -Yariv reagent binds to AG polysaccharides, an increase in intracellular Ca^{2+} is triggered (Pickard and Fujiki, 2005) and PT growth is inhibited (Roy et al., 1998). This suggested an involvement of AGPs, AG glycans and Ca^{2+} in reproduction.

When the PT physically interacts with the receptive synergid, a cytosolic Ca^{2+} oscillation is induced in the synergid cell (Denninger et al., 2014; Dresselhaus et al., 2016). Throughout PT discharge, Ca^{2+} signatures are triggered in the egg and central cells activating the female gametes (Denninger et al., 2014; Tekleyohans et al., 2017). During gamete fusion, a second peak of Ca^{2+} occurs in the egg cell (Denninger et al., 2014; Tekleyohans et al., 2017).

1.3.1. GlcA and Ca^{2+} - a deep connection

Firstly, AGPs seemed to act as signaling molecules per se but AGP receptors have not been found (Lamport et al., 2014). Recently, it was proposed that AGPs can bind and store Ca^{2+} through GlcA in a reversible and pH-dependent way at the plasma membrane (Fig. 2). It was demonstrated that *in vitro* AGPs could hold Ca^{2+} in a pH of 4-5 and as the pH was lowered Ca^{2+} was released. AGPs fully released Ca^{2+} at pH 2.5. Many reproductive processes rely on Ca^{2+} oscillations and therefore, some AGPs expressed in the reproductive tissues may be surface Ca^{2+} capacitors (Lamport and Várnai, 2013). The CAZy GT14 family (Lombard et al., 2014) plays an important role in Ca^{2+} signaling as it guarantees the presence of GlcA on AGPs. It is estimated that AGPs extracts have about 10% of GlcA (Tryfona et al., 2012). Knoch and colleagues (2013) suggested that GlcA is a terminator of the AGPs chains elongation since *glcat14a* mutant showed an increase in Gal and a reduction in Ara when compared to wt, an evidence of further elongation of the galactan backbone.

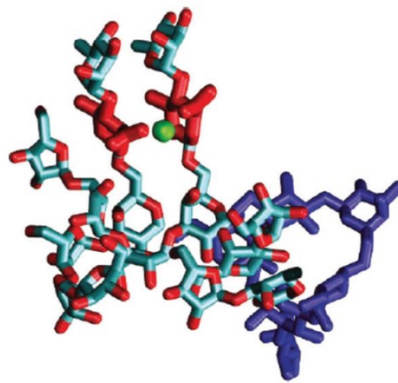


Fig. 2 - Three-dimensional molecular model simulating an Hyp-AG with bound Ca^{2+} .

The galactan backbone is in dark blue, the sidechains are in light blue, GlcA residues are in red and Ca^{2+} ion in green. Taken from Lamport et al. (2014).

4-O-methylation of GlcA changes the calcium binding affinity, thus modulating the calcium release response to pH. Moreover, the addition of the methyl group to GlcA prevents the addition of 4-linked sugars, such as Rha, and will also prevent extension of 4-linked side chains to the GlcA of AG as seen in APAP1 (Temple et al., 2019). APAP1 (ARABINOXYLAN PECTIN ARABINO GALACTAN PROTEIN1; At3g45230) is the only AGP covalently linked to pectin by a GlcA residue and hemicellulose via an Ara residue (Tan et al., 2013).

The disaccharide 4-Me-GlcA- β -(1-6)-Gal, AMOR, is responsible for PTs to respond to LURE attractant peptides secreted by the synergid cells in *Torenia fournieri* (Mizukami et al., 2016; Jiao et al., 2017). AMOR is likely derived from AGPs and denotes the importance of [Me]GlcA.

The triple mutant *glcat14abd-1* contained 80% less [Me]GlcA than wt AGPs and bound less 83% of Ca^{2+} than wt plants (Lopez-Hernandez, 2018), validating the hypothesis of AGP- Ca^{2+} interaction via [Me]GlcA.

1.3.2. PT growth and Ca^{2+} - a co-dependent relationship

PTs are one of the fastest growing eukaryotic cells (Rounds and Bezanilla, 2013; Johnson et al., 2019). PTs have evolved in order to enter in one ovule as fast as they can, so they can deliver the sperm cells and transmit its information to the next plant generation (Williams et al., 2016; Johnson et al., 2019). The growth machinery in a PT is located in the tube's apex where new plasma membrane and cell wall components are delivered through the secretory pathway (Johnson et al., 2019). The wall materials such as methylesterified pectins are deposited at the tip and their maturation occurs along the subapical and shank regions through the de-esterification of pectins by pectin

methylesterases (PMEs). This reaction allows pectins to form Ca^{2+} -salt bridges with other polymers, creating a pectin gel that helps rigidify the wall (Johnson et al., 2019).

The signal guiding PT growth may be the Ca^{2+} gradient generated by Ca^{2+} dissociation from AGPs in the transmitting tissue (Lamport et al., 2014). AGPs have higher affinity to Ca^{2+} than pectin at low pH (Lamport et al., 2018). PTs have been reported to acidify their growth medium when grown *in vitro* (Feijó et al., 1995). Therefore, *in vivo* PTs may dissociate Ca^{2+} from AGPs of the transmitting tissue. PT growth has been explained using Hechtian strands and the AGP- Ca^{2+} capacitor model (Lamport et al., 2018). As PT grows, Hechtian strands are stretched leading to the acidification of the apoplast via activation of H^+ -ATPase proton pump. Consequently, Ca^{2+} is dissociated of AGPs resulting in a Ca^{2+} influx that activates exocytosis of cell wall precursors.

1.4. Objectives

Since AGPs are decorated with 90% of AG polysaccharides which coat their molecular surface it is likely that these sugars are essential to AGP function. Furthermore, the [Me]GlcA residues of AGPs can bind and store Ca^{2+} at the plasma membrane. Therefore, the aim of this chapter was to clarify the AGP- Ca^{2+} connection through the [Me]GlcA during the sexual reproduction in Arabidopsis. To accomplish this objective, the expression of GlcAT genes was evaluated, the subcellular localization of GlcAT was assessed and GlcAT mutants were functionally characterized during Arabidopsis sexual reproduction.

2. MATERIAL AND METHODS

2.1. Plant material and growth conditions

2.1.1. Plant lines

All *A. thaliana* (L.) Heynh. plants used in this work were in Col-0 ecotype. Wt seeds were obtained from NASC. A T-DNA insertion line for *At4g27480* was selected using the SIGnAL database and was obtained from NASC (Table 2). The *glcat14b-1*, *glcat14e-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2* and *glcat14a±be* mutant lines were kindly shared by professor Paul Dupree (University of Cambridge, England). The *glcat14a-2* mutant line used in this study was previously characterized by Knoch et al. (2013).

Table 2 - List of *A. thaliana* T-DNA insertion lines used.

Accession number	Locus	Gene name	Mutant name	T-DNA insertion	NASC Stock ID
NM_123362	At5g39990	<i>GlcAT14A</i>	<i>glcat14a-2</i>	SALK_043905	N543905
NM_121509	At5g15050	<i>GlcAT14B</i>	<i>glcat14b-1</i>	SALK_080923	N580923
NM_113309	At3g24040	<i>GlcAT14D</i>	<i>glcat14d-1</i>	GK363F05	N377379
			<i>glcat14d-2</i>	GK508D01	N326773
NM_112398	At3g15350	<i>GlcAT14E</i>	<i>glcat14e-1</i>	SALK_022820	N522820
NM_118883	At4g27480	-	<i>at4g27480-1</i>	SALK_105434C	N686608

2.1.2. Growth in soil

Wt *Nicotiana benthamiana* and Arabidopsis plants were sown directly on soil (Seed raising mix, Growth Better). Arabidopsis plants were kept for 48 h at 4°C in the dark to induce stratification. Afterwards, all plants were grown under a long day conditions (16 h light at 22°C and 8 h darkness at 18°C) with 50-60% relative humidity and light intensity at 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

For phosphinothricin acetyltransferase selection, the seedlings comprising the promotor fusions were sprayed with 100 mg l⁻¹ of glufosinate ammonium (BASTA®; Bayer Crop Science) for three times during a nine-day period.

2.1.3. Hydroponic Growth

Hydroponic growth followed the method described by (Lopez-Hernandez, 2018). Seeds were sterilized (Chapter 1; 2.1.3.1) and sown on plates containing MS medium (Chapter 1; 2.1.3.2) as described previously (Fig. 3A). New squared pots with drainage holes

(9.5x8.5x9.5 cm; Garden City Plastics) were filled in with a layer of 1 cm of vermiculite (Brunnings) and a square (8x8x7 cm) of rockwool (Grodan Classic). The drainage holes were filled with soft foam to avoid the loss of vermiculite while allowing the solution to flow through (Fig. 3B). The prepared pots were washed with deionized water for two days, changing the water once a day. The washed pots were covered with aluminum foil to avoid evaporation and contamination (Fig. 3C). Two-week-old plants grown *in vitro* were transferred to the pots through openings in the foil. Four plants were transplanted to each pot and in the middle of the pot an opening was used to water each pot with 250 mL of nutrient solution containing 0.2 mM (low calcium nutrient solution - LCNS) or 2 mM (basal nutrient solution - BNS) of Ca^{2+} , adjusted to pH 5.8 using KOH (Sigma-Aldrich), every fifth day. The full solutions composition is described in Table 3. All plants were grown in a long-day growth room (16 h light at 22°C and 8 h darkness at 18°C) with 50-60% relative humidity and light intensity at $180 \mu\text{mol m}^{-2} \text{s}^{-1}$ (Fig. 3D).

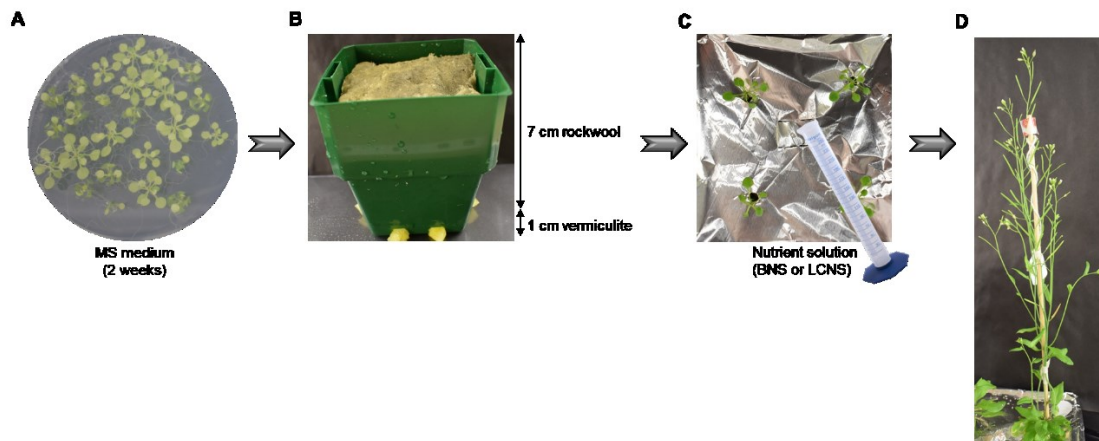


Fig. 3 - Representation of the *in vitro* and hydroponic growth.

(A) Seed germination and seedling growth on MS medium plates for two weeks. (B) Pots with drainage holes filled with soft foam, a layer of 1 cm of vermiculite and a square (8x8x7 cm) of rockwool. (C) Seedling transplantation to pots covered with aluminum foil. The pots were watered with 250 mL of nutrient solution every fifth day. (D) Aspect of a plant grown in these conditions.

Table 3 - Nutrient solutions composition.

	Final Concentration	
	LCNS	BNS
Macronutrients		
KNO ₃ (Sigma-Aldrich)		5.0 mM
MgSO ₄ ·7H ₂ O (VWR)		2.0 mM
Ca(NO ₃) ₂ ·4H ₂ O (Sigma-Aldrich)	0.2 mM	2.0 mM
K ₂ HPO ₄ (Sigma-Aldrich) + KH ₂ PO ₄ (Sigma-Aldrich) pH5.5		2.5 mM
NaFe(III)EDTA (Sigma-Aldrich)		50.0 µM
Micronutrients		
H ₃ BO ₃ (VWR)		70.0 µM
MnCl ₂ ·4H ₂ O (Sigma-Aldrich)		14.0 µM
CuSO ₄ ·5H ₂ O (Sigma-Aldrich)		0.5 µM
ZnSO ₄ ·7H ₂ O (Sigma-Aldrich)		1.0 µM
Na ₂ MoO ₄ ·2H ₂ O (Sigma-Aldrich)		0.2 µM
NaCl (Sigma-Aldrich)		10.0 µM
CoCl ₂ ·6H ₂ O (Sigma-Aldrich)		0.1 µM

2.2. Constructs generation

2.2.1. DNA amplification and purification

gDNA was extracted according to Edwards et al. (1991). Genomic promotor fragments up to 2500 bp positioned upstream of the start codon of *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* were amplified with the primer pairs described in Supplementary table 4. The coding sequence (CDS) of *GlcAT14B*, *GlcAT14E* (*AT3G15350.1*), *At4g27480.1*, *At1g71070*, *At1g53100* and *At1g03520.1* were amplified with the primer pairs described in supplementary table 5.

All DNA fragments were amplified by PCR using Platinum™ Taq DNA Polymerase (Invitrogen™) and the respective annealing temperatures were estimated using the Tm Calculator (Thermo Fisher Scientific). All PCR products were separated in a 1% (w/v) agarose (Grisp) gel in 1X TAE [40 mM Tris-HCl (Sigma-Aldrich) and 1 mM EDTA (Sigma-Aldrich), pH 8.3] with SYBR™ Safe DNA Gel Stain (Invitrogen™). DNA Gel Loading Dye (6X) (Thermo Scientific™) was added to each sample prior to loading in gel and 1 Kb Plus DNA Ladder (Invitrogen) was used as a molecular weight marker. The electrophoretic separation was conducted in a horizontal electrophoresis system (Bio-Rad) at 80-100V for 20-35 min. The amplified DNA fragments were purified from the agarose gel using PureLink™ Quick Gel Extraction Kit (Invitrogen™) in line with the

manufacturer's procedure. The DNA was quantified in a NanoDrop™ One UV-Vis Spectrophotometer (Thermo Scientific™).

2.2.2. Cloning

The amplified fragments were cloned into pCR™8/GW/TOPO™ (Invitrogen™). The resulting promoter fragments were subsequently transferred into pMT589, a modified version of pMDC111 (Curtis and Grossniklaus, 2003) with a triple nuclear localization signal (NLS) of yellow fluorescent proteins (YFP) and phosphinothricin acetyltransferase selection, and CDS fragments were transferred into pMDC83 (Curtis and Grossniklaus, 2003) by an LR reaction using the Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen™). All constructs were confirmed by DNA sequencing and the sequencing data were analyzed using Geneious Prime (<https://www.geneious.com>).

2.2.3. Bacteria chemical transformation

The entry clone was transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen™) following the instructions manual with the following modifications. 250 µL of pre-warmed LB medium (BD Difco™) 2.5% (w/v) was added for cell recovery. The transformed cells carrying the entry clone were plated in LB agar (BD Difco™) 4% (w/v) containing spectinomycin (50 µg/mL) while the ones carrying the expression clone were plated in LB agar supplemented with kanamycin (50 µg/mL).

A. tumefaciens AGL1 chemically competent cells were transformed according to Sparkes et al. (2006) with minor changes. For cell recovery, 1 mL of warmed LB medium (BD Difco™) 2.5% (w/v) was added. The cells were spread on LB agar (BD Difco™) 4% (w/v) plates containing rifampicin (50 µL/mL) and kanamycin (50 µL/mL).

2.2.4. Restriction analysis

To confirm the presence and the orientation of the desired fragment in the entry clone, a restriction digestion assay was performed using restriction endonucleases from New England Biolabs according to the manufacturer's instructions. The enzymes used and the expected fragments sizes obtained after digestion are discriminated on Table 4. The digestion fragments were separated by agarose gel electrophoresis as mentioned in the section 2.2.1.

The plasmid DNA of positive colonies carrying the entry and expression clones were sequenced and the sequencing data were analyzed using Geneious Prime

Table 4 - Enzymes used in the restriction analyses of constructs and expected fragments sizes.

Constructs	Enzymes	Fragments Sizes
<i>pCR8:GlcAT14Bprom</i>	EcoRV-HF®	3245 bp + 295 bp
<i>pCR8:At4g27480prom</i>		2925 bp + 1381 bp
<i>pCR8:At1g53100prom</i>		3661 bp + 856 bp
<i>pCR8:At1g71070</i>		3879 bp + 648 bp
<i>pCR8:At1g03520</i>		3650 bp + 770 bp
<i>pCR8:At1g71070prom</i>	EcoRV-HF® + SacI-HF®	3429 bp + 222 bp
<i>pCR8:GlcAT14E</i>		4253 bp + 194 bp
<i>pCR8:At4g27480</i>	EcoRV-HF® + Scal-HF®	3824 bp + 473 bp
<i>pCR8:At1g53100</i>		3627 bp + 994 bp
<i>pCR8:GlcAT14Eprom</i>	EcoRV-HF® + HindIII-HF®	4409 bp + 898 bp
<i>pCR8:At1g03520prom</i>		2944 bp + 814 bp
<i>pCR8:GlcAT14B</i>	EcoRV-HF® + BamHI-HF®	2778 bp + 960 bp + 711 bp
<i>pCR8</i>	EcoRV-HF® + SacI-HF® + Scal-HF® + HindIII-HF® + BamHI-HF®	2817 bp

(<https://www.geneious.com>). The bacterial colonies of interest were stored in 25% (v/v) glycerol (Sigma-Aldrich) at -80°C.

2.3. Transient expression and subcellular localization in *N. benthamiana*

Six-week-old *N. benthamiana* leaves were infiltrated with *A. tumefaciens* AGL1 as described (Sparkes *et al.*, 2006) with the following modifications. The infiltration medium was composed of 10 mM MgCl₂ (Sigma-Aldrich), 10 mM MES (Sigma-Aldrich) and 100 µM acetosyringone (Sigma-Aldrich).

The GmMan1 soybean α-1,2-mannosidase I, cytoplasmic tail and transmembrane domain fused to mCherry was used as a Golgi marker. The signal peptide of AtWAK2 (Arabidopsis wall-associated kinase 2) at the N-terminus of the mCherry and the ER retention signal His-Asp-Glu-Leu at its C-terminus served as ER marker. The full-length coding region of a plasma membrane aquaporin (AtPIP2A) fused to mCherry was utilized as plasma membrane marker (Nelson *et al.*, 2007). The used OD₆₀₀ was 0.3 for all constructs. Fluorescent signals in leaf epidermal sections were analyzed after 3 days of infiltration by confocal laser scanning microscopy using a Leica SP5 Spectral Scanning Confocal Microscope with a 63 x water immersion objective. For imaging the expression of GFP constructs, the excitation line was 488 nm whereas for mCherry constructs, the excitation line was 561 nm.

2.4. Protein and genomic sequences

The CAZy database (Lombard et al., 2014; <http://www.cazy.org/>) was used to identify the members of the GT14 family in Arabidopsis and obtain the respective accession numbers. Genomic and protein sequences were retrieved from the NCBI (At5g39990, NP_198815; At5g15050, NP_197009; At2g37585, NP_565866; At3g24040, NP_189046; At3g15350, NP_974319; At4g27480, NP_194478; At1g71070, NP_565009; At1g53100, NP_175718; At1g03520, NP_171851; At3g03690, NP_187019; At4g03340, NP_192243).

Arabidopsis GT14 homologue protein sequences of the species *P. patens* (Pp3c10_2430; Pp3c14_2510; Pp3c17_12000; Pp3c19_19350; Pp3c26_670; Pp3c9_900; Pp3c9_903), *S. moellendorffii* (SMO351G0567; SMO367G0106), *H. vulgare* (HVVU0036G1048; HVU0037G0199; HVU0037G3800; HVU0038G0657; HVU0038G1074; HVU0038G1425; HVU0040G0096; HVU0040G1221; HVU0041G0231; HVU0045G0946; HVU0045G0988) and *P. trichocarpa* (Potri.001G053800; Potri.001G399500; Potri.004G130300; Potri.006G263000; Potri.008G006500; Potri.008G128000; Potri.009G003300; Potri.010G114800; Potri.010G251900; Potri.011G119600; Potri.013G066200; Potri.013G145500; Potri.015G107100; Potri.017G075600; Potri.019G040600; Potri.019G104600) were retrieved from PLAZA (Monocots PLAZA 4.0) from homologous gene family HOM04M000287 using *At5g39990* as a bait (Van Bel et al., 2018).

3. RESULTS

3.1. In silico analysis of GT14 family

The CAZy GT14 family is composed of 11 members in *A. thaliana*. Five of these members (GlcAT14A, GlcAT14B, GlcAT14C, GlcAT14D and GlcAT14E) have been described with β -GlcAT activity specific of type II AGs of AGPs (Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez, 2018).

The AGI codes of the GT14 family genes were identified in the CAZy database (<http://www.cazy.org/>; Lombard et al., 2014) and the respective genomic and protein sequences were retrieved from the NCBI. Full length of gDNA sequences of GT14 family diverged between 2181 bp (*GlcAT14C*) and 2946 bp (*GlcAT14E*) (Table 5). Regarding the ORF, it varied between 1137 bp (*UNE7*) and 1347 bp (*At4g03340*). Subsequently, the length of GT14 proteins sequences ranged from 378 (43.8 kDa) to 448 (51.7 kDa) amino acids long. Six genes had only one transcription variant, four genes had two and *At4g27480* had three.

Table 5 - Arabidopsis GT14 coding region information.

Gene	Locus	Full length genomic DNA (bp)	Coding region (bp)	Amino acids	Molecular weight (kDa)	Splice variant
<i>GlcAT14A</i>	At5g39990	2571	1344	447	51.1	AT5G39990.1
<i>GlcAT14B</i>	At5g15050	2153	1305	434	49.6	AT5G15050.1
<i>GlcAT14C</i>	At2g37585	2181	1155	384	44.5	AT2G37585.1
<i>GlcAT14D</i>	At3g24040	2413	1254	417	47.2	AT3G24040.1
<i>GlcAT14E</i>	At3g15350	2946	1275	424	48.5	AT3G15350.1 AT3G15350.2
<i>At4g27480</i>	At4g27480	2261	1266	421	47.8	AT4G27480.1 AT4G27480.2 AT4G27480.3
<i>At1g71070</i>	At1g71070	2338	1188	395	44.9	AT1G71070.1
<i>At1g53100</i>	At1g53100	2195	1272	423	48.2	AT1G53100.1
<i>At1g03520</i>	At1g03520	2399	1344	447	51.5	AT1G03520.1 AT1G03520.2
<i>UNE7</i>	At3g03690	2464	1137	378	43.8	AT3G03690.1 AT3G03690.2
<i>At4g03340</i>	At4g03340	2354	1347	448	51.7	AT4G03340.1 AT4G03340.2

A search using Pfam (El-Gebali et al., 2019) revealed only one predicted domain, branch domain (PF02485), in the Arabidopsis GT14 enzymes (Table 6). This protein family includes core-2/I-branching enzymes. The 11 protein sequences were submitted in the TMHMM 2.0 (Sonnhammer et al., 1998) and ARAMEMNON (Schwacke et al., 2003) servers to predict the presence of TMDs. All sequences were predicted to have a single

hydrophobic TMD close to the N-terminus (Table 6). The presence of a signal peptide was also tested by running the sequences in the SignalP 5.0 server (Almagro Armenteros et al., 2019). Apart from GlcAT14C, all analyzed proteins were not predicted to have a signal peptide in the N-terminus (Table 6). GlcAT14C was predicted to have a cleavage site between position 33 and 34. Profilescan (Sigrist et al., 2010) recognized the potential N-glycosylation sites for the GT14 proteins (Table 6).

Table 6 - Arabidopsis GT14 predicted domains and sites.

TMHMM 2.0 (Sonnhammer et al., 1998) and ARAMEMNON (Schwacke et al., 2003) predicted the amino terminus (cytoplasmatic), transmembrane and luminal positions. SignalP 5.0 (Almagro Armenteros et al., 2019) predicted the presence of signal peptides. Pfam (El-Gebali et al., 2019) predicted the presence and position of the branch domain (PF02485). Profilescan (Sigrist et al., 2010) predicted the N-glycosylation sites.

Protein	Amino Terminus	Signal Peptide	Transmembrane	Lumen	Branch (PF02485)	N-glycosylation sites
GlcAT14A	1-33	-	34-56	57-447	101-361	151, 200, 329, 405
GlcAT14B	1-21	-	22-44	45-434	88-348	138, 187, 316, 392
GlcAT14C	1-11	1-33	12-34	35-384	59-316	156, 285, 306
GlcAT14D	1-11	-	12-29	30-417	69-329	197, 310, 374
GlcAT14E	1-11	-	12-34	35-424	81-341	54, 180, 235, 335
At4g27480	1-11	-	12-34	35-421	80-340	48, 53, 65, 179
At1g71070	1-6	-	7-29	30-395	49-309	170, 277
At1g53100	1-23	-	24-46	47-423	83-341	56, 65, 182, 237, 247, 354
At1g03520	1-36	-	37-59	60-447	103-363	91, 202, 222
UNE7	1-6	-	7-26	27-378	54-313	48, 152, 207, 281
At4g03340	1-35	-	36-58	59-448	104-364	72, 91, 203, 223

The full protein sequences were aligned using Clustal W (Thompson et al., 1994) and a sequence identity matrix was generated (Fig. 4). Examining the percentage of identity between the sequences, three pairs of proteins point out. At1g03520 and At1g03340 shared nearly 79% of sequence identity. In like manner, GlcAT14A and GlcAT14B shared almost 77% of sequence identity. GlcAT14E and At1g53100 shared a 75% of sequence identity. At4g27480 also shared 68% and 66% of sequence identity with GlcAT14E and At1g53100, respectively. GlcAT14D and GlcAT14E only shared 39% of sequence identity. However, both enzymes have GlcAT activity (Lopez-Hernandez, 2018) suggesting that regardless of a low percentage of sequence identity between these family members, the 11 enzymes may all have a GlcAT activity.

Protein	GlcAT14A	GlcAT14B	GlcAT14C	GlcAT14D	GlcAT14E	UNE7	At4g27480	At1g71070	At1g53100	At1g03520	At1g03340
GlcAT14A											
GlcAT14B	76,92										
GlcAT14C	55,85	52,91									
GlcAT14D	44,74	44,33	42,25								
GlcAT14E	58,75	57,32	51,61	39,4							
UNE7	55,5	55,32	50,69	41,26	56,42						
At4g27480	59,52	59,76	48,66	38,1	68,33	55,76					
At1g71070	46,77	46,04	45,23	44,53	46,35	44,26	44,79				
At1g53100	54,57	53,98	47,85	39,69	75	55,96	65,52	42,93			
At1g03520	54,11	53,74	51,19	41,85	52,18	52,01	53,41	44,59	49,76		
At1g03340	52,39	52,21	51,05	41,02	51,57	54,01	51,82	45,24	50,36	78,52	

Fig. 4 - Percent of identity matrix between aligned GT14 protein sequences from Arabidopsis.

The percent identity was calculated using Clustal W (Thompson et al., 1994). The values in blue represent high sequence identity whereas the values in red represent low sequence identity.

The sequence identity was reflected in the phylogenetic tree obtained from the Arabidopsis GT14 aligned protein sequences (Fig. 5). A few pairs of paralog genes were evident such as *GlcAT14D/At1g71070*, *At1g03520/At4g03340*, *GlcAT14A/GlcAT14B* and *GlcAT14E/At1g53100*. The eleven protein sequences clustered into 4 clades: *GlcAT14D/At1g71070/GlcAT14C* (green), *At1g03520/At4g03340* (pink), *GlcAT14A/GlcAT14B* (orange) and *UNE7/At4g27480/GlcAT14E/At1g53100* (blue) (Fig. 5).

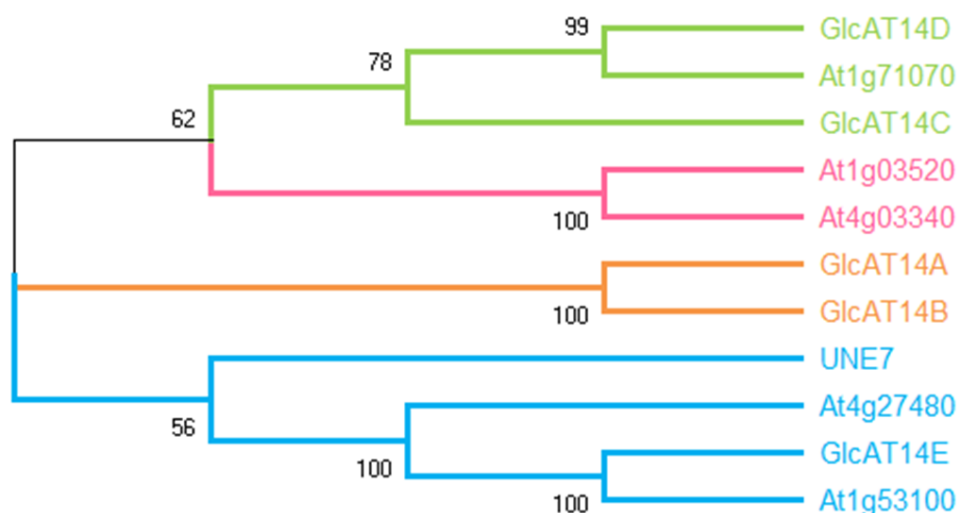


Fig. 5 - Phylogenetic tree of Arabidopsis GT14 proteins sequences.

The 11 GT14 protein sequences form four clusters (green, pink, orange and blue). The phylogenetic tree was obtained with MEGA-X (Kumar et al., 2018) using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. This analysis involved 11 amino acid sequences. After deletion of ambiguous positions, there were a total of 467 positions in the final dataset.

The distribution of the GT14 genes along the five chromosomes of *Arabidopsis* displayed no evidence of clustering (Fig. 6). Chromosome 2 had 1 gene (*GlcAT14C*), chromosomes 4 and 5 had 2 genes (*At4g03340* and *At4g27480*, *GlcAT14B* and *GlcAT14A*, respectively), and chromosomes 1 and 3 had 3 genes (*At1g03520*, *At1g53100* and *At1g71070*; *UNE7*, *GlcAT14E* and *GlcAT14D*; respectively).

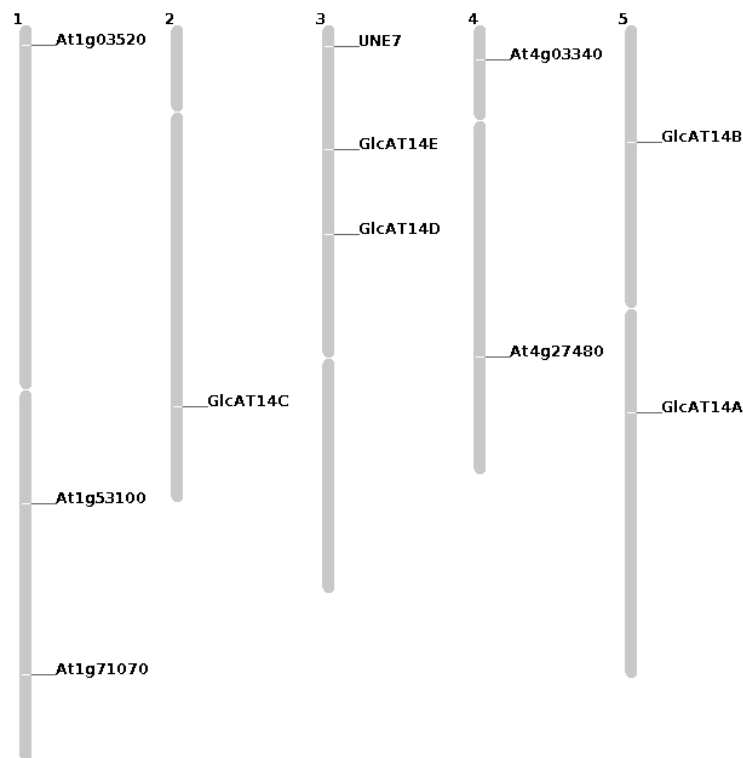


Fig. 6 - GT14 genes distribution on *A. thaliana* chromosomes.

The localization of the chromosomes was obtained on the Chromosome Map Tool available at TAIR.

To identify recent functional divergence from ancient proteins, homologous protein sequences to the GT14 family from early land plants to higher plants were retrieved from PLAZA (Van Bel et al., 2018). Protein sequences from *A. thaliana*, *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa* were aligned using Clustal W (Thompson et al., 1994) and a phylogenetic tree was generated using MEGA X (Kumar et al., 2018) (Fig. 7). However, three protein sequences (Potri.019G040600, HVU0037G3800 and HVU0036G1048) were not used to generate the tree as their sequence did not align with majority. The number of homologous proteins varied between these species: *P. patens* (7), *S. moellendorffii* (2), *H. vulgare* (11) and *P. trichocarpa* (16). GT14 genes seem to be present in different plants among embryophytes. GT14 genes seem to increase from mosses to angiosperms except for *S. moellendorffii*, a lycophyte, which only had two putative GT14 genes.

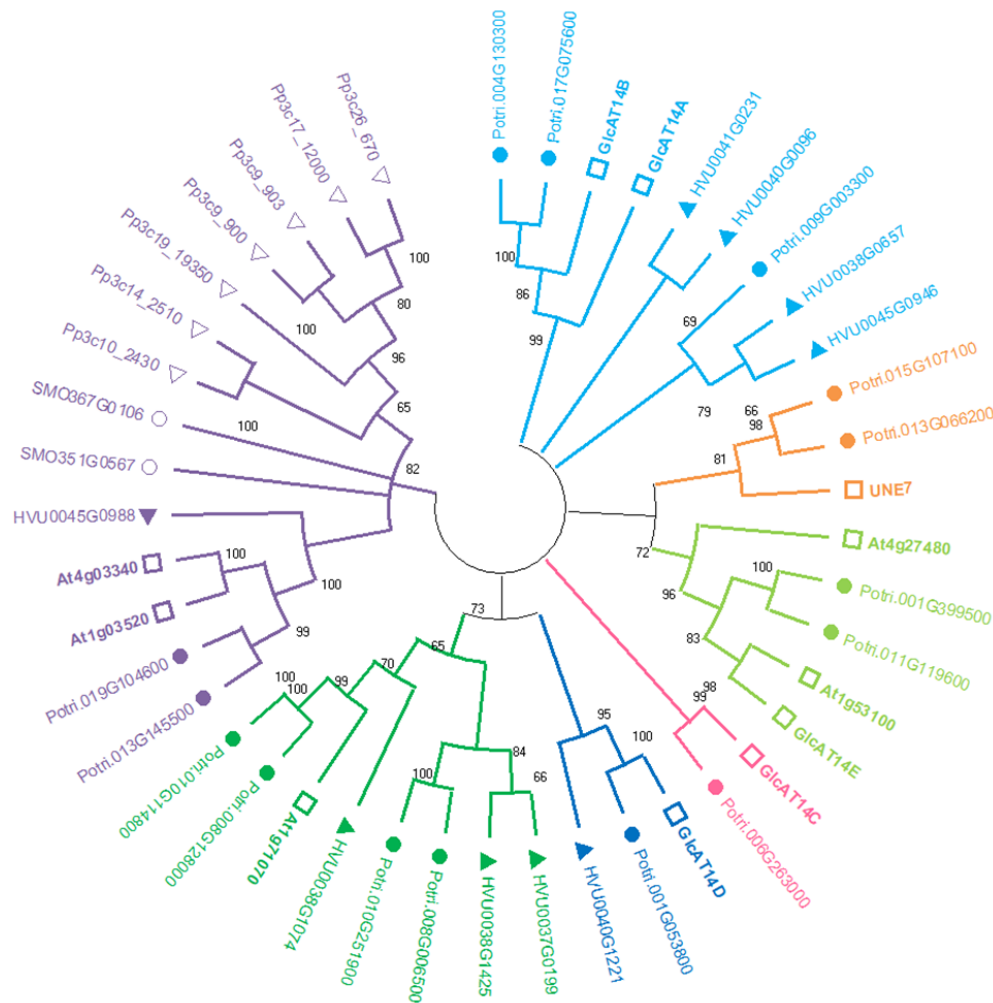


Fig. 7 - Phylogenetic tree of GT14 proteins and putative homologues in *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa*.

The phylogenetic tree was created using MEGA X (Kumar et al., 2018) using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. This analysis involved 44 amino acid sequences. After deletion of ambiguous positions, there were a total of 581 positions in the final dataset. The tree is divided into seven clades (light blue, orange, light green, pink, blue, green and violet). Each plant species is identified with a different shape (Δ , *P. patens*; $\circ$, *S. moellendorffii*, $\blacktriangle$, *H. vulgare*; $\bullet$, *P. trichocarpa*; $\square$, *A. thaliana*). GT14 proteins from *A. thaliana* are highlighted in bold.

The protein sequences of *A. thaliana*, *P. patens*, *S. moellendorffii*, *H. vulgare* and *P. trichocarpa* clustered into 7 groups (Fig. 7). The protein sequences of *P. patens* (7) and *S. moellendorffii* (2) all grouped into one unique clade (violet) along with one *H. vulgare*, two Arabidopsis and two *P. trichocarpa* sequences, supporting that they share a common ancestor. Besides, GlcAT14B, UNE7 and At1g71070 may be duplicated in *P. trichocarpa* (Potri.004G130300/Potri.017G075600, Potri.013G066200/Potri.015G107100 and Potri.010G114800/Potri.008G128000, respectively).

3.2. Expression profile of GT14 genes

In order to explore the functional diversity of GT14 genes in Arabidopsis, a spatial and developmental expression analysis was performed. RNA was isolated from different tissues of Arabidopsis and analyzed by qPCR (Supplementary figure 6). The qPCR fragments were purified and quantified by HPLC (Supplementary figure 7-9). The size in base pairs and identity of qPCR product was confirmed by sequencing (Supplementary fig. 10-12). qPCR data indicated that GT14 genes were broadly expressed and had overlapping but distinct expression patterns (Supplementary figure 6). In Fig. 8, qPCR data was expressed as a ratio of maximal levels observed. The expression data was gathered into three different groups to allow easy interpretation and to check if the clades in Fig. 5 shared the same expression pattern.

GlcAT14A, *GlcAT14B*, *At1g03520* and *At4g03340* were expressed at their maximal rate in the stems (Fig. 8A). *GlcAT14B* was also highly expressed in cauline leaves whereas *GlcAT14A* was moderately transcribed in seedlings, cauline leaves and siliques. The levels of *At1g03520* and *At4g03340* were generally low. In the RNA-seq data, *GlcAT14A* and *At1g03520* were most transcribed in roots and in mature leaves, respectively (Fig. 8B). Additionally, high levels of transcripts of *At1g03520* were detected in internode, flower 1 and silique 4. *At4g03340* was also highly transcribed in roots. According to the qPCR data, *At1g03520* and *At4g03340* shared the same expression pattern (Fig. 8A).

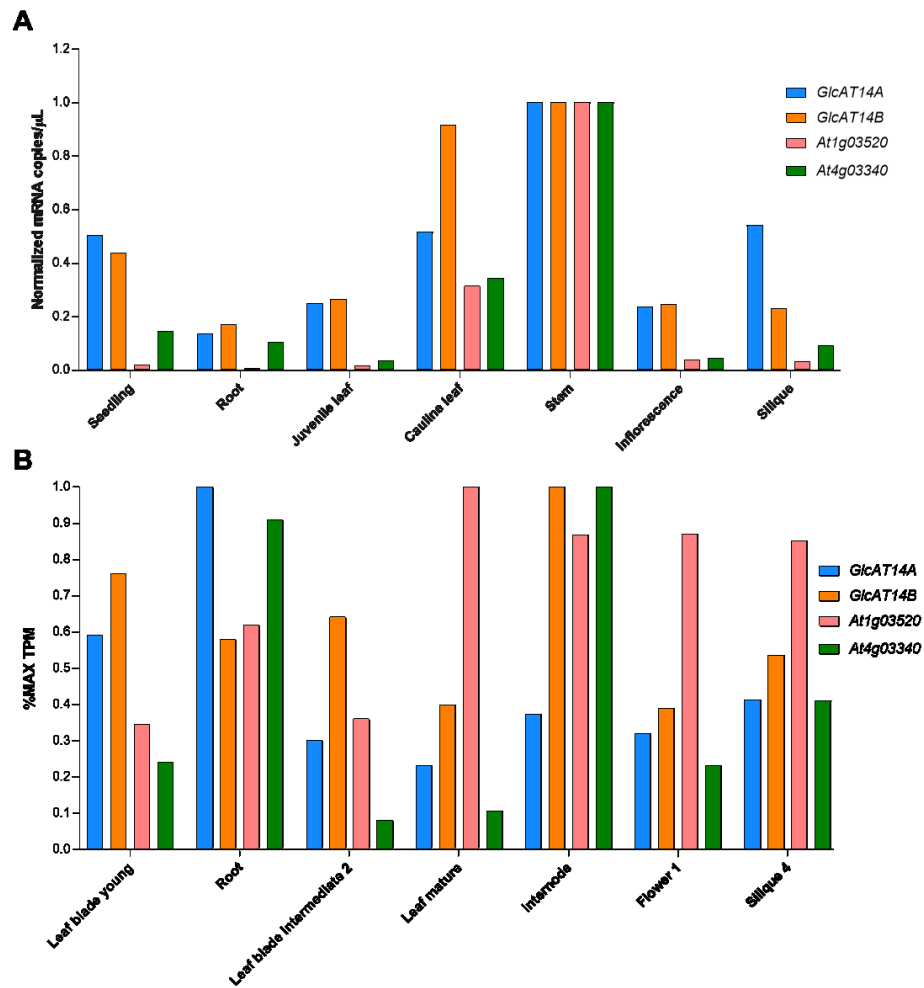


Fig. 8 - Expression pattern of GT14 genes in different tissues in Arabidopsis.

(A) Normalized transcript levels of *GlcAT14A*, *GlcAT14B*, *At1g03520* and *At4g03340* in a range of tissues. Means of three biological replicates are expressed as a ratio of maximal levels observed. Values of 1.0 therefore indicate the tissue in which the gene is transcribed at its maximal rate. (B) Percentage of TPM of *GlcAT14A*, *GlcAT14B*, *At1g03520* and *At4g03340* expressed as a ratio of maximal levels observed for the specific gene based on RNA-seq data (Klepikova et al., 2016).

GlcAT14E, *At4g27480* and *UNE7* were most transcribed in stems while *At1g53100* was most transcribed in siliques (Fig. 9A). *GlcAT14E* was moderately transcribed in seedlings, cauline leaves, inflorescences and siliques. Lower levels of expression of *At4g27480* and *At1g53100* were found in seedlings, roots, juvenile leaves and inflorescences. In contrast, the RNA-seq data showed that *At4g27480* and *At1g53100* were transcribed at their maximal rate in mature leaves whereas *UNE7* was most transcribed in roots (Fig. 9B). High levels of expression for *GlcAT14E* and *At4g27480* were expected in roots and in flowers. This group of four genes did not share the same pattern regarding gene expression.

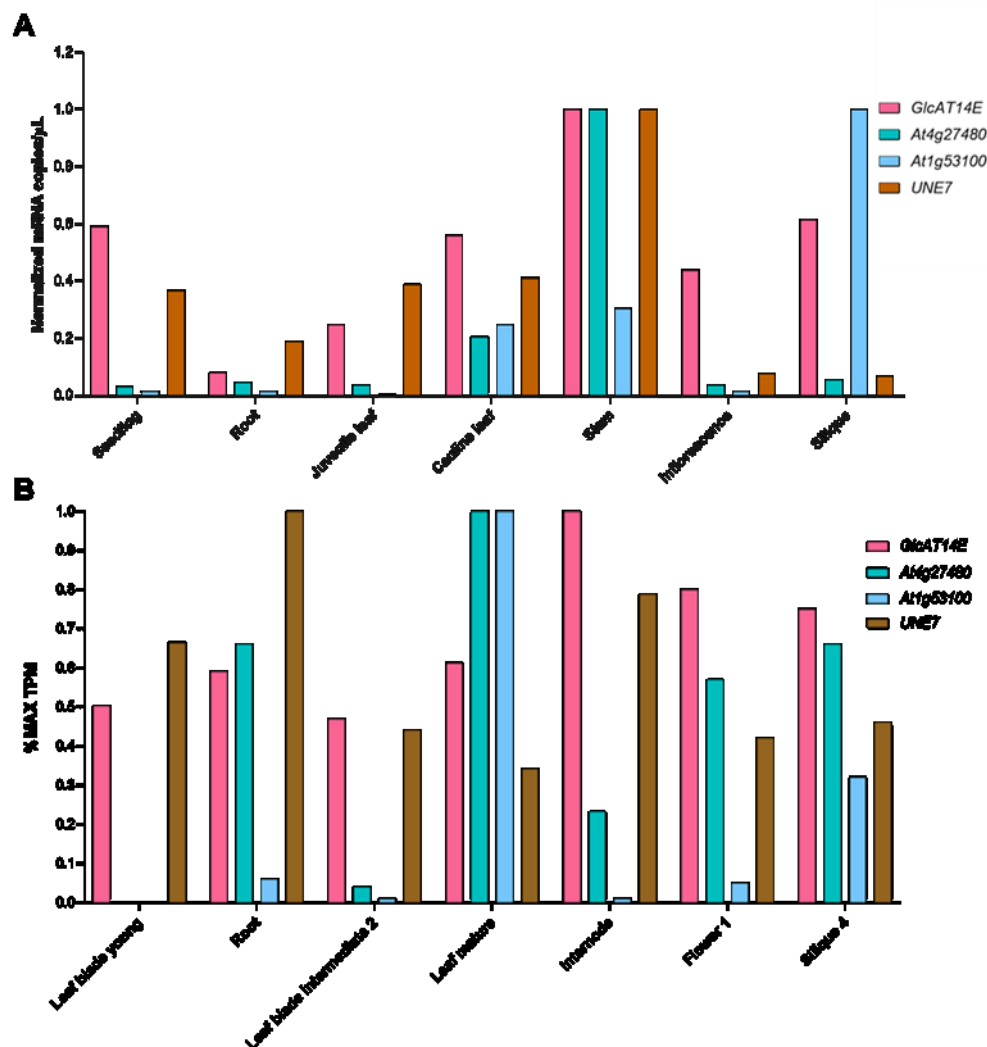


Fig. 9 - Expression pattern of GT14 genes in different tissues in Arabidopsis.

(A) Normalized transcript levels of *GlcAT14E*, *At4g27480*, *At1g53100* and *UNE7* in a range of tissues. Means of three biological replicates are expressed as a ratio of maximal levels observed. Values of 1.0 therefore indicate the tissue in which the gene is transcribed at its maximal rate. (B) Percentage of TPM of *GlcAT14E*, *At4g27480*, *At1g53100* and *UNE7* expressed as a ratio of maximal levels observed for the specific gene based on RNA-seq data (Klepikova et al., 2016).

Furthermore, *GlcAt14C* was transcribed at its maximal rate in seedlings while *GlcAt14D* and *At1g71070* was transcribed at their maximal rate in stems (Fig. 10A). High levels of expression of *GlcAt14C* were found in juvenile leaves and stems whereas *GlcAT14D* was in seedlings and cauline leaves. Some qPCR results were opposed to the RNA-seq data. *GlcAt14C* and *At1g71070* were most transcribed in internodes and roots, respectively (Fig. 10B). High levels of expression were expected for *GlcAt14D* and *At1g71070* in roots. Apparently, the expression pattern of *GlcAT14C* and *At1g71070* had some similarities. The genes most transcribed in flowers were *GlcAT14A*, *GlcAT14B*, *GlcAT14D* and *GlcAT14E*.

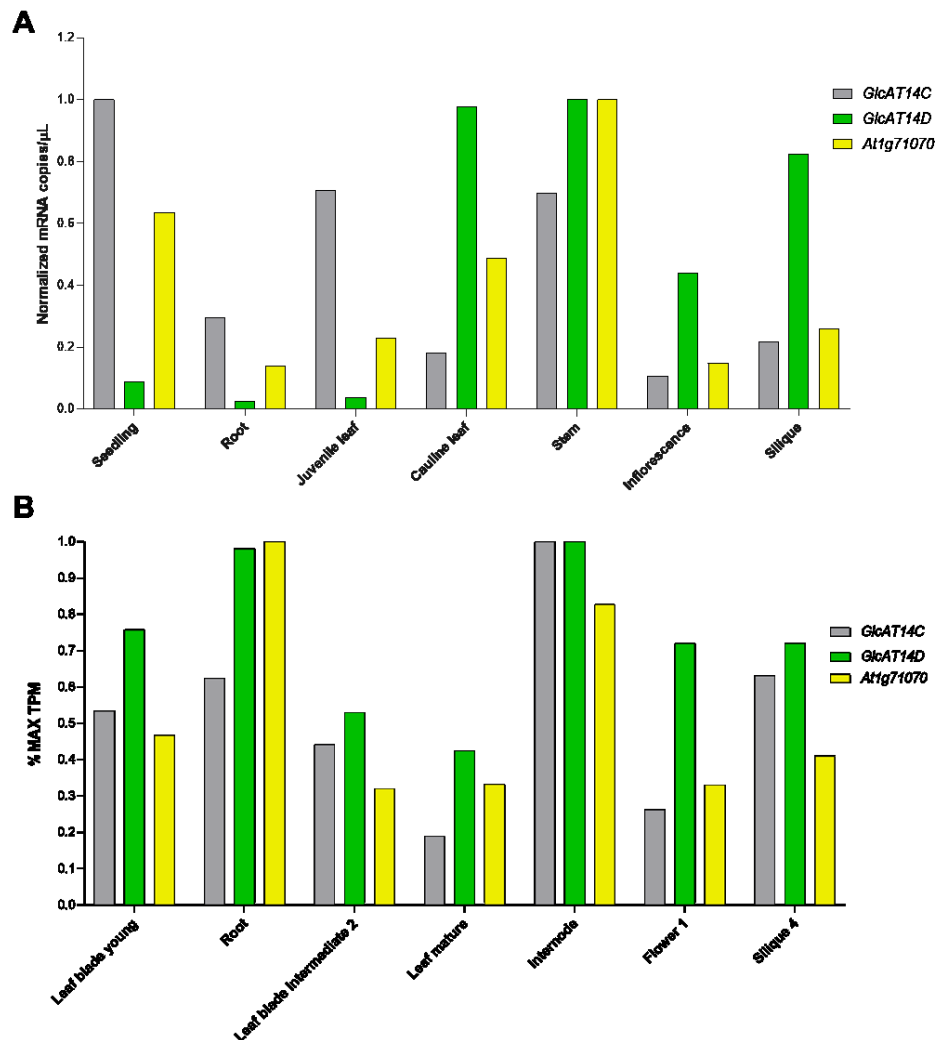


Fig. 10 - Expression pattern of GT14 genes in different tissues in Arabidopsis.

(A) Normalized transcript levels of *GlcAT14C*, *GlcAT14D* and *At1g71070* in a range of tissues. Means of three biological replicates are expressed as a ratio of maximal levels observed. Values of 1.0 therefore indicate the tissue in which the gene is transcribed at its maximal rate. (B) Percentage of TPM of *GlcAT14C*, *GlcAT14D* and *At1g71070* expressed as a ratio of maximal levels observed for the specific gene based on RNA-seq data (Klepikova et al., 2016).

3.3. Plant marker lines

Due to time availability, this part of the study was focused only on six members of the GT14 family in Arabidopsis: *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520*. In order to complement the previous data from gene expression analysis by qPCR, plant marker lines with promoters fused to a NLS and three tandem fluorescent protein (YFP) sequences were created. Promotor fragments of *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* (Fig. 11A) were cloned into pCRTM8/GW/TOPOTM (InvitrogenTM). A restriction digestion analyses was performed to confirm the present and the orientation of the inserts (Fig. 11B). *pCR8:GlcAt14Bprom*, *pCR8:At4g27480prom* and *pCR8:At1g53100prom* were cut with EcoRV, *pCR8:GlcAt14Eprom* and *pCR8:At1g03520prom* were cut with EcoRV and HindIII, and *pCR8:At1g71070prom* was cut with EcoRV and SacI. pCR8 vector without any insert and a negative control without DNA were also cut with the same enzymes. The corresponding digestion fragments had the expected sizes (Table 4). Then, the promotor fragments were subcloned into pMT589 to create prom:YFP constructs (Fig. 12A). The constructs were confirmed by DNA sequencing (Supplementary fig. 13-14) and transformed into Arabidopsis plants by the floral dip method. Due to restriction of time, the plants containing the *At1g71070prom:YFP*, *At1g53100prom:YFP* and *At1g03520prom:YFP* were not yet examined.

In all transgenic plants bearing the different pMT589:YFP constructs, the YFP fluorescent signal was targeted to the nuclei as expected. All the flowers analyzed were between stages 9-10, when cell division and differentiation ends, and stages 12-13, when anthesis occurs according to (Smyth et al., 1990).

YFP expression driven by the *GlcAT14B* promotor was strong in the pistil tissues and in the transmitting tissue but was not detected in the stigma (Fig. 12B-G). The fluorescent signal was found in the stamens in the wall of anthers and in the filaments (Fig. 12H-I), and in the nucellus and chalaza region of ovules from flowers in stages 9-10 (Fig. 12J-K). YFP expression was detected in the funiculus that attaches the ovule to the placenta (Fig. 12L-M) and in the chalazal region of the ovules (Fig. 12L-M) of flowers in stages 12-13. YFP expression was not detected in pollen grains (Fig. 12N-O).

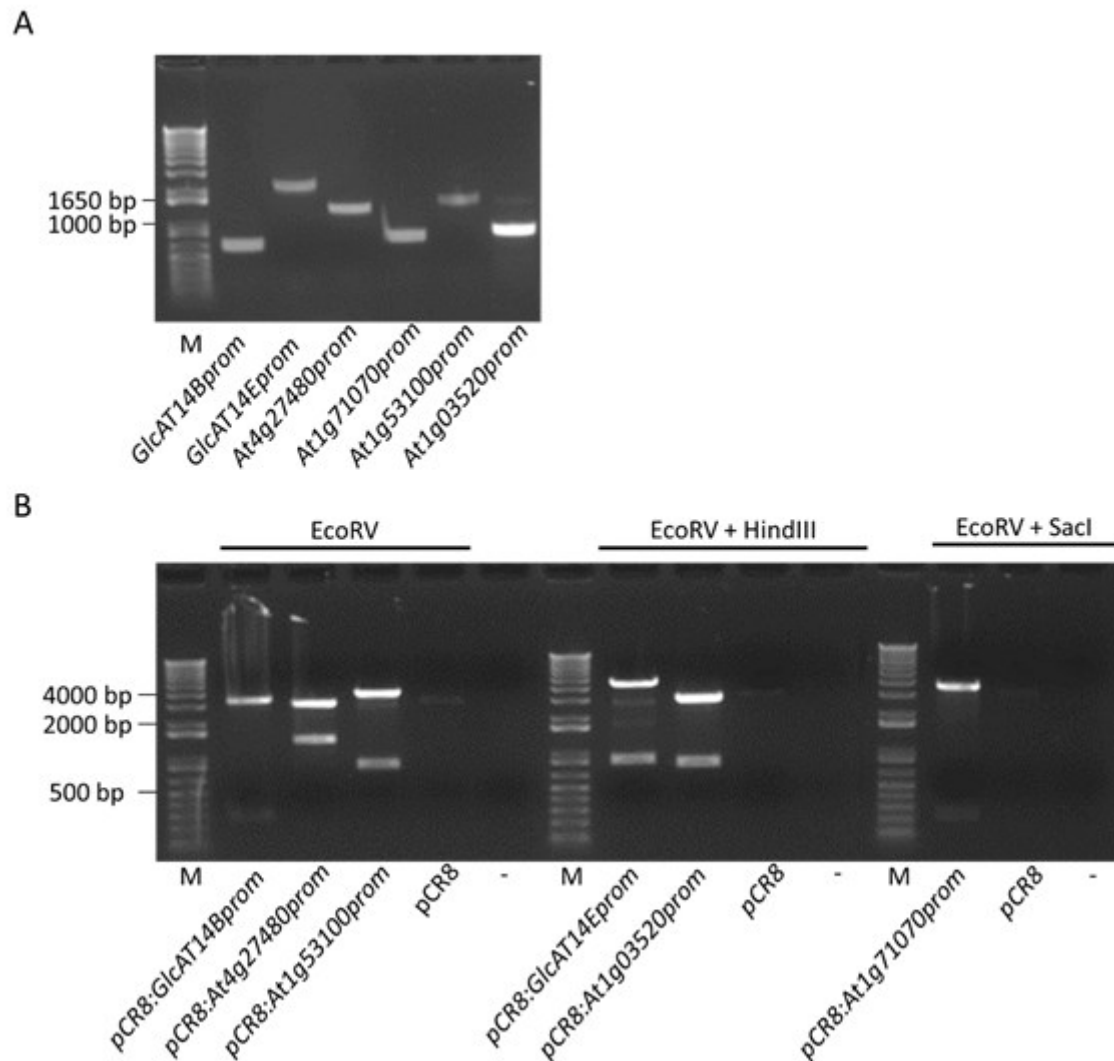


Fig. 11 - Preparation of plant marker lines with YFP gene under regulation of GT14 promoters.

(A) Electrophoretic separation of GT14 promoters fragments amplified by PCR. Promotor fragments of *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* were amplified with the primer pairs described in Supplementary table 4. (B) Electrophoretic separation of DNA fragments from restriction digestion analysis performed on plasmid DNA of pCR8 with GT14 promoters. *pCR8:GlcAT14Bprom*, *pCR8:At4g27480prom* and *pCR8:At1g53100prom* were cut with EcoRV, *pCR8:GlcAT14Eprom* *pCR8:At1g03520prom* were cut with EcoRV and HindIII, and *pCR8:At1g71070prom* was cut with EcoRV and SacI. The resulting DNA fragments had the expected sizes (Table 4). Abbreviations: M, DNA ladder; pCR8, vector without insert; -, negative control without DNA.

GlcAT14E promotor guided the expression of YFP strongly to rosette (Fig. 12A-B) and caulinar leaves (Fig. 12C-D), sepals (Fig. 12E-F) and petals (Fig. 12G-H). The fluorescent signal was present in the wall of anthers but not in pollen grains from flowers at stages 9-10 (Fig. 12I-J). YFP was found to be expressed in the pistil tissues and in the transmitting tissue (Fig. 12M-N), in the funiculus (Fig. 12K-L) and in the chalazal pole of the ovules (Fig. 12K-L) of flowers at stages 12-13.

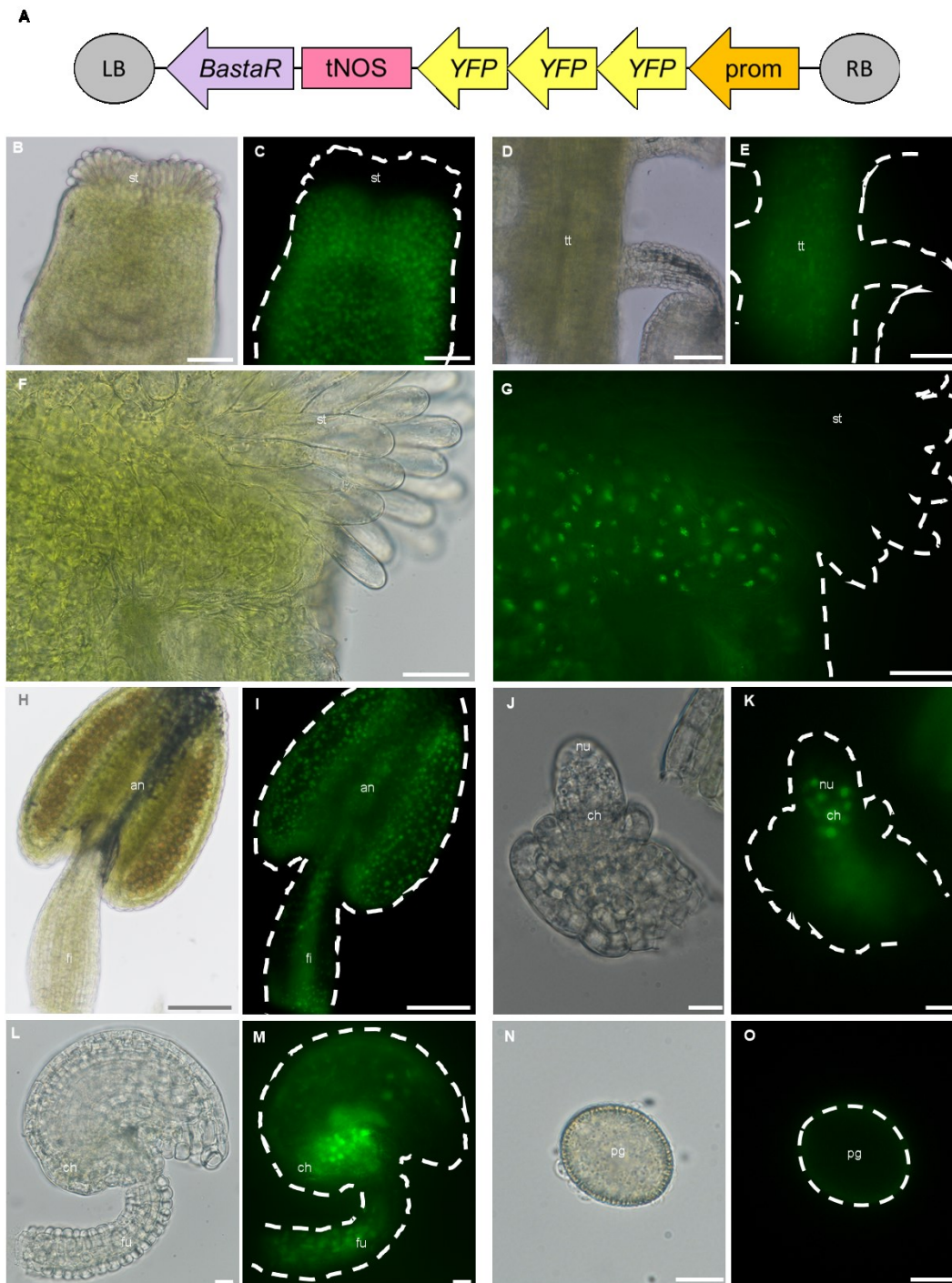


Fig. 12 - *GlcAt14Bprom:YFP* expression pattern in Arabidopsis.

(A) Expression cassette showing the relative position of the GT14 promoter (prom), the three tandem YFP sequences, the terminator Nos (tNOS) and the resistance to phosphinothricin acetyltransferase gene (BastaR). (B, D, F, H, J, L and N) Bright-field microscopy images and (C, E, G, I, K, M and O) fluorescence microscopy images. YFP expression was detected in young (B-C) and mature pistils (D-G), in anthers (H-I), in the young (J-K) and mature ovules (L-M). YFP expression was not detected in pollen grains (N-O). All the flowers used in these observations were at stages 9-10 or 12-13 according to Smyth et al. (1990). Abbreviations: an, anther; ch, chalaza; fu, funiculus; fi, filament; nu, nucellus; pg, pollen grain; st, stigma; tt, transmitting tissue. Scale bars: 20 µm (J-O); 50 µm (D-G); 100 µm (B-C, H-I).

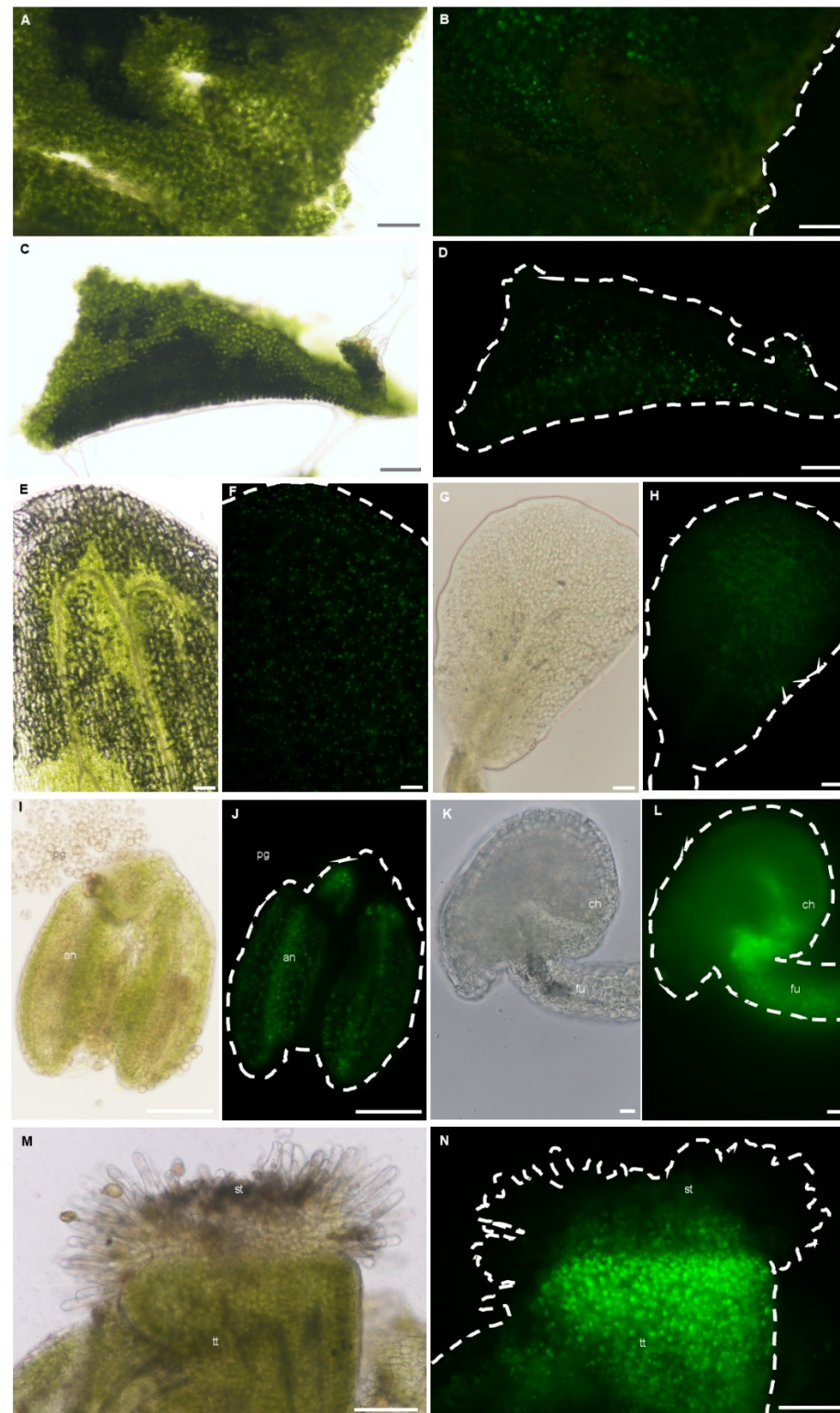


Fig. 13 - *GlcAt14Eprom:YFP* expression pattern in Arabidopsis.

(A, C, E, G, I, K and M) Bright-field microscopy images and (B, D, F, H, J, L and N) fluorescence microscopy images. YFP expression was detected in rosette (A-B) and caulinar leaves (C-D), in sepals (E-F), in petals (G-H), in anthers (I-J), in mature ovules (K-L) and in mature pistils (M-N). All the flowers used in these observations were at stages 9-10 or 12-13 according to Smyth et al. (1990). Abbreviations: an, anther; ch, chalaza; fu, funiculus; pg, pollen grain; st, stigma; tt, transmitting tissue. Scale bars: 20 μ m (K-L); 100 μ m (A-J, M-N).

Plants transgenic for the *At4g27480prom:YFP* expression cassette exhibited YFP expression in petals (Fig. 14A-B), in the wall of anthers from flowers at stages 9-10 (Fig. 14C-D) and in filaments of flowers at stages 12-13 (Fig. 14E-F). YFP expression was not found in pollen grains (Fig. 14G-H).

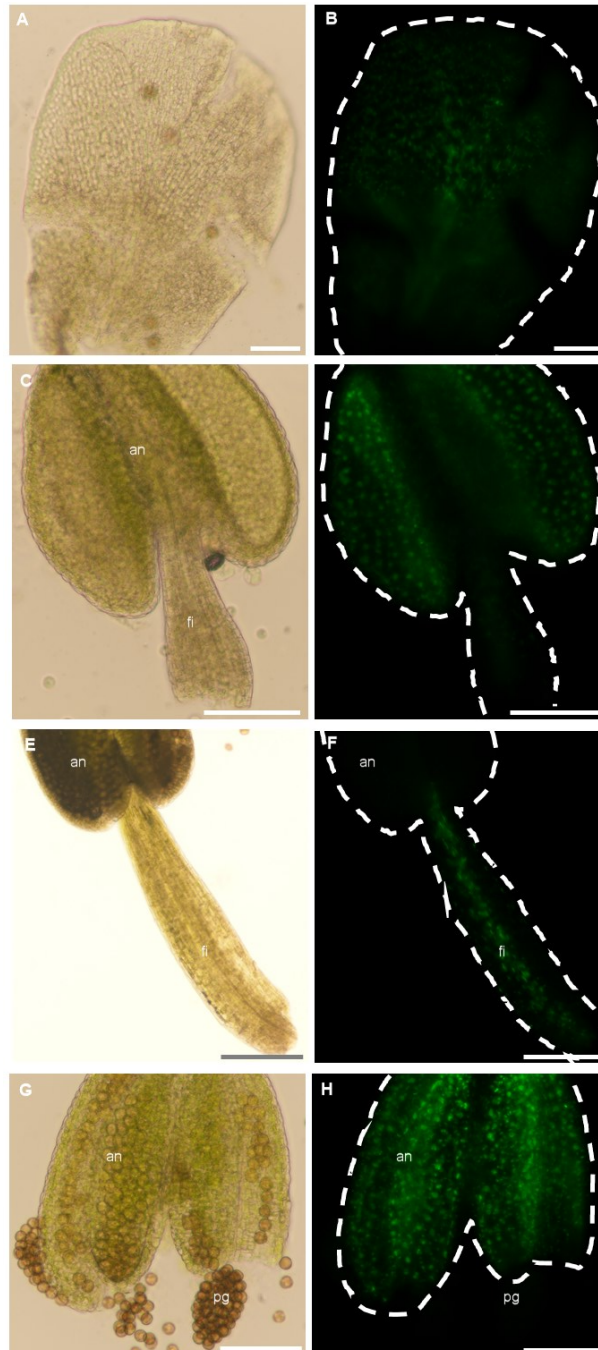


Fig. 14 - *At4g27480prom:YFP* expression pattern in Arabidopsis.

(A, C, and E) Bright-field microscopy images and (B, D and F) fluorescence microscopy images. YFP expression was detected in petals (A-B), in young (C-D) and mature anthers (E-F). YFP expression was not detected in pollen grains (G-H). All the flowers used in these observations were at stages 9-10 or 12-13 according to Smyth et al. (1990). Abbreviations: an, anther; fi, filament. Scale bars: 100 μ m (A-H).

3.4. Subcellular localization

The 11 sequence proteins of GT14 family were run in several subcellular localization prediction programs: TargetP 1.1 (Emanuelsson et al., 2000), ARAMEMNON (Schwacke et al., 2003) and SUBA (Hooper et al., 2017). GlcAT14D, GlcAT14E, At4g27480, At1g71070 and UNE7 seem to be localized to the Golgi (Table 7). GlcAT14B, At1g03520 and At4g03340 were probably targeted to the mitochondrion while GlcAT14C was localized in the chloroplast. For GlcAT14A and At1g53100, the three servers predicted different subcellular locations.

Table 7 - Predicted subcellular localization of GT14 proteins in Arabidopsis.

Subcellular localization was predicted with 3 different programs: TargetP 1.1 (Emanuelsson et al., 2000), ARAMEMNON (Schwacke et al., 2003) and SUBA (Hooper et al., 2017). Abbreviations: C, chloroplast; Ex, extracellular; G, Golgi; M, mitochondrion; PM, plasma membrane; S, secretory pathway; -, any other location.

Protein	Predicted Subcellular Localization		
	TargetP 1.1	ARAMEMNON	SUBA
GlcAT14A	-	M	G
GlcAT14B	M	M	G
GlcAT14C	C	C	M
GlcAT14D	S	S	G
GlcAT14E	S	S	Ex
At4g27480	S	S	PM
At1g71070	S	S	G
At1g53100	-	S	PM
At1g03520	M	M	M
UNE7	S	S	Ex
At4g03340	M	M	M

To confirm the subcellular localization previsions, the CDS of the six GT14 genes were fused to GFP and transiently expressed in *N. benthamiana* leaves. The amplified CDS of *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* (Fig. 15A) were cloned into pCR™8/GW/TOPO™ (Invitrogen™). To check the presence and the orientation of inserts, the constructs were digested with restriction enzymes (Fig. 15B). In this analysis, pCR8 vector without any insert and a negative control without DNA were included as controls. *pCR8:At1g71070* and *pCR8:At1g03520* were cut with EcoRV, *pCR8:GlcAT14B* was cut with EcoRV and BamHI, *pCR8:GlcAT14E* was cut with EcoRV and SacI, and *pCR8:At4g27480* and *pCR8:At1g53100* were cut with EcoRV and Scal. The expected sizes (Table 4) were consistent with the digestion fragments (Fig. 15B). Afterward, the CDS fragments were subcloned into pMDC83 (Curtis and Grossniklaus, 2003) to create *CDS:GFP* constructs (Fig. 16A).

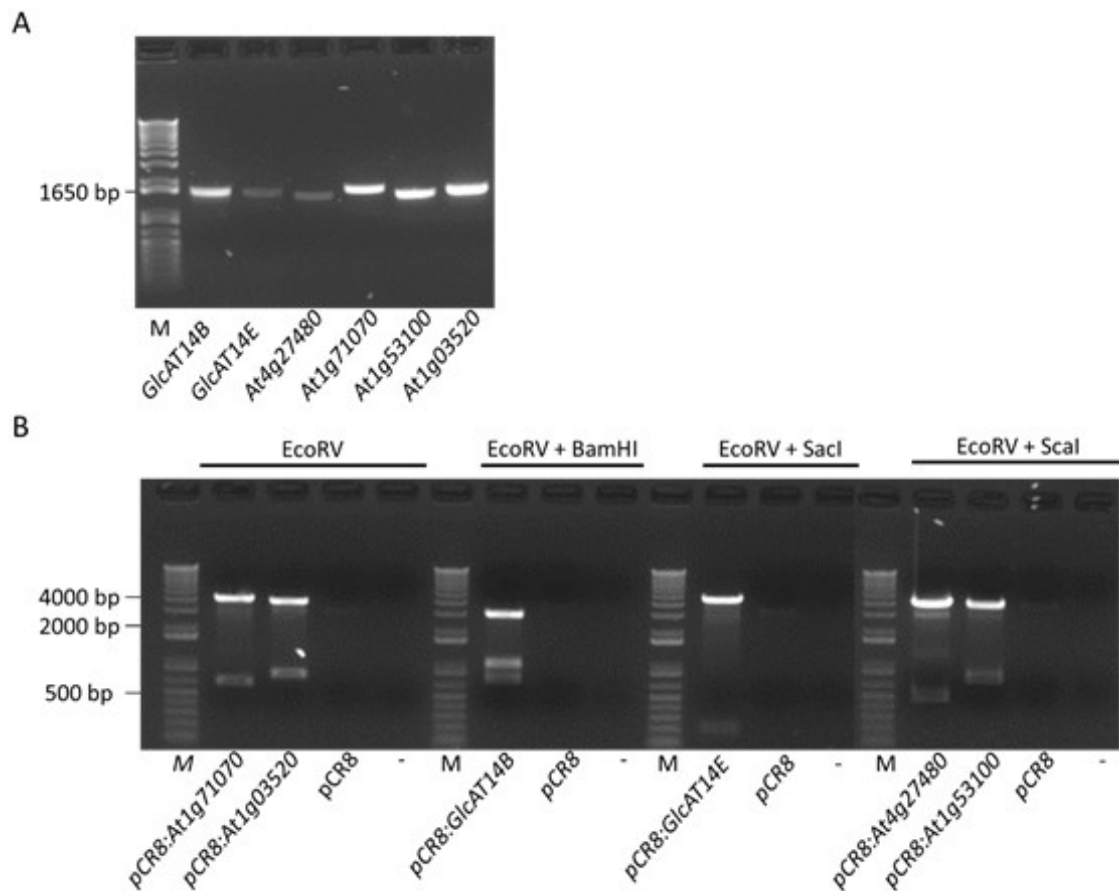


Fig. 15 - Preparation of plant marker lines with GT14 genes and GFP gene under regulation of 35S promoter. (A) Electrophoretic separation of GT14 CDS fragments amplified by PCR. CDS fragments of *GlcAT14B*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100* and *At1g03520* were amplified with the primer pairs described in Supplementary table 5. (B) Electrophoretic separation of DNA fragments from restriction digestion analysis performed on plasmid DNA of pCR8 with CDS of GT14. *pCR8:At1g71070* and *pCR8:At1g03520* were cut with EcoRV, *pCR8:GlcAT14B* was cut with EcoRV and BamHI, *pCR8:GlcAT14E* was cut with EcoRV and SacI, and *pCR8:At4g27480* and *pCR8:At1g53100* were cut with EcoRV and Scal. The resulting DNA fragments had the expected sizes (Table 4). Abbreviations: M, DNA ladder; pCR8, vector without insert; -, negative control without DNA.

The constructs were confirmed by DNA sequencing (Supplementary fig. 15-16) and then transiently expressed in *N. benthamiana*. Transient expression of *GlcAT14B:GFP* (Fig. 16B), *GlcAT14E:GFP* (Fig. 16C), *At4g27480:GFP* (Fig. 16D) and *At1g53100:GFP* (Fig. 16F) after 3 days of infiltration showed a typical pattern of Golgi localization with discrete punctate structures as shown in the Golgi marker fused to mCherry (Fig. 16I). *At1g71070:GFP* (Fig. 16E) and *At1g03520:GFP* (Fig. 16G) a very weak punctate signal that may be characteristic of Golgi localization. Due to time restrictions, there was no possibility of repeating the experiment for these two constructs. No reticulate structures typical of ER localization (Fig. 16J) or the plasma membrane marker were observed (Fig. 16K).

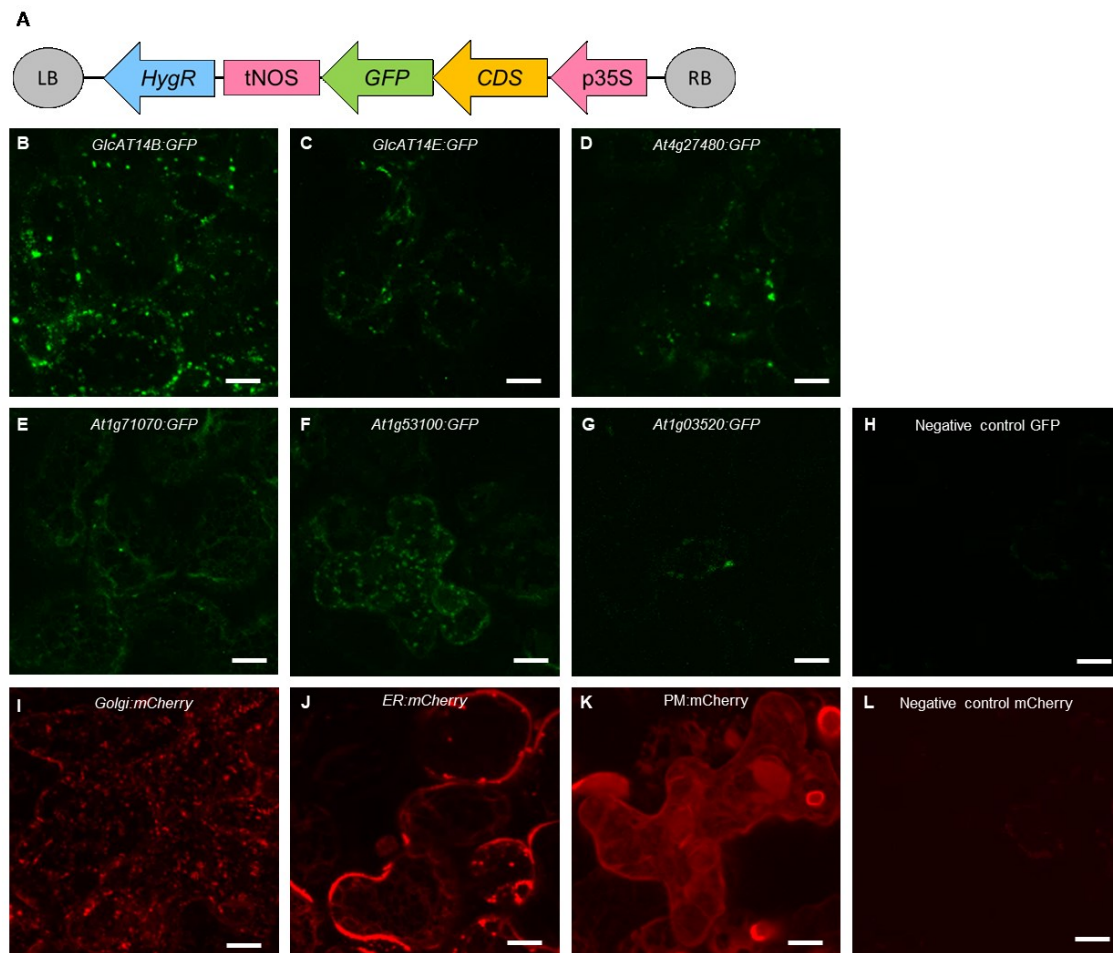


Fig. 16 - Subcellular localization of six GT14 proteins in *N. benthamiana* leaf epidermal cells.

(A) Expression cassette showing the relative position of the 35S promoter (p35S), the GT14 CDS, the GFP gene, the terminator Nos (tNOS) and the resistance to hygromycin gene (HygR). Transiently expressed *GlcAT14B:GFP* (B), *GlcAT14E:GFP* (C), *At4g27480:GFP* (D), *At1g71070:GFP* (E), *At1g53100:GFP* (F) and *At1g03520:GFP* (G) localized in the Golgi apparatus (I) after 3 days of infiltration. A Golgi marker (I), an ER marker (J) and a plasma membrane (PM) marker (K) fused to mCherry were utilized. (H) and (L) correspond to negative controls (leaves not infiltrated) of GFP and mCherry, respectively. These constructs were examined by laser-scanning confocal microscopy under fluorescent light. Size bar = 10 μ m.

3.5. Functional analysis of glucuronosyltransferase genes

To investigate a role for GT14 family *in vivo*, a reverse genetic approach was adopted, and several mutants of Arabidopsis were analyzed. The identification of phenotypes in mutants of GTs has proven to be a hard task in single mutants due to the expected redundancy in the activity of GTs. Thus, creating high order mutants may be essential to see obvious phenotypes. The previously reported insertion mutants *glcat14b-1*, *glcat14e-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1* and *glcat14abd-2* (Lopez-Hernandez, 2018) and *glcat14a±be* were kindly provided by prof. Paul Dupree (University of Cambridge, England). To identify a knock-out mutant for *At4g27480*, a T-DNA insertion

line was selected using the SIGnAL database and was obtained from NASC. Due to time restrictions, the mutants *glcat14a-2*, *glcat14d-1* and *glcat14d-2* were not analyzed although these lines were used to create the double and triple mutants at Dupree's laboratory.

The homozygous T-DNA insertion line *glcat14a-2* (SALK_043905) was described in Knoch et al. (2013). The line *glcat14a-2* has a T-DNA insertion in the 2th exon of *GlcAT14A* (Fig. 17A). Homozygous mutants were identified by PCR (Fig. 17B) with primers listed in Supplementary table 1 and confirmed by sequencing (Supplementary figure 17). qPCR analysis indicates a lower transcript abundance in *glcat14abd-1*, *glcat14abd-2* and *glcat14ab* in comparison to wt (Fig. 17C; Supplementary figure 18). *glcat14a±be* has a similar expression to wt as it is heterozygous for *GlcAt14A*.

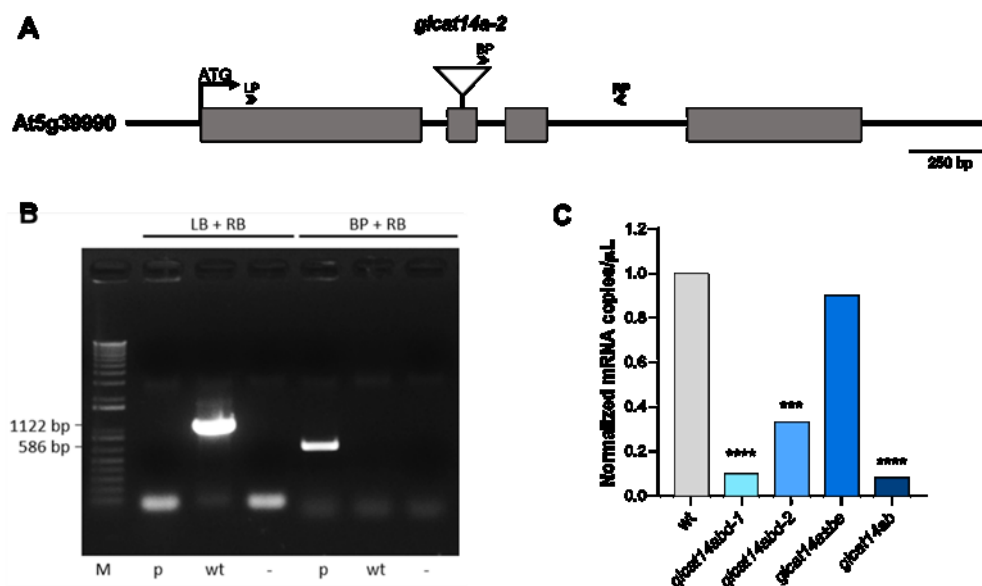


Fig. 17 - Analysis of *GlcAT14A* T-DNA insertion lines.

(A) Schematic gene model structure of *GlcAT14A* (At5g39990) and T-DNA insertion site in *glcat14a-2* (SALK_043905). The grey box represents the exons and the black line represents the untranslated regions and introns. *glcat14a-2* insertion site is indicated by an inverted triangle. The arrows indicate the genotyping primers (LP - left genomic primer; RP - right genomic primer; BP - T-DNA border primer). (B) Genotyping electrophoresis gel of a *glcat14a-2* and a wt plant. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (C) qPCR analysis to detect the presence of *GlcAT14A* transcript in the inflorescences of wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be* and *glcat14ab* plants. Means of three biological replicates are expressed as a ratio of maximal levels observed in wt. Values of 1.0 therefore indicate the wt transcription of *GlcAT14A*. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

Lopez-Hernandez (2018) described a mutant line for *GlcAT14B*, *glcat14b-1* (SALK_080923), with an insertion in the 4th coding exon of *GlcAT14B* (Fig. 18A). Homozygous lines were identified by PCR analysis (Fig. 18B) with primers enumerated in Supplementary table 1 and T-DNA insertion sites were confirmed by DNA sequencing

(Supplementary figure 17). When comparing to wt, a lower amount of *GlcAT14B* transcripts were found in the mutants *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1* and *glcat14b-1* (Fig. 18C; Supplementary figure 18).

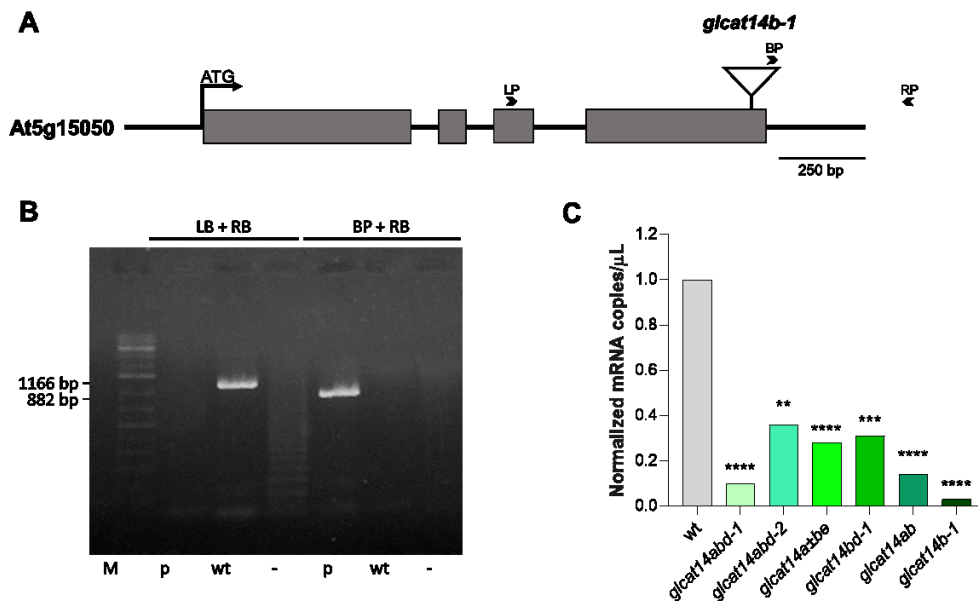


Fig. 18 - Analysis of *GlcAT14B* T-DNA insertion lines.

(A) Schematic gene model structure of *GlcAT14B* (At5g15050) and T-DNA insertion site in *glcat14b-1* (SALK_080923). The grey box represents the exons and the black line represents the untranslated regions and introns. *glcat14a-2* insertion site is indicated by an inverted triangle. The arrows indicate the genotyping primers (LP – left genomic primer; RP – right genomic primer; BP – T-DNA border primer). (B) Genotyping electrophoresis gel of a *glcat14b-1* and a wt plant. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (C) qPCR analysis to detect the transcript abundance of *GlcAT14B* in the inflorescences of wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1* and *glcat14b-1* plants. Means of three biological replicates are expressed as a ratio of maximal levels observed in wt. Values of 1.0 therefore indicate the wt transcription of *GlcAT14B*. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

The lines *glcat14d-1* (GK363F05) and *glcat14d-2* (GK508D01) described by (Lopez-Hernandez, 2018). Both lines have a T-DNA insertion in the 3rd intron of *GlcAT14D* (Fig. 19A). Homozygous lines were confirmed by PCR (Fig. 19B-C) with primers registered in Supplementary table 1 and confirmed by sequencing (Supplementary table 19). qPCR results showed lower transcript abundance in *glcat14abd-1*, *glcat14abd-2* and *glcat14bd-1* in comparison to wt (Fig. 19D; Supplementary figure 18).

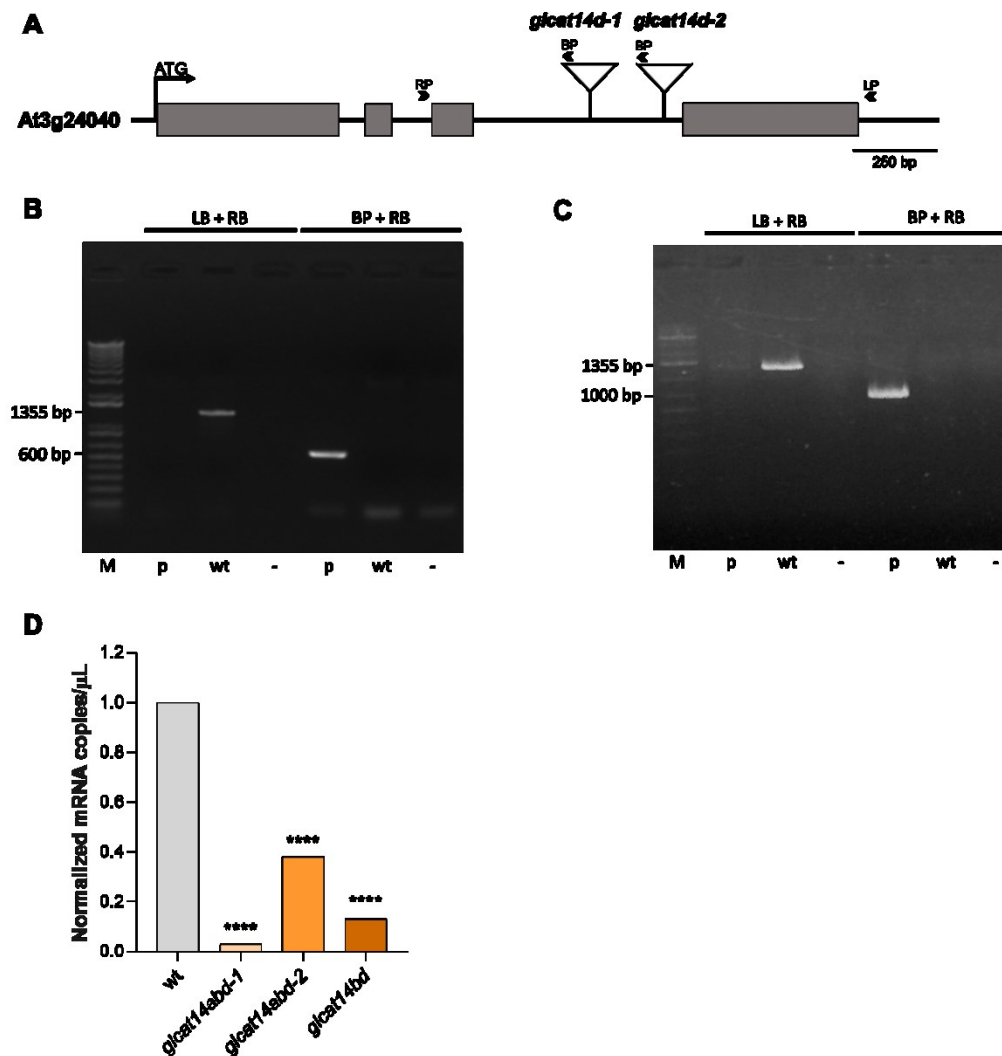


Fig. 19 - Analysis of *GlcAT14D* T-DNA insertion lines.

(A) Schematic gene model structure of *GlcAT14D* (At3g24040) and T-DNA insertion sites in *glcat14d-1* (GK363F05) and *glcat14d-2* (GK508D01). The grey box represents the exons and the black line represents the untranslated regions and introns. *glcat14a-2* insertion site is indicated by an inverted triangle. The arrows indicate the genotyping primers (LP – left genomic primer; RP – right genomic primer; BP – T-DNA border primer). (B) Genotyping electrophoresis gel of a *glcat14d-1* and a wt plant. (C) Genotyping electrophoresis gel of an *glcat14d-2* and a wt plant. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (C) qPCR analysis to detect the transcript abundance of *GlcAT14D* in the inflorescences of wt, *glcat14abd-1*, *glcat14abd-2* and *gcat14bd-1* plants. Means of three biological replicates are expressed as a ratio of maximal levels observed in wt. Values of 1.0 therefore indicate the wt transcription of *GlcAT14D*. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

Another line was described by Lopez-Hernandez (2018) for *GlcAT14E*: *glcat1e-1* (SALK_022820). The insertion is situated in the 2nd intron of *GlcAT14E* (Fig. 20A). PCR analysis with primers enumerated in Supplementary table 1 identified homozygous lines (Fig. 20B) and DNA sequencing confirmed the T-DNA insertion positions (Supplementary table 20). When comparing to wt, a lower amount of *GlcAT14E*

transcripts were found in the mutants *glcat14a±be* and *glcat14e-1* (Fig. 20C; Supplementary figure 18).

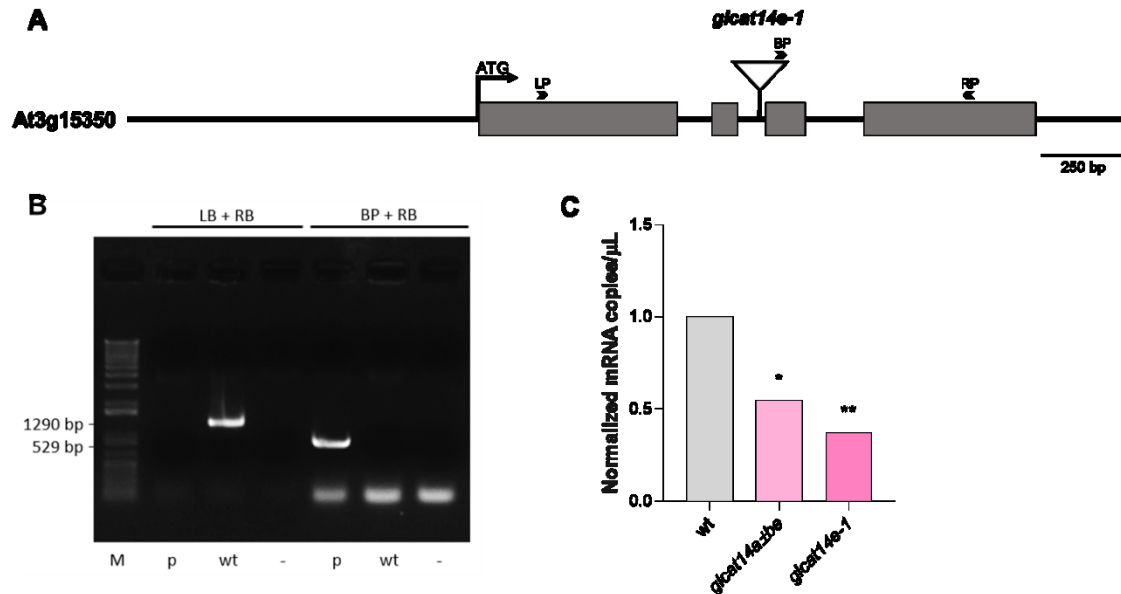


Fig. 20 - Analysis of *GlcAT14E* T-DNA insertion lines.

(A) Schematic gene model structure of *GlcAT14E* (At3g15350) and T-DNA insertion site in *glcat14e-1* (SALK_022820). The grey box represents the exons and the black line represents the untranslated regions and introns. *glcat14a-2* insertion site is indicated by an inverted triangle. The arrows indicate the genotyping primers (LP – left genomic primer; RP – right genomic primer; BP – T-DNA border primer). (B) Genotyping electrophoresis gel of a *glcat14e-1* and a wt plant. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (C) qPCR analysis to detect the transcript abundance of *GlcAT14E* in the inflorescences of wt, *glcat14a±be* and *atgcat14e-1* plants. Means of three biological replicates are expressed as a ratio of maximal levels observed in wt. Values of 1.0 therefore indicate the wt transcription of *GlcAT14E*. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

A T-DNA insertion line (SALK_105434C) for *At4g27480* was characterized. SALK_105434C contains a T-DNA insertion in the 1st intron of *At4g27480* (Fig. 21A). The homozygous line was isolated and named as *at4g27480-1*. Copies of the inserts were amplified from the genomic DNA (Fig. 21B) with primers listed in Supplementary table 1 and DNA sequencing confirmed the T-DNA insertion positions (Supplementary figure 20). qPCR detected only a residual level of *AT4G27480* expression in *at4g27480-1* (Fig. 21C; Supplementary figure 18).

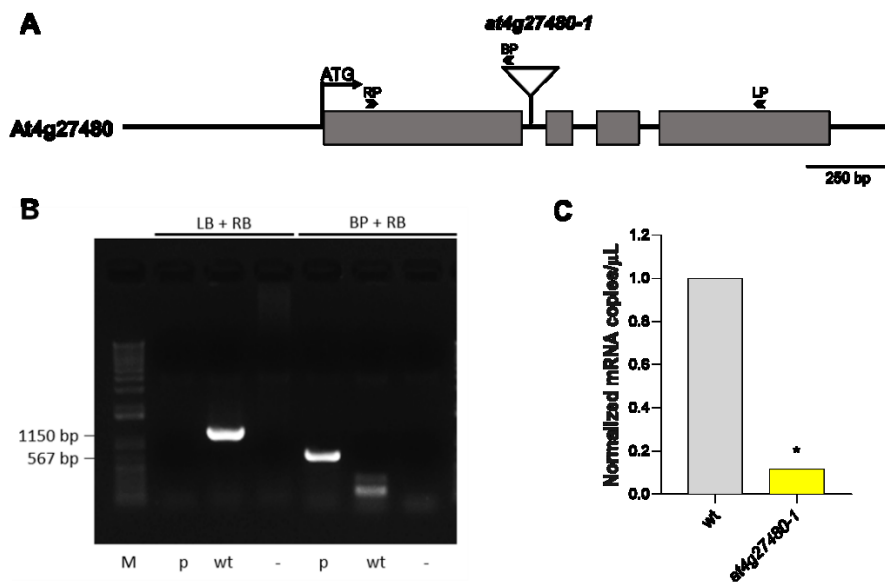


Fig. 21 - Analysis of *At4g27480* T-DNA insertion lines.

(A) Schematic gene model structure of *At4g27480* and T-DNA insertion site in *at4g27480-1* (SALK_105434C). The grey box represents the exons and the black line represents the untranslated regions and introns. *glcat14a-2* insertion site is indicated by an inverted triangle. The arrows indicate the genotyping primers (LP – left genomic primer; RP – right genomic primer; BP – T-DNA border primer). (B) Genotyping electrophoresis gel of an *at4g27480-1* and a wt plant. Abbreviations: M, DNA ladder; p, mutant plant; wt, wild-type; -, negative control without DNA. (C) qPCR analysis to detect the transcript abundance of *GlcAT14E* in the inflorescences of wt and *at4g27480-1* plants. Means of three biological replicates are expressed as a ratio of maximal levels observed in wt. Values of 1.0 therefore indicate the wt transcription of *At4g27480*. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

Here it was confirmed that *glcat14b-1*, *glcat14e-1*, *at4g27480-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2* and *glcat14a±be* are knockdown mutants.

The mutants were grown in normal conditions in soil and no visible phenotype was visible (data not shown). Some growth phenotypes of mutants may be conditional requiring specific conditions of growth such as high concentrations of sodium chloride (NaCl) (Liang et al., 2013; Tryfona et al., 2014) in the growth medium or temporal treatments with chemicals such as isoxaben (Showalter and Basu, 2016a). Recent evidence show that [Me]GlcA of AGPs can bind to Ca^{2+} (Lamport and Várnai, 2013). Lopez-Hernandez (2018) tested the sensitivity of the phenotypes to Ca^{2+} . The AG-polysaccharides of AGPs of the triple mutant *glcat14abd-1* bind less Ca^{2+} than wt AGs suggesting that this mutant is deficient in Ca^{2+} binding in the apoplast. The mutants were grown in plant growth medium with differing concentrations of Ca^{2+} and some growth Ca^{2+} -sensitive phenotypes were found as described previously (Lopez-Hernandez, 2018).

Therefore, in this work the same hydroponic system was used to find phenotypes in the mutants related to sexual reproduction (Lopez-Hernandez, 2018). Wt, *glcat14b-1*,

glcat14e-1, *at4g27480-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2* and *glcat14a±be* were grown on solidified MS medium for 15 days. Seedlings were then transferred to pots containing a thin layer of vermiculite and a cube of rockwool covered with aluminum foil to avoid evaporation and contamination with soil. Every five days the pots were individually watered with nutrient solutions, LCNS and BNS, containing 0.2 mM or 2 mM of Ca²⁺, respectively (Table 3). Fig. 22 shows five-week old plants grown using LCNS (Fig. 22 A) and BNS (Fig. 22 B). The triple mutants (*glcat14abd-1*, *glcat14abd-2* and *glcat14a±be*) grown in LCNS (Fig. 22A) were dwarfed relative to wt. No noteworthy differences were observed in *glcat14b-1*, *glcat14e-1*, *at4g27480-1*, *glcat14ab* and *glcat14bd-1* relative to wt using LCNS (Fig. 22 A) or BNS (Fig. 22 B). Some growth deficiencies were also noticed in *glcat14abd-1* relatively to wt with BNS (Fig. 22 B).

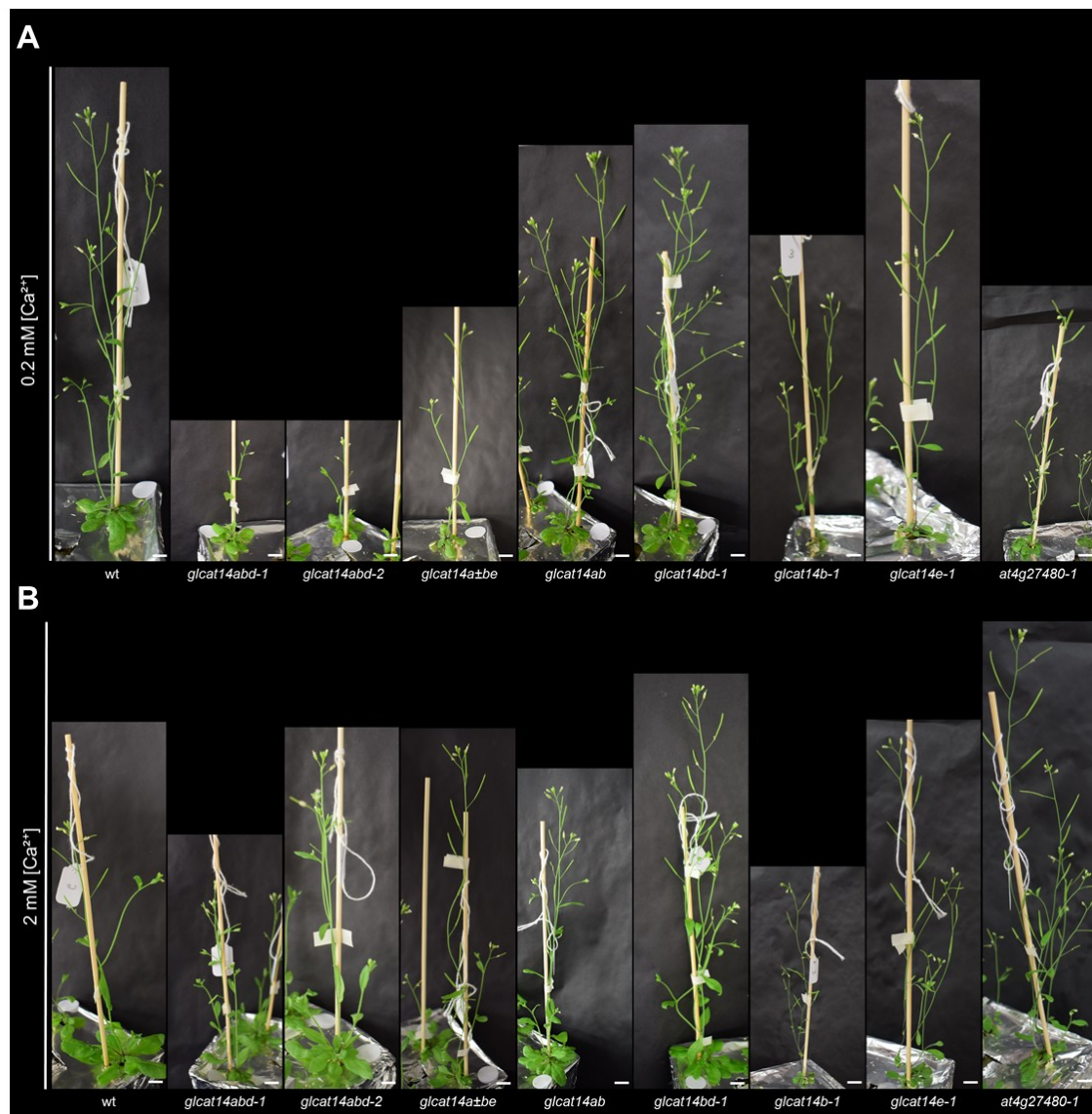


Fig. 22 - Five-week old Arabidopsis wt and mutant plants grown using hydroponic solutions.

Wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* plants grown with LCNS (A) and BNS (B). Size bar = 0.5 cm.

Lopez-Hernandez (2018) reported that the siliques of *glcat14abd-1* had reduced size and presented some aborted or non-fertilized ovules. Therefore, the siliques of *glcat14b-1*, *glcat14e-1*, *at4g27480-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be* and wt plants were examined. The 10th oldest siliques in the stems of each mutant were collected and fixed in ethanol. Seed set was investigated and large areas with no seeds likely corresponding to aborted ovules and seeds were detected in the triple mutants *glcat14abd-1* and *glcat14abd-2* grown in LCNS (Fig. 23A). Some spots with no seeds were also noticed in *glcat14a±be* siliques grown in LCNS (Fig. 23A) and in *glcat14abd-1* and *glcat14abd-2* grown in BNS (Fig. 23B). No defects were seen in all other siliques.

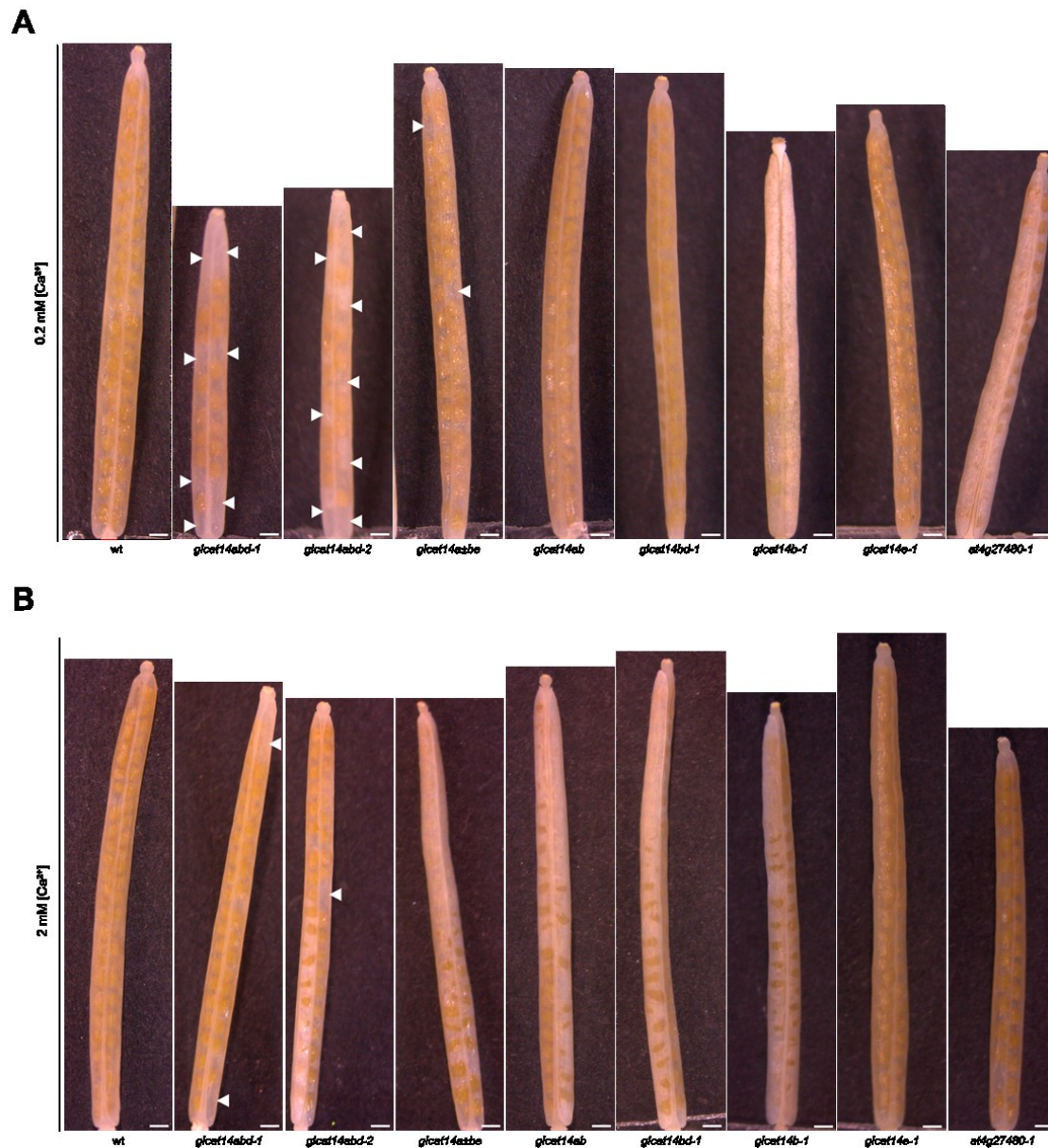


Fig. 23 - Seed set analysis of Arabidopsis wt and mutant siliques in plants grown using hydroponic solutions. Wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* ethanol-fixed siliques of plants grown with LCNS (A) and BNS (B). White arrows indicate areas with no seeds. Size bar = 0.5 mm.

The siliques from the mutant and wt plants were measured (Fig. 24). The silique lengths of *glcat14e-1*, *glcat14ab*, *glcat14bd-1* and *glcat14a±be* did not vary significantly with wt when grown in LCNS (Fig. 24A) and BNS (Fig. 24B). However, mature siliques of *glcat14abd-1*, *glcat14abd-2*, *glcat14b-1* and *at4g27480-1* plants were shorter than wt plants when grown in LCNS (Fig. 23A and Fig. 24A).

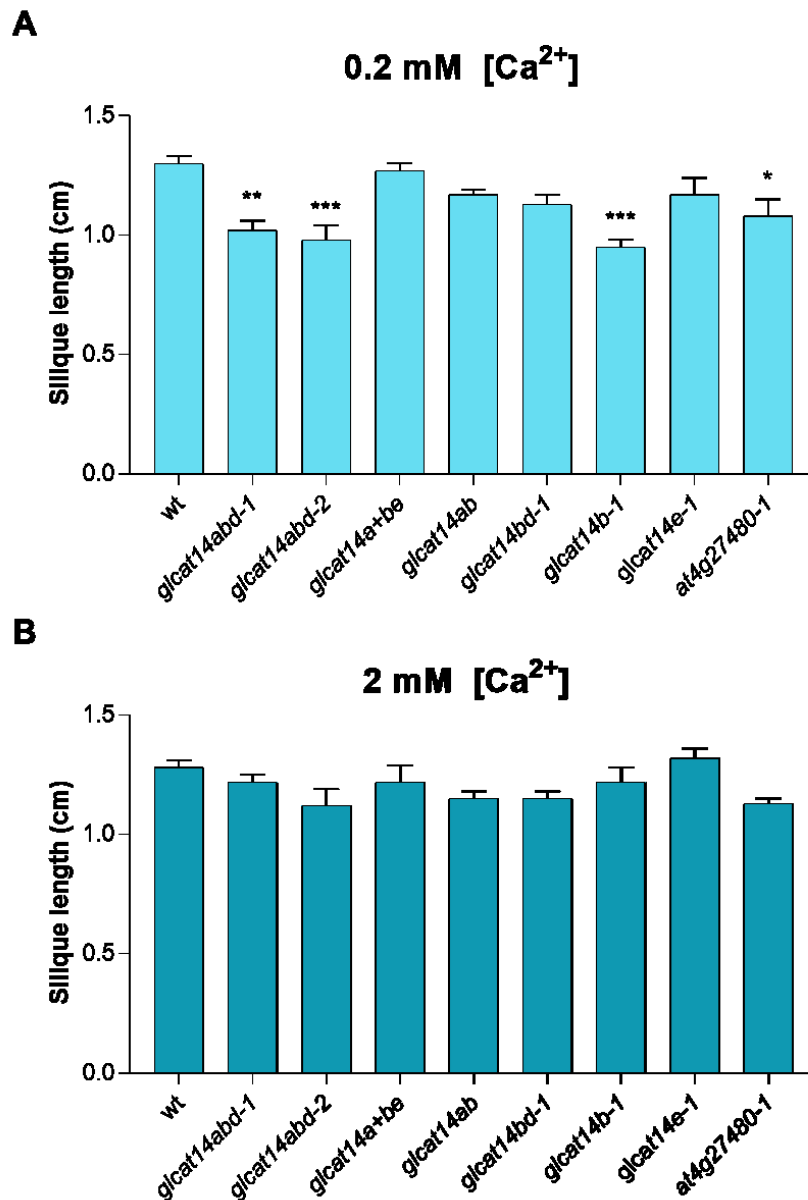


Fig. 24 - Silique length of six-week old *Arabidopsis* mutant and wt plants grown using hydroponic solutions.

Silique lengths of wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a+be*, *glcat14ab*, *glcat14bd-1*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* plants grown with LCNS (A) and BNS (B). Means $\pm$ SEM of three biological replicates are shown. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

The number of viable green seeds, aborted ovules, aborted seeds and white seeds inside the siliques was analyzed, revealing significant lower number of green mature seeds in *glcat14abd-1* and *glcat14abd-2* siliques from wt siliques grown in LCNS (Fig. 25A). The mutant *glcat14ab* grown in BNS had significant lower number of green seeds and aborted ovules compared to wt (Fig. 25B).

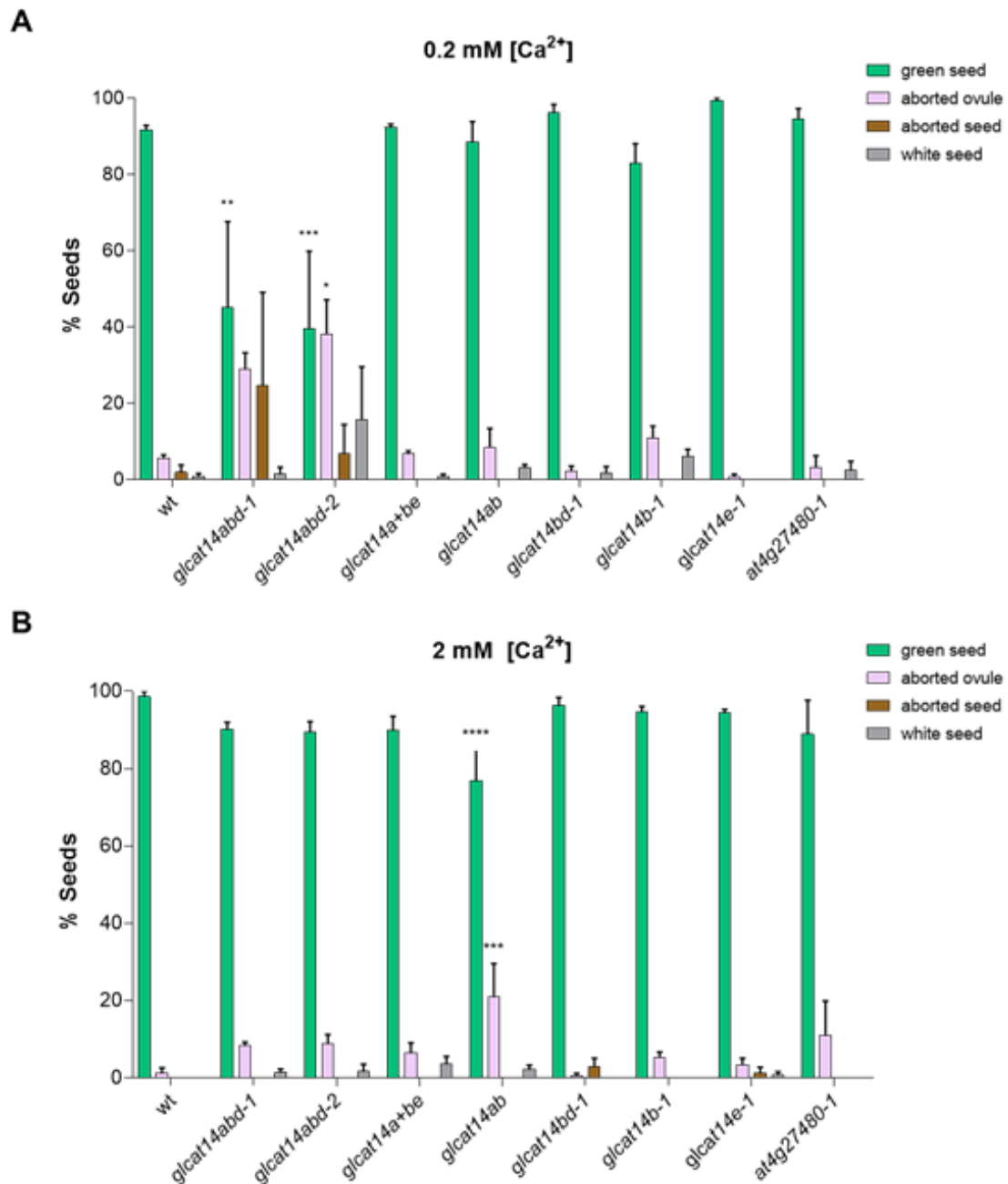


Fig. 25 - Percentage of seeds per silique of six-week old Arabidopsis mutant and wt plants grown using hydroponic solutions.

Percentage of seed per silique of wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* plants grown with LCNS (A) and BNS (B). Means ± SEM of three biological replicates are shown. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

To understand the reason of the presence of unfertilized ovules, the embryo sac of the mutants was cleared and observed in a DIC microscope. 100 ovules of each genotype were observed. The embryo sac cells in *glcat14b-1*, *glcat14e-1*, *at4g27480-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be* and wt plants grown in LCNS and

BNS develop normally (Fig. 26; Supplementary figure 21). No differences were observed between the development of mutant and wt ovules.

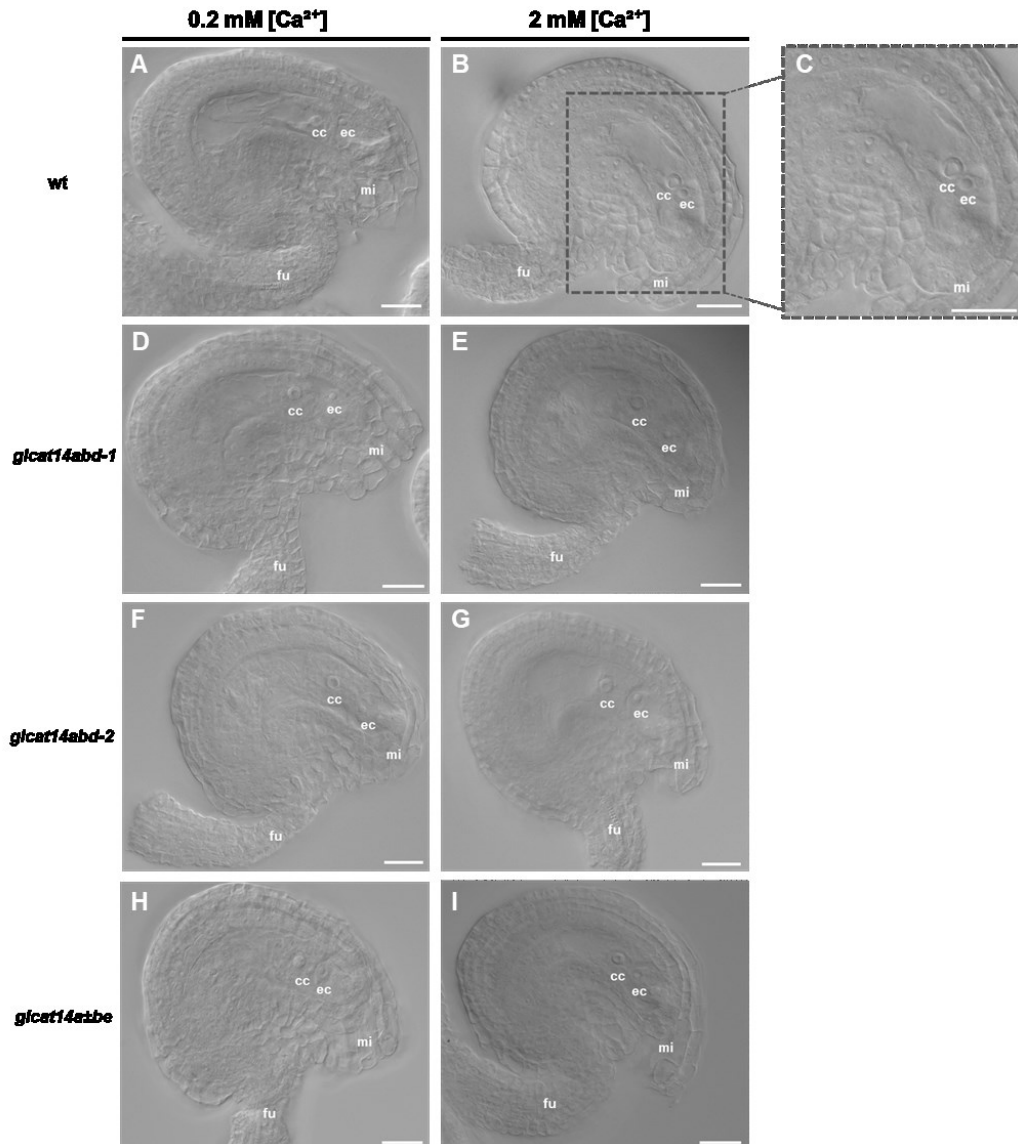


Fig. 26 - Cleared ovules of Arabidopsis mutant and wt plants grown using hydroponic solutions. Cleared ovules of wt (A-B), *glat14abd-1* (D-E), *glat14abd-2* (F-G) and *glat14a±be* (H-I) plants grown with LCNS (left column) and BNS (right column). (C) Detailed of the central and egg cells in a wild-type ovule grown in BNS. Abbreviations: cc, central cell; ec, egg cell; fu, funiculus; mi, micropyle. Size bar, 20 μ m.

As the female germline was developing correctly, the viability of the pollen was studied resorting to Alexander's staining (Fig. 26; Supplementary figure 22). Alexander's staining allows to distinguish pollen viable from non-viable based on the color of the pollen stained. Pollen walls are stained with malachite green dye and cytoplasm is stained with acid fuchsin. Thus, viable pollen stains in red-magenta while non-viable pollen stains in

green-blue (Alexander, 1969). The majority of pollen grains from *glcat14abd-1* were non-viable in plants grown in LCNS (Fig. 27C). Some pollen grains were also stained green in *glcat14abd-2* plants grown in LCNS (Fig. 27E) and in *glcat14abd-1* (Fig. 27D) plants grown in BNS.

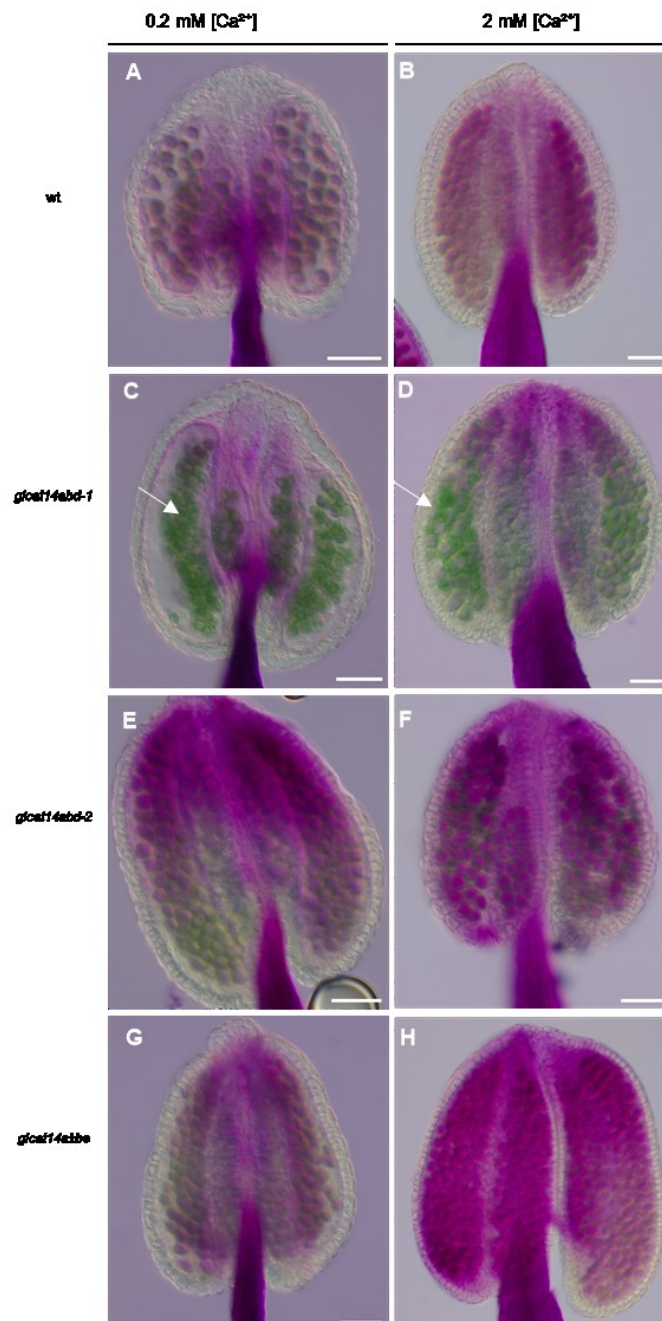


Fig. 27 - Pollen from Arabidopsis mutant and wt plants grown using hydroponic solutions stained with Alexander's staining. Anthers with pollen grains of wt (A-B), *glcat14abd-1* (C-D), *glcat14abd-2* (E-F) and *glcat14a±be* (G-H) plants grown with LCNS (left column) and BNS (right column). Pink-red stained pollen grains are viable. White arrows indicate green-blue stained non-viable pollen. Size bar, 50 μ m.

The output of this analyses was highly variable. Aborted pollen grains were also found in *atglcat14a±be*, *glcat14bd-1*, *glcat14ab*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* grown in LCNS and in *glcat14abd-2* grown in BNS (data not shown). The short length of the siliques and the absence of seeds in *glcat14abd-1* and *atglcat14abd-2* showed that the pollen grains stained in pink in these anthers may not be able to fertilize the ovules. To test this hypothesis, auto-fertilized pistils were analyzed with aniline blue staining in order to discover if the pollen grains are germinating in the stigma and consequently, if the PT grows until it finds an embryo sac (Fig. 26; Supplementary figure 23). Aniline blue stains a polysaccharide present in the PTs, callose, under ultraviolet light. Therefore, with the aid of the fluorescence microscopy, the PTs in the female tissues are clearly perceptible and can be followed until they enter the ovules. Aniline blue staining showed a low number of pollen grains germinated in the stigma of *glcat14abd-1* and *glcat14abd-2* plants grown in LCNS (Fig. 28C,E) when compared to wt grown in LCNS (Fig. 28A). However, the other mutants grown in LCNS or BNS displayed a similar number of PTs in the transmitting tissue when compared to wt (Fig. 28; Supplementary figure 23). Besides, the number of fertilized and non-fertilized ovules in each mutant pistil were counted (Fig. 29). An ovule is considered fertilized when a unique PT enter the micropyle (Fig. 30A) while a non-fertilized ovule shows no PT entering the ovule (Fig. 30B). Most ovules from *glcat14abd-1*, *glcat14abd-2* and *glcat14a±be* grown in LCNS (Fig. 29A) are non-fertilized. Whereas *glcat14b-1* grown in LCNS (Fig. 29A) have almost 50% of non-fertilized ovules. In the other mutants, no significant differences were found when compared to wt.

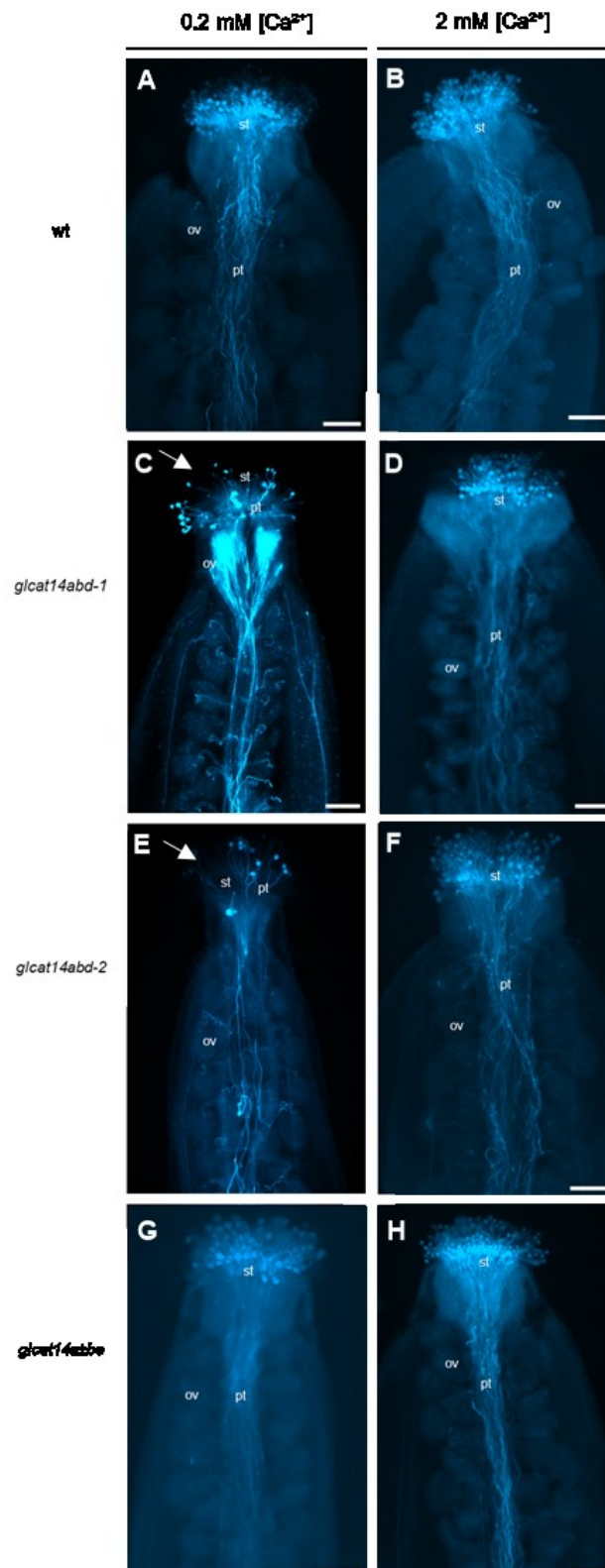


Fig. 28 - Aniline blue staining of Arabidopsis mutant and wt pistils of plants grown using hydroponic solutions. Pistils of wt (A-B), *glcat14abd-1* (C-D), *glcat14abd-2* (E-F) and *glcat14a±be* (G-H) plants grown with LCNS (left column) and BNS (right column). Abbreviations: ov, ovule; pt, pollen tube; st, stigma. White arrows indicate stigmas with low number of germinated pollen grains. Size bar, 100 μm.

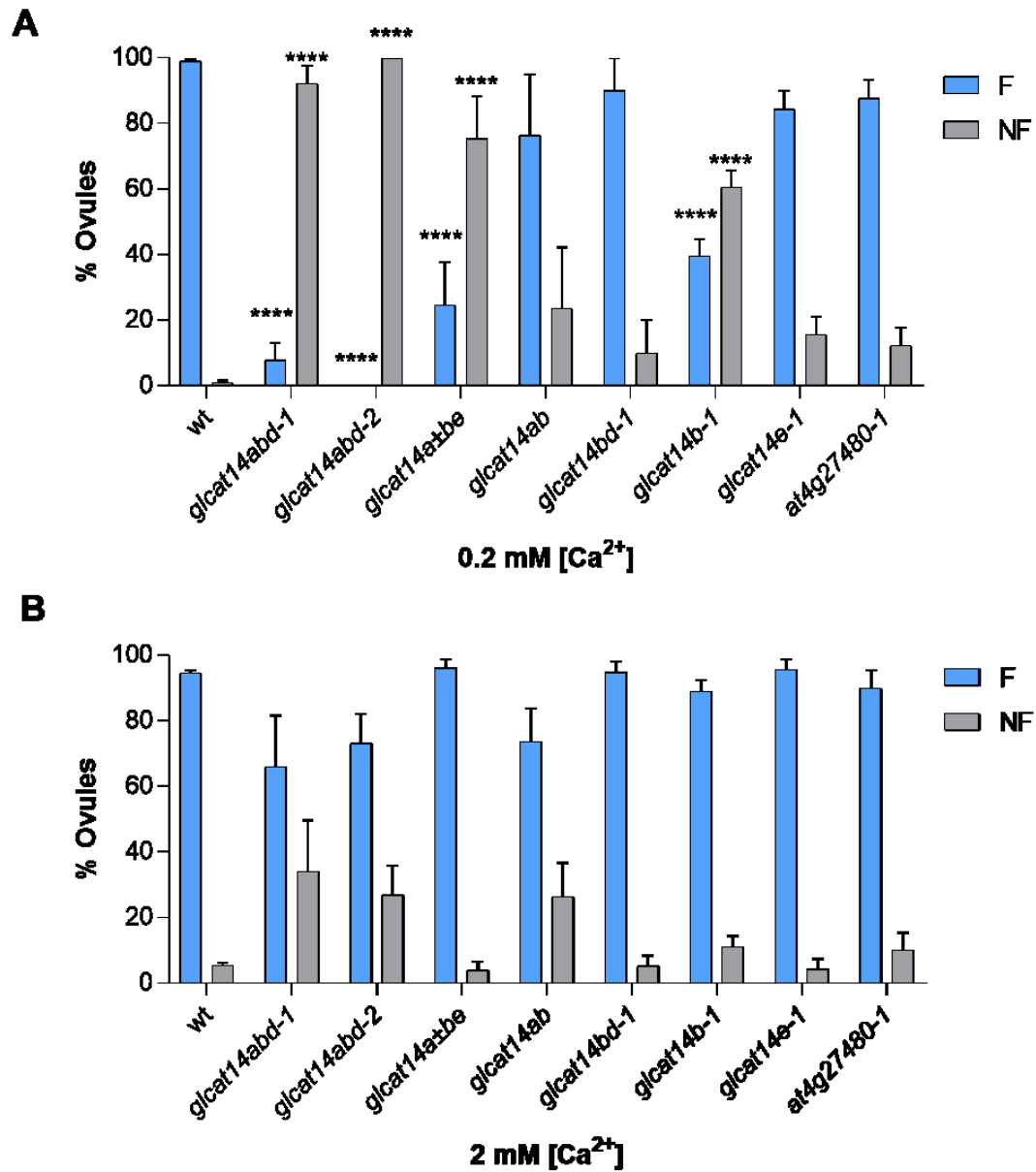


Fig. 29 - Percentage of fertilized and unfertilized ovules per pistil of Arabidopsis mutant and wt plants grown using hydroponic solutions.

Percentage of fertilized and unfertilized ovules per pistil of wt, *glcat14abd-1*, *glcat14abd-2*, *glcat14a±be*, *glcat14ab*, *glcat14bd-1*, *glcat14b-1*, *glcat14e-1* and *at4g27480-1* plants grown with LCNS (A) and BNS (B). Means of three biological replicates are shown. Abbreviations: F, fertilized ovules; N, non-fertilized ovules. Means $\pm$ SEM of three biological replicates are shown. Asterisks indicate value significantly different from wt expression according to a Dunnett's test (*, $p < 0.033$; **, $p < 0.002$; ***, $p < 0.0002$; ****, $p < 0.0001$).

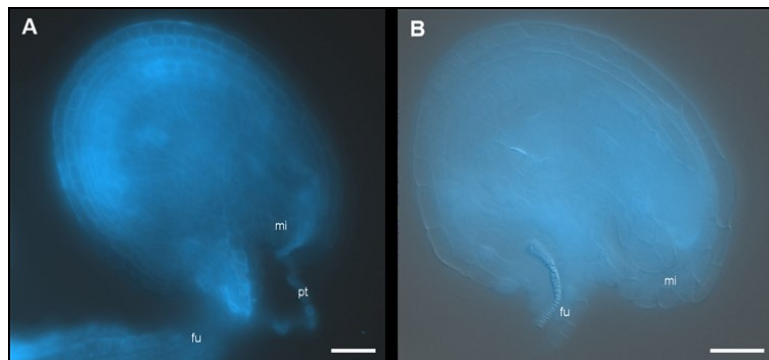


Fig. 30 - Wt ovules stained with aniline blue.

(A) Pollen tube entering the embryo sac of a wt ovule. (B) Non-fertilized wt ovule. Abbreviations: fu, funiculus; mi; micropyle; pt, pollen tube. Size bar, 20 μ m.

4. DISCUSSION

4.1. In silico analysis of GT14 family

Five proteins of the eleven members of GT14 family in Arabidopsis have been biochemical characterized and present β -GlcAT activity specific of type II AGs of AGPs (Knoch et al., 2013; Dilokpimol and Geshi, 2014; Hernández, 2018).

The sequence identity matrix suggested that the enzyme pairs At1g03520/At1g03340, GlcAT14A/GlcAt14B and GlcAT14E/At1g53100 might have similar activity as the amino acid sequences have high sequence identity. Knoch et al. (2013) and Lopez-Hernandez (2018) described a 73% of sequence identity shared by GlcAT14A and GlcAT14B, confirming these results. In like manner, GlcAT14C shares almost 56% of sequence identity with GlcAT14A and 53% with GlcAT14B. However, GlcAT14D and GlcAT14E only share 40% of sequence identity but they both have GlcAT activity as described by Lopez-Hernandez (2018). These results suggested that even enzymes sharing a lower percentage of sequence identity may still have the same activity. However, no biochemical evidence of their function has been reported to date. Therefore, the eleven members of GT14 family are all putative GlcAT specific of type II AGs of AGPs. The phylogenetic analysis reflected this similarity, clustering together GlcAt14D/At1g71070, At1g03520/At4g03340, GlcAT14A/GlcAt14B and GlcAT14E/At1g53100.

The GlcAT activity on type II AGs of AGPs seems to be conserved across embryophytes. *P. patens* is a bryophyte, a non-vascular terrestrial plant, which contains AGPs with GlcA recognized by the LM2 antibody (Berry et al., 2016; Ruprecht et al., 2017). *S. moellendorffii* is a lycophyte, a basal vascular plant, in which GlcA of AGPs was also identified by LM2 (Harholt et al., 2012). *H. vulgare* belongs to the grass family and is one of the first cultivated grains. AGPs were also recognized by LM2 in this species (Marzec et al., 2015). Finally, *P. trichocarpa* was the first tree to have its genome sequenced. A recent bioinformatic approach identify 162 AGPs in *P. trichocarpa* (Showalter et al., 2016).

4.2. Expression profile of GT14 genes

qPCR analysis of GT14 genes was performed to examine their expression and provide information relevant to phenotypic analysis. Most genes were widely expressed. Within a given organ or tissue, the GT14 genes exhibited both temporal and spatial differences in their expression patterns. Some of these patterns were corroborated by RNA-seq data (Klepikova et al., 2016). *GlcAT14A*, *GlcAT14B*, *GlcAT14D* and *GlcAT14E* were the

genes most transcribed in inflorescences and therefore, these four genes were the most promising for mutant analysis in the reproductive process in Arabidopsis.

4.3. Plant marker lines

To corroborate the GT14 expression patterns, plants containing the constructs *GlcAT14Bprom:YFP*, *GlcAT14Eprom:YFP*, *At4g27480prom:YFP*, *At1g71070prom:YFP*, *At1g53100prom:YFP* and *At1g03520prom:YFP* were obtained. Due to restriction of time, only the plants containing the *GlcAT14Bprom:YFP*, *GlcAT14Eprom:YFP* and *At4g27480prom:YFP* constructs were examined. The three genes showed expression in the wall of anthers. *AGP6*, *AGP11*, *ENODL12* and *ENODL15* are expressed in the wall of anthers (Levitin et al., 2008; Mashiguchi et al., 2009). One of the layers of the wall of anthers is the tapetum which is involved in the synthesis of sporopollenin that composes the exine outer layer of mature pollen walls. Hence, these four AGPs might be involved in pollen development.

A variety of AGPs are expressed in pollen grains and PTs such as *AGP6*, *AGP11*, *AGP23*, *AGP40*, *FLA3*, *FLA5* and *FLA14* (Pereira et al., 2006; Li et al., 2010; Nguema-Ona et al., 2012; Pereira et al., 2014; Miao et al., 2018). However, none of the GlcATs analyzed was expressed in the pollen. Studies with LM2 antibody, that specifically recognizes Gal residues bound to GlcA (Ruprecht et al., 2017), showed a strong labelling at the tip of PTs (Pereira et al., 2006; Dardelle et al., 2010). Therefore, it is expected that one or more GlcATs be present in PTs.

In addition, *GlcAT14Bprom:YFP* and *GlcAT14Eprom:YFP* were expressed in the transmitting tissue. Both GlcATs add GlcA onto AGPs (Lopez-Hernandez, 2018). AGPs like *AGP1*, *JAGGER*, *AGP15* and *ENODL15* are expressed in the transmitting tissue (Pereira et al., 2014; Hou et al., 2016; Pereira et al., 2016b) and are proposed to sustain PT growth (Cheung et al., 1995).

YFP expression driven by *GlcAT14B* and *GlcAT14E* promoters was detected in the funiculus where PTs turn to an ovule after growing in the transmitting tissue. *AGP1*, *AGP12*, *FLA8* and *ENODL15* (Pereira et al., 2014; Hou et al., 2016; Costa et al., 2019) are expressed in the funiculus and might play a signaling role in this tissue (Pereira et al., 2016a).

GlcAT14B and *GlcAT14E* showed expression at the chalazal tissues of the ovules. Several AGPs are expressed in the chalaza such as *AGP1*, *AGP12*, *FLA8* and *FLA10* (Pereira et al., 2014; Costa et al., 2019). These glycoproteins might have a function in nutrition uptake into the endosperm as it occurs via the chalazal pole (Ingram, 2010).

Overall, the expression of the three GlcATs was consistent with the distribution of AGPs and the respective function.

4.4. Subcellular localization

Glycosylation of AGPs occurs in Golgi vesicles (Kieliszewski, 2001). Localization studies of other Arabidopsis AGP GTs also localized in the Golgi (Qu et al., 2008; Wu et al., 2010; Basu et al., 2013; Geshi et al., 2013; Gille et al., 2013; Knoch et al., 2013; Dilokpimol et al., 2014; Basu et al., 2015a; Basu et al., 2015b; Ogawa-Ohnishi and Matsubayashi, 2015; Temple et al., 2019). As GlcAT14A was localized in the Golgi apparatus (Knoch et al., 2013), Arabidopsis GT14 enzymes were expected to reside in the Golgi vesicles.

All eleven members of GT14 family were predicted as type II membrane proteins, containing a short cytoplasmic N-terminal region followed by a hydrophobic transmembrane domain and a long non-cytoplasmic C-terminal region containing the branch domain. GlcAT14E, At4g27480, At1g71070 and UNE7 were predicted to be localized to the Golgi as reported by (Ye et al., 2011). However, some discrepancies were found relative to the other GlcATs subcellular localization predictions probably due to the use of different programs. Nevertheless, all subcellular localization predictions should be confirmed by *in vitro* experiments. Thus, the predictions were confirmed by heterologous expression of fluorescently tagged protein fusions in *N. benthamiana* leaves. *GlcAT14B:GFP*, *GlcAT14E:GFP*, *At4g27480:GFP*, *At1g71070:GFP*, *At1g53100:GFP* and *At1g03520:GFP* are localized to Golgi vesicles consistent with bioinformatics predictions (Ye et al., 2011). Nevertheless, co-expression experiments should be performed between GT14 CDS fusions and a Golgi marker.

4.5. Functional analysis of glucuronosyltransferase genes

The availability of genomic resources and improvement in molecular techniques have allowed the identification of several GTs. However, only a small part has been functionally characterized. The insertion mutants *glcat14b-1*, *glcat14e-1*, *at4g27480-1*, *glcat14ab*, *glcat14bd-1*, *glcat14abd-1*, *glcat14abd-2* and *glcat14a±be* were characterized. As AGPs were proposed to bind and store Ca^{2+} through GlcA residues, these mutants were grown using two nutrient solutions LCNS ($[\text{Ca}^{2+}] = 0.2 \text{ mM}$) and BNS ($[\text{Ca}^{2+}] = 2 \text{ mM}$). All single mutants were able to grow and reproduce normally, indicating that these genes are not essential for plant survival as reported by Knoch et al. (2013) and Lopez-Hernandez (2018). Single mutants showed no detectable growth phenotypes

because some functional redundancy is expected in this gene family and one gene can compensate another knocked out gene. Besides, single mutants still presented enough quantity of glycosylated AGPs. However, as double and triple mutants have a reduction in AGP glycosylation some pleiotropic phenotypes in growth and development are displayed. The triple mutants *glcat14abd-1* and *glcat14abd-2* grown in LCNS were smaller plants with smaller siliques comprising a noteworthy amount of aborted ovules and seeds relative to wt. It is also noteworthy that this phenotype was complemented when grown in BNS.

Phenotypic analysis of the mutants grown in LCNS revealed no impact of the mutations on female gametophyte development as the embryo sacs reached maturity and were ready to be fertilized. However, the triple mutants bear defects in the male tissues, having non-viable pollen grains. This defect was reflected in the low number of fertilized ovules comparing to wt. This suggested that pollen viability may not be the only reason for partial sterility of triple mutants. PT guidance through the pistil to the ovule might also be affected.

During the hydroponic growth experiment with nutrient solutions, the utmost care was taken to avoid Ca^{2+} contamination. However, some discrepant results may result from external contamination.

GlcAT14A, GlcAT14B and GlcAT14D seem to be important for pollen development and in PT attraction to the ovule indicating that AG polysaccharides are required for these functions. In fact, the disaccharide 4-Me-GlcA- β -(1-6)-Gal, named AMOR, induces competency of the PT to respond to ovular attractant LURE peptides in *Torenia fournieri* (Mizukami et al., 2016).

Therefore, this suggests that GlcAT14A, GlcAT14B and GlcAT14D are involved in glycosylation of pollen specific AGPs. Several AGPs are reported to be expressed in the pollen including *AGP6*, *AGP11*, *AGP23*, *AGP40*, *FLA3* and *FLA14* (Pereira et al., 2006; Li et al., 2010; Nguema-Ona et al., 2012; Pereira et al., 2014; Miao et al., 2018). Indeed, two GTs specific of AG polysaccharides were shown to be involved in pollen development and PT growth. KNS4, a β -1,3-GALT, participates in the synthesis of cell wall components required for the development of microspores and pollen cell walls (Li et al., 2017; Suzuki et al., 2017). Similarly, UNE7 is involved in PT elongation and attraction to the ovule (Pagnussat et al., 2005; Boavida et al., 2009). Consequently, UNE7 might also transfer GlcA onto AGPs.

AGPs are modified with [Me]GlcA which can hold Ca^{2+} in a reversible manner (Lamport and Várnai, 2013). Most AGPs are predicted to be anchored to the plasma membrane

via a GPI anchor and thus serving as an extracellular capacitor of Ca^{2+} , generating a layer of Ca^{2+} available to the cell to facilitate Ca^{2+} signaling processes (Lamport and Várnai, 2013; Lamport et al., 2014). However, Ca^{2+} is only released from AGPs when a localized drop of pH occurs in the apoplastic side of the plasma membrane.

Mutants of GlcATs have reduced quantity of [Me]GlcA on AGPs and consequently reduced Ca^{2+} -binding capacity *in vitro* and likely *in vivo* (Lopez-Hernandez, 2018) which causes deficiencies in Ca^{2+} -signaling pathways. However, when these plants were grown with nutrient solutions supplemented with Ca^{2+} (BNS), such deficiencies were reduced. Multiple GlcAT mutants will display more pronounced phenotypes and therefore, future analysis will provide more information about this family.

Growing PTs have a tip-focused gradient of $[\text{Ca}^{2+}]$ maintained by the influx of extracellular Ca^{2+} through Ca^{2+} channels (Feijó et al., 1995). AGPs located in the PTs upon a pH drop in the apoplast might release Ca^{2+} that is uptake through Ca^{2+} channels and help maintain the gradient of $[\text{Ca}^{2+}]$.

Like PT growth, diverse growth and development processes rely on Ca^{2+} gradients or oscillations. AGPs might be a source of extracellular Ca^{2+} and may regulate the release of Ca^{2+} in different cells/tissues at different times.

5. CONCLUSIONS

Polysaccharides cover the surface of AGPs and account for almost 90% of the molecular mass of AGPs. Thus, the carbohydrate moiety may play a critical role in AGPs function. Hence, in order to understand AGPs it is very important to study the enzymes covering the sweet proteins, the GTs. With new advance techniques appearing every day, plant biologists are overwhelmed with new candidate genes encoding GTs. A lot of effort will be required in order to analyze biochemically and functionally each putative GT involved in AGP glycosylation in Arabidopsis. Biochemical assays deciphering substrate specificity of GTs will provide important information to address this effort. GTs mutants' studies may elucidate AGPs function. GTs are highly redundant enzymes and therefore, it will be necessary to create and analyze double, triple, quadruple and quintuple mutants to understand their specific role.

The GT14 family has 11 members and almost half have been characterized (Knoch et al., 2013; Dilokpimol and Geshi, 2014; Lopez-Hernandez, 2018). It is likely that the other six members of this family have GlcAT activity on AGs.

The aim of this chapter was to identify the role of [Me]GlcA of AGPs during sexual reproduction in Arabidopsis. This was accomplished via characterization of the expression pattern of GT14 genes, investigation of subcellular localization of GT14 enzymes and identification and analysis of GT14 mutants.

The analysis of mutants of GlcATs suggested GlcATs are not essential for plant survival and reproduction however [Me]GlcA on AGPs is very important and double fertilization relies on AGP-Ca²⁺ interaction. The phenotypes found in the reproductive tissues were suppressed when plants grown in normal concentrations of Ca²⁺. Ca²⁺-binding proteins and Ca²⁺-dependent cellular processes could be studied in pollen of the triple mutants.

Further characterization of other GT14 members will provide more information on the activity and specificity of this family. To determine the specific substrate required by GlcATs, *in vitro* assays should be performed. Moreover, other GT14 promoters may be fused with a fluorescent protein to identify in which cells and tissues GlcATs are expressed. Furthermore, tissue sections of GT14 mutants may be studied to observe the lack of [Me]GlcA on AGs of AGPs by using the LM2 antibody which specifically binds to GlcA terminally linked to $\beta(1-6)$ -galactans (Ruprecht et al., 2017).

Additionally, to determine if some GlcAT interact with each other, photobleaching FRET technique may be used to indicate plausible protein–protein interaction. Besides, AGPs fused to a fluorescent protein can be heterologously expressed and purified to verify which

AGPs carry [Me]GlcA. These fusions could also be expressed in the triple mutant background and serve as a negative control. This task will also be informative about the AGs structure in AGPs.

These results improve the study of these sweet proteins in the sexual plant reproductive process, opening doors for new studies to continue to understand the biosynthesis and function of AGPs.

6. REFERENCES

- Alexander MP** (1969) Differential staining of aborted and nonaborted pollen. *Biotech Histochem* **44**: 117–122 doi: 10.3109/10520296909063335
- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H** (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**: 420–423 doi: 10.1038/s41587-019-0036-z
- Basu D, Liang Y, Liu X, Himmeldirk K, Faik A, Kieliszewski M, Held M, Showalter AM** (2013) Functional identification of a hydroxyproline-O-galactosyltransferase specific for arabinogalactan protein biosynthesis in arabidopsis. *J Biol Chem* **288**: 10132–10143 doi: 10.1074/jbc.M112.432609
- Basu D, Tian L, DeBrosse T, Poirier E, Emch K, Herock H, Travers A, Showalter AM** (2016) Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) mediates root growth and seed mucilage adherence via a cell wall receptor-like kinase (FEI1/FEI2) Pathway in Arabidopsis. *PLoS One* **11**: e0145092 doi: 10.1371/journal.pone.0145092
- Basu D, Tian L, Wang W, Bobbs S, Herock H, Travers A, Showalter AM** (2015a) A small multigene hydroxyproline-O-galactosyltransferase family functions in arabinogalactan-protein glycosylation, growth and development in Arabidopsis. *BMC Plant Biol* **15**: 1–23 doi: 10.1186/s12870-015-0670-7
- Basu D, Wang W, Ma S, DeBrosse T, Poirier E, Emch K, Soukup E, Tian L, Showalter AM** (2015b) Two hydroxyproline galactosyltransferases, GALT5 and GALT2, function in arabinogalactan-protein glycosylation, growth and development in Arabidopsis. *PLoS One* **10**: e0125624 doi: 10.1371/journal.pone.0125624
- Berry EA, Tran ML, Dimos CS, Budziszek MJ, Scavuzzo-Duggan TR, Roberts AW** (2016) Immuno and affinity cytochemical analysis of cell wall composition in the moss *Physcomitrella patens*. *Front Plant Sci* **7**: 248 doi: 10.3389/fpls.2016.00248
- Bierhuizen MFA, Mattei MG, Fukuda M** (1993) Expression of the developmental I antigen by a cloned human cDNA encoding a member of a β -1,6-N-acetylglucosaminyltransferase gene family. *Genes Dev* **7**: 468–478 doi: 10.1101/gad.7.3.468
- Boavida LC, Shuai B, Yu H-J, Pagnussat GC, Sundaresan V, McCormick S** (2009) A collection of Ds insertional mutants associated with defects in male gametophyte development and function in *Arabidopsis thaliana*. *Genetics* **181**: 1369–85 doi:

- 10.1534/genetics.108.090852
- Cheung AY, Wang H, Wu H ming** (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. *Cell* **82**: 383–393 doi: 10.1016/0092-8674(95)90427-1
- Chou YH, Pogorelko G, Zabolina OA** (2012) Xyloglucan xylosyltransferases XXT1, XXT2, and XXT5 and the glucan synthase CSLC4 form Golgi-localized multiprotein complexes. *Plant Physiol* **159**: 1355–1366 doi: 10.1104/pp.112.199356
- Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S** (2019) In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reprod* **32**: 353–370 doi: 10.1007/s00497-019-00376-7
- Coutinho PM, Deleury E, Davies GJ, Henrissat B** (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**: 307–317 doi: 10.1016/S0022-2836(03)00307-3
- Curtis MD, Grossniklaus U** (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol* **133**: 462–469 doi: 10.1104/pp.103.027979
- Dardelle F, Lehner A, Ramdani Y, Bardor M, Lerouge P, Driouich A, Mollet JC** (2010) Biochemical and immunocytological characterizations of arabidopsis pollen tube cell wall. *Plant Physiol* **153**: 1563–1576 doi: 10.1104/pp.110.158881
- Denninger P, Bleckmann A, Lausser A, Vogler F, Ott T, Ehrhardt DW, Frommer WB, Sprunck S, Dresselhaus T, Grossmann G** (2014) Male-female communication triggers calcium signatures during fertilization in arabidopsis. *Nat Commun* **5**: 1–12 doi: 10.1038/ncomms5645
- Dilokpimol A, Geshi N** (2014) Arabidopsis thaliana glucuronosyltransferase in family GT14. *Plant Signal Behav* **9**: e28891 doi: 10.4161/psb.28891
- Dilokpimol A, Poulsen CP, Vereb G, Kaneko S, Schulz A, Geshi N** (2014) Galactosyltransferases from arabidopsis thaliana in the biosynthesis of type II arabinogalactan: Molecular interaction enhances enzyme activity. *BMC Plant Biol* **14**: 1–14 doi: 10.1186/1471-2229-14-90
- Dodd AN, Org Kudla J, Sanders D** (2010) The language of calcium signaling. **61**: doi: 10.1146/annurev-arplant-070109-104628
- Dresselhaus T, Sprunck S, Wessel GM** (2016) Fertilization mechanisms in flowering plants. *Curr Biol* **26**: R125–R139 doi: 10.1016/j.cub.2015.12.032
- Edel KH, Marchadier E, Brownlee C, Kudla J, Hetherington AM** (2017) The Evolution

- of Calcium-Based Signalling in Plants. *Curr Biol* **27**: R667–R679 doi: 10.1016/j.cub.2017.05.020
- Edwards K, Johnstone C, Thompson C** (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* **19**: 1349 doi: 10.1093/nar/19.6.1349
- El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, et al** (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427–D432 doi: 10.1093/nar/gky995
- Emanuelsson O, Nielsen H, Brunak S, Von Heijne G** (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* **300**: 1005–1016 doi: 10.1006/jmbi.2000.3903
- Eudes A, Mouille G, Thévenin J, Goyallon A, Minic Z, Jouanin L** (2008) Purification, cloning and functional characterization of an endogenous beta-glucuronidase in *Arabidopsis thaliana*. *Plant Cell Physiol* **49**: 1331–1341 doi: 10.1093/pcp/pcn108
- Feijó JA, Malhó R, Obermeyer G** (1995) Ion dynamics and its possible role during in vitro pollen germination and tube growth. *Protoplasma* **187**: 155–167 doi: 10.1007/BF01280244
- Geshi N** (2014) Arabinogalactan glycosyltransferases: enzyme assay, protein-protein interaction, subcellular localization, and perspectives for application. *Adv Bot* **2014**: 1–7 doi: 10.1155/2014/434979
- Geshi N, Johansen JN, Dilokpimol A, Rolland A, Belcram K, Verger S, Kotake T, Tsumuraya Y, Kaneko S, Tryfona T, et al** (2013) A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in *Arabidopsis*. *Plant J* **76**: 128–137 doi: 10.1111/tpj.12281
- Gille S, Sharma V, Baidoo EEK, Keasling JD, Scheller HV, Pauly M** (2013) Arabinosylation of a Yariv-precipitable cell wall polymer impacts plant growth as exemplified by the *arabidopsis* glycosyltransferase mutant ray1. *Mol Plant* **6**: 1369–1372 doi: 10.1093/mp/sst029
- Gloster TM** (2014) Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol* **28**: 131–141 doi: 10.1016/j.sbi.2014.08.012
- Hansen SF, Bettler E, Wimmerová M, Imberty A, Lerouxel O, Breton C** (2009) Combination of several bioinformatics approaches for the identification of new putative glycosyltransferases in *arabidopsis*. *J Proteome Res* **8**: 743–753 doi: 10.1021/pr800808m
- Hansen SF, Harholt J, Oikawa A, Scheller H V.** (2012) Plant glycosyltransferases

- beyond CAZy: A perspective on DUF families. *Front Plant Sci* **3**: 1–10 doi: 10.3389/fpls.2012.00059
- Harholt J, Sørensen I, Fangel J, Roberts A, Willats WGT, Scheller HV, Petersen BL, Banks JA, Ulvskov P** (2012) The glycosyltransferase repertoire of the spikemoss *selaginella moellendorffii* and a comparative study of its cell wall. *PLoS One* **7**: e35846 doi: 10.1371/journal.pone.0035846
- Henrissat B, Davies G** (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**: 637–644 doi: 10.1016/S0959-440X(97)80072-3
- Henrissat B, Davies GJ** (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol* **124**: 1515–1519 doi: 10.1104/pp.124.4.1515
- Hooper CM, Castleden IR, Tanz SK, Aryamanesh N, Millar AH** (2017) SUBA4: The interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res* **45**: D1064–D1074 doi: 10.1093/nar/gkw1041
- Hou Y, Guo X, Cyprys P, Zhang Y, Bleckmann A, Cai L, Huang Q, Luo Y, Gu H, Dresselhaus T, et al** (2016) Maternal ENODLs Are Required for Pollen Tube Reception in Arabidopsis. *Curr Biol* **26**: 2343–2350 doi: 10.1016/j.cub.2016.06.053
- Ingram GC** (2010) Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195–214 doi: 10.1007/s00709-010-0184-y
- Jiao J, Mizukami AG, Sankaranarayanan S, Yamguchi J, Itami K, Higashiyawma T** (2017) Structure-activity relation of AMOR sugar molecule that activates pollen-tubes for ovular guidance. *Plant Physiol* **173**: 354–363 doi: 10.1104/pp.16.01655
- Johnson MA, Harper JF, Palanivelu R** (2019) A fruitful journey: pollen tube navigation from germination to fertilization. *Annu Rev Plant Biol* **70**: 809–837 doi: 10.1146/annurev-arplant-050718-100133
- Kieliszewski MJ** (2001) The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* **57**: 319–323 doi: 10.1016/S0031-9422(01)00029-2
- Klepikova A V., Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA** (2016) A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant J* **88**: 1058–1070 doi: 10.1111/tpj.13312
- Knoch E, Dilokpimol A, Geshi N** (2014) Arabinogalactan proteins: Focus on carbohydrate active enzymes. *Front Plant Sci* **5**: 1–9 doi: 10.3389/fpls.2014.00198
- Knoch E, Dilokpimol A, Tryfona T, Poulsen CP, Xiong G, Harholt J, Petersen BL,**

- Ulvskov P, Hadi MZ, Kotake T, et al** (2013) A β -glucuronosyltransferase from *Arabidopsis thaliana* involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. *Plant J* **76**: 1016–1029 doi: 10.1111/tpj.12353
- Kuhn J, Götting C, Schnölzer M, Kempf T, Brinkmann T, Kleesiek K** (2001) First isolation of human UDP-D-Xylose: proteoglycan core protein β -D-Xylosyltransferase secreted from cultured JAR choriocarcinoma cells. *J Biol Chem* **276**: 4940–4947 doi: 10.1074/jbc.M005111200
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K** (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547–1549 doi: 10.1093/molbev/msy096
- Lamport DTA, Tan L, Held MA, Kieliszewski MJ** (2018) Pollen tube growth and guidance: Occam's razor sharpened on a molecular arabinogalactan glycoprotein Rosetta Stone. *New Phytol* **217**: 491–500 doi: 10.1111/nph.14845
- Lamport DTA, Várnai P** (2013) Periplasmic arabinogalactan glycoproteins act as a calcium capacitor that regulates plant growth and development. *New Phytol* **197**: 58–64 doi: 10.1111/nph.12005
- Lamport DTA, Varnai P, Seal CE** (2014) Back to the future with the AGP-Ca²⁺ flux capacitor. *Ann Bot* **114**: 1069–1085 doi: 10.1093/aob/mcu161
- Levitin B, Richter D, Markovich I, Zik M** (2008) Arabinogalactan proteins 6 and 11 are required for stamen and pollen function in *Arabidopsis*. *Plant J* **56**: 351–363 doi: 10.1111/j.1365-313X.2008.03607.x
- Li J, Yu M, Geng LL, Zhao J** (2010) The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of *Arabidopsis*. *Plant J* **64**: 482–497 doi: 10.1111/j.1365-313X.2010.04344.x
- Li WL, Liu Y, Douglas CJ** (2017) Role of glycosyltransferases in pollen wall primexine formation and exine patterning. *Plant Physiol* **173**: 167–182 doi: 10.1104/pp.16.00471
- Liang Y, Basu D, Pattathil S, Xu WL, Venetos A, Martin SL, Faik A, Hahn MG, Showalter AM** (2013) Biochemical and physiological characterization of fut4 and fut6 mutants defective in arabinogalactan-protein fucosylation in *Arabidopsis*. *J Exp Bot* **64**: 5537–5551 doi: 10.1093/jxb/ert321
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B** (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495 doi: 10.1093/nar/gkt1178

- Lopez-Hernandez F** (2018) Identification of the role of [methyl]glucuronic acid on arabinogalactan polysaccharides in *Arabidopsis thaliana*. PhD dissertation, University of Cambridge, UK. doi: doi.org/10.17863/CAM.23619
- Ma Y, Zeng W, Bacic A, Johnson K** (2018) AGPs Through Time and Space. *Annu Plant Rev* **1**: 1–38 doi: 10.1002/9781119312994.apr0608
- Marzec M, Szarejko I, Melzer M** (2015) Arabinogalactan proteins are involved in root hair development in barley. *J Exp Bot* **66**: 1245–1257 doi: 10.1093/jxb/eru475
- Mashiguchi K, Asami T, Suzuki Y** (2009) Genome-wide identification, structure and expression studies, and mutant collection of 22 early nodulin-like protein genes in *arabidopsis*. *Biosci Biotechnol Biochem* **73**: 2452–2459 doi: 10.1271/bbb.90407
- Miao Y, Lin S, Cao J** (2018) The fasciclin-like arabinogalactan protein gene, FLA14, is involved in microspore development of *Arabidopsis*. Poster presented at 25th International Congress on Sexual Plant Reproduction; 2018 Jun 11–16; Gifu, Japan
- Mizukami AG, Inatsugi R, Jiao J, Kotake T, Kuwata K, Ootani K, Okuda S, Sankaranarayanan S, Sato Y, Maruyama D, et al** (2016) The AMOR arabinogalactan sugar chain induces pollen-tube competency to respond to ovular guidance. *Curr Biol* **26**: 1091–1097 doi: 10.1016/j.cub.2016.02.040
- Nelson BK, Cai X, Nebenführ A** (2007) A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J* **51**: 1126–1136 doi: 10.1111/j.1365-313X.2007.03212.x
- Nguema-Ona E, Coimbra S, Vitré-Gibouin M, Mollet JC, Driouich A** (2012) Arabinogalactan proteins in root and pollen-tube cells: distribution and functional aspects. *Ann Bot* **110**: 383–404 doi: 10.1093/aob/mcs143
- Ogawa-Ohnishi M, Matsubayashi Y** (2015) Identification of three potent hydroxyproline O-galactosyltransferases in *Arabidopsis*. *Plant J* **81**: 735–746 doi: 10.1111/tpj.12764
- Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V** (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* **132**: 603–14 doi: 10.1242/dev.01595
- 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**: 1–15 doi: 10.3389/fpls.2016.01895
- Pereira AM, Masiero S, Nobre MS, Costa ML, Solís 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 doi: 10.1093/jxb/eru300
- 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 doi: 10.1016/j.molp.2016.01.002
- Pereira LG, Coimbra S, Oliveira H, Monteiro L, Sottomayor M** (2006) Expression of arabinogalactan protein genes in pollen tubes of *Arabidopsis thaliana*. *Planta* **223**: 374–380 doi: 10.1007/s00425-005-0137-4
- Pickard BG, Fujiki M** (2005) Ca²⁺ pulsation in BY-2 cells and evidence for control of mechanosensory Ca²⁺-selective channels by the plasmalemmal reticulum. *Funct Plant Biol* **32**: 863–879 doi: 10.1071/FP05045
- Qu Y, Egelund J, Gilson PR, Houghton F, Gleeson PA, Schultz CJ, Bacic A** (2008) Identification of a novel group of putative *Arabidopsis thaliana* β -(1,3)-galactosyltransferases. *Plant Mol Biol* **68**: 43–59 doi: 10.1007/s11103-008-9351-3
- Rounds CM, Bezanilla M** (2013) Growth mechanisms in tip-growing plant cells. *Annu Rev Plant Biol* **64**: 243–265 doi: 10.1146/annurev-arplant-050312-120150
- Roy S, Jauh GY, Hepler PK, Lord EM** (1998) Effects of Yariv phenylglycoside on cell wall assembly in the lily pollen tube. *Planta* **204**: 450–458 doi: 10.1007/s004250050279
- Ruprecht C, Bartetzko MP, Senf D, Dallabernadina P, Boos I, Andersen MCF, Kotake T, Knox JP, Hahn MG, Clausen MH, et al** (2017) A synthetic glycan microarray enables epitope mapping of plant cell wall glycan-directed antibodies. *Plant Physiol* **175**: 1094–1104 doi: 10.1104/pp.17.00737
- Schwacke R, Schneider A, Van Der Graaff E, Fischer K, Catoni E, Desimone M, Frommer WB, Flügge UI, Kunze R** (2003) ARAMEMNON, a novel database for *Arabidopsis* integral membrane proteins. *Plant Physiol* **131**: 16–26 doi: 10.1104/pp.011577
- Showalter AM, Basu D** (2016a) Glycosylation of arabinogalactan-proteins essential for development in *Arabidopsis*. *Commun Integr Biol* **9**: e1177687 doi: 10.1080/19420889.2016.1177687
- Showalter AM, Basu D** (2016b) Extensin and arabinogalactan-protein biosynthesis: Glycosyltransferases, research challenges, and biosensors. *Front Plant Sci* **7**: 1–9 doi: 10.3389/fpls.2016.00814
- Showalter AM, Keppler BD, Liu X, Lichtenberg J, Welch LR** (2016) Bioinformatic

- Identification and Analysis of Hydroxyproline-Rich Glycoproteins in *Populus trichocarpa*. *BMC Plant Biol* **16**: 1–34 doi: 10.1186/s12870-016-0912-3
- Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N** (2010) PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Res* **38**: D161-6 doi: 10.1093/nar/gkp885
- Smyth DR, Bowman JL, Meyerowitz EM** (1990) Early flower development in *Arabidopsis*. *Plant Cell* **2**: 755–767 doi: 10.1105/tpc.2.8.755
- Sonnhammer EL, von Heijne G, Krogh A** (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* **6**: 175–182
- Sparkes IA, Runions J, Kearns A, Hawes C** (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**: 2019–2025 doi: 10.1038/nprot.2006.286
- Su S, Higashiyama T** (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. *Plant Reprod* **31**: 67–75 doi: 10.1007/s00497-018-0329-2
- Suzuki T, Narciso JO, Zeng W, van de Meene A, Yasutomi M, Takemura S, Lampugnani ER, Doblin MS, Bacic A, Ishiguro S** (2017) KNS4/UPEX1: A type II arabinogalactan β -(1,3)-galactosyltransferase required for pollen exine development. *Plant Physiol* **173**: 183–205 doi: 10.1104/pp.16.01385
- Tan L, Eberhard S, Pattathil S, Warder C, Glushka J, Yuan C, Hao Z, Zhu X, Avci U, Miller JS, et al** (2013) An *Arabidopsis* cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* **25**: 270–287 doi: 10.1105/tpc.112.107334
- Tekleyohans DG, Mao Y, Kägi C, Stierhof YD, Groß-Hardt R** (2017) Polyspermy barriers: a plant perspective. *Curr Opin Plant Biol* **35**: 131–137 doi: 10.1016/j.pbi.2016.11.012
- Temple H, Mortimer JC, Tryfona T, Yu X, Lopez-Hernandez F, Sorieul M, Anders N, Dupree P** (2019) Two members of the DUF579 family are responsible for arabinogalactan methylation in *Arabidopsis*. *Plant Direct* **3**: 1–4 doi: 10.1002/pld3.117
- Thompson JD, Higgins DG, Gibson TJ** (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680

doi: 10.1093/nar/22.22.4673

- Tryfona T, Liang HC, Kotake T, Tsumuraya Y, Stephens E, Dupree P** (2012) Structural characterization of Arabidopsis leaf arabinogalactan polysaccharides. *Plant Physiol* **160**: 653–666 doi: 10.1104/pp.112.202309
- Tryfona T, Theys TE, Wagner T, Stott K, Keegstra K, Dupree P** (2014) Characterisation of FUT4 and FUT6 α -(1→2)-fucosyltransferases reveals that absence of root arabinogalactan fucosylation increases arabidopsis root growth salt sensitivity. *PLoS One* **9**: e93291 doi: 10.1371/journal.pone.0093291
- Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van De Peer Y, Coppens F, Vandepoele K** (2018) PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res* **46**: D1190–D1196 doi: 10.1093/nar/gkx1002
- Vanneste S, Friml J** (2013) Calcium: The missing link in auxin action. *Plants* **2**: 650–675 doi: 10.3390/plants2040650
- Williams JH, Edwards JA, Ramsey AJ** (2016) Economy, efficiency, and the evolution of pollen tube growth rates. *Am J Bot* **103**: 471–483 doi: 10.3732/ajb.1500264
- Wilson IBH** (2002) Functional characterization of *Drosophila melanogaster* peptide O-xylosyltransferase, the key enzyme for proteoglycan chain initiation and member of the core 2/I N-acetylglucosaminyltransferase family. *J Biol Chem* **227**: 21207–21212 doi: 10.1074/jbc.M201634200
- Wu Y, Williams M, Bernard S, Driouich A, Showalter AM, Faik A** (2010) Functional identification of two nonredundant Arabidopsis α (1,2)fucosyltransferases specific to arabinogalactan proteins. *J Biol Chem* **285**: 13638–13645 doi: 10.1074/jbc.M110.102715
- Ye CY, Li T, Tuskan GA, Tschaplinski TJ, Yang X** (2011) Comparative analysis of GT14/GT14-like gene family in Arabidopsis, Oryza, Populus, Sorghum and Vitis. *Plant Sci* **181**: 688–695 doi: 10.1016/j.plantsci.2011.01.021
- Yu LC, Twu YC, Chang CY, Lin M** (2001) Molecular basis of the adult i phenotype and the gene responsible for the expression of the human blood group I antigen. *Blood* **98**: 3840–3845 doi: 10.1182/blood.V98.13.3840
- Zhou Y, Li S, Qian Q, Zeng D, Zhang M, Guo L, Liu X, Zhang B, Deng L, Liu X, et al** (2009) BC10, a DUF266-containing and Golgi-located type II membrane protein, is required for cell-wall biosynthesis in rice (*Oryza sativa* L.). *Plant J* **57**: 446–462 doi: 10.1111/j.1365-3113.2008.03703.x

APPENDIX

Supplementary table 1 - List of primers used to genotype T-DNA insertion lines.

*, primers sequences from Knoch et al. (2013); **, primers sequences from Hernández (2018).

Mutant lines	Primers	Sequence (5' - 3')	Product size
<i>fla5-1</i>	1. SALK_139382 LP	CTGTTCTTTGTTTCTGCCTCG	1215 bp (1+2)
	2. SALK_139382 RP	ATTGCAGCCATCGACAAATAC	860 bp (2+LBb1.3)
<i>fla5-2</i>	3. SALK_024772 LP	AAGCAAACCGATGCATAAGTG	1189 bp (3+4)
	4. SALK_024772 RP	CTTTTAAACCCACTCTTCGCC	910 bp (3+LBb1.3)
<i>glcat14a-2</i>	5. SALK_043905 LP	ATTGGTTCAATCTTCGCTTTG*	1122 bp (5+6)
	6. SALK_043905 RP	TCAACCAATGAGAAATGGAGC*	586 bp (6+LBb1.3)
<i>glcat14b-1</i>	7. SALK_080923 LP	AGTGACCATCAATAGGCCCTC**	1166 bp (7+8)
	8. SALK_080923 RP	TGATGTTTTCTGGGTTTCACAG**	882 bp (8+LBb1.3)
<i>glcat14d-1</i>	9. GK-363F05 LP	TTCGTGTCTTGACAACCTATTTACTC**	1355 bp (9+10)
	10. GK-363F05 RP	TCAGTGTAACATTTTCTCTTGTTTC**	600 bp (10+o8409)
<i>glcat14d-2</i>	11. GK-508D01 LP	TTCGTGTCTTGACAACCTATTTACTC**	1355 bp (11+12)
	12. GK-508D01 RP	TCAGTGTAACATTTTCTCTTGTTTC**	1000 bp (12+o8409)
<i>glcat14e-1</i>	13. SALK_022820 LP	AGAGTCTAAAGTTGCTAGACAAACCC**	1290 bp (13+14)
	14. SALK_022820 RP	TATCATTCTCCGTTTAAGTAATTCC**	529 bp (14+LBb1.3)
<i>at4g27480-1</i>	15. SALK_105434C LP	TTATCCAAGGCAGGATCATTG	1150 bp (15+16)
	16. SALK_105434C RP	TCTATCGACCACCAACGAGAC	567 bp (16+LBb1.3)
T-DNA SALK LB	LBb1.3	ATTTTGCCGATTTTCGGAAC	
T-DNA SAIL LB	LB1	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC	
T-DNA GK LB	o8409	ATATTGACCATCATACTCATTGC	

Supplementary table 2 - List of primers used in gene expression analysis.

Gene	Primer	Sequence (5' - 3')	Product size
ACT8	ACT8_fw	CTCAGGTATTGCAGACCGTATGAG	292 bp (gDNA)
	ACT8_rv	CTGGACCTGCTTCATCATACTCTG	185 bp (cDNA)
RUB1	RUB1_fw	CTGTTACACGGAACCCAATTC	133 bp
	RUB1_rv	GGAAAAAGGTCTGACCGACA	
FLA5	qFLA5_fw	ACACCAGAGGCTGATTCTCC	258 bp
	qFLA5_rv	ACTTCCAATGGCCGTGTC	
GlcAT14A	qGlcAT14A_fw	GCTTGGATGGCGTTATCTCG	139 bp
	qGlcAT14A_rv	CAGCGTTGCAGAGAACAGTG	
GlcAT14B	qGlcAT14B_fw	GAGTCACGGGATGGTCACTC	117 bp
	qGlcAT14B_rv	CGTTTAGCTCCAAGACCAGG	
GlcAT14C	qGlcAT14C_fw	CTCGAGTTTTGCATTTGGGG	194 bp
	qGlcAT14C_rv	GTTACATTCAAGGTGCGCTG	
GlcAT14D	qGlcAT14D_fw	CAACTATCCTCTGCAACACGG	192 bp
	qGlcAT14D_rv	CGATGCGATCAAGGACAGTG	
GlcAT14E	qGlcAT14E_fw	GCTGGCCTCTCGGATTAATAAC	120 bp
	qGlcAT14E_rv	GCGTGAAGAGTATTTGCCAC	
At4g27480	qAt4g27480_fw	CATCCTCCTCCAGTACAGCC	182 bp
	qAt4g27480_rv	CTCACACGTTTAGCCAGCTC	
At1g71070	qAt1g71070_fw	CGCCTTGGATTGTACTAAGC	125 bp
	qAt1g71070_rv	CGTGTGGAAGTAGCATTC	
At1g53100	qAt1g53100_fw	CACCGGTTGATGAGAGATCAG	134 bp
	qAt1g53100_rv	CGTATTCGCCACCATTGTTG	
At1g03520	qAt1g03520_fw	GTAATGCATTTGGGGATGGG	179 bp
	qAt1g03520_rv	GGTGTGCTTTGGAGGACTG	
At3g03690	qAt3g03690_fw	GTGAACCACGACCTCCATTAC	189 bp
	qAt3g03690_rv	CGACCCAAGTTTCCGTTTCC	
At4g03340	qAt4g03340_fw	CCAAAGCAACACCCAAACTC	174 bp
	qAt4g03340_rv	CGAGCTACCAATGCACCAAG	
ARP9	ARP9_fw	GAGTTCTTACGCGATACCTCCA	180 bp
	ARP9_rv	GACCACCTTTATTAACCCCATTTACCA	
CYCLOPHILIN	Cyclophilin_fw	TGGCGAACGCTGGTCCTAATACA	223 bp
	Cyclophilin_rv	CAAAAACCTCTGCCCCAATCAA	
EF1B α-subunit 1	EF1B_fw	AAGAACTCACAATCATGATGACAA	206 bp
	EF1B_rv	AAGAAGAGACTTTAGCGAAACACAA	
TUA2	TUA2_fw	ATGTGGGTCAGGGTATGGAA	143 bp
	TUA2_rv	CCGACAACCTTCTTAGTCTCCTCT	

Supplementary table 3 - List of primers used in *FLA5* promoter cloning.

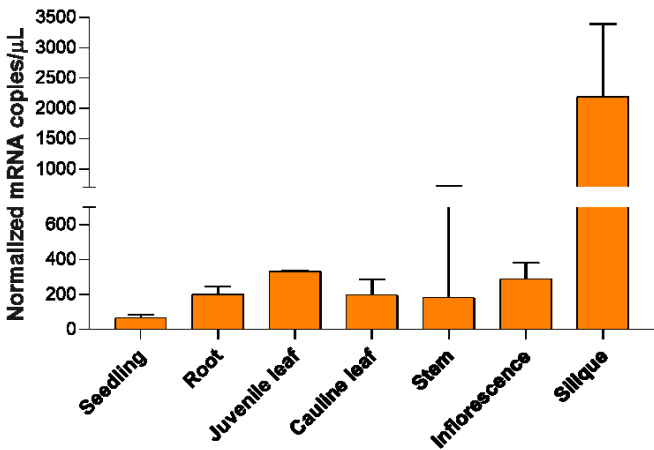
Gene	Primer	Sequence (5' - 3')	Product size
<i>FLA5</i>	pFLA5_fw	GGGGACAAGTTTGTACAAAAAAGCAGGCTTAAGGGTGGAAACGTGGAACA	2271 bp
	pFLA5_rv	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCGTGTCTGTGTTCTCTT	
	AtP_1487	GTGGCTGTTGTAGTTGTACTCC	

Supplementary table 4 - List of primers used in GT14 promoter cloning.

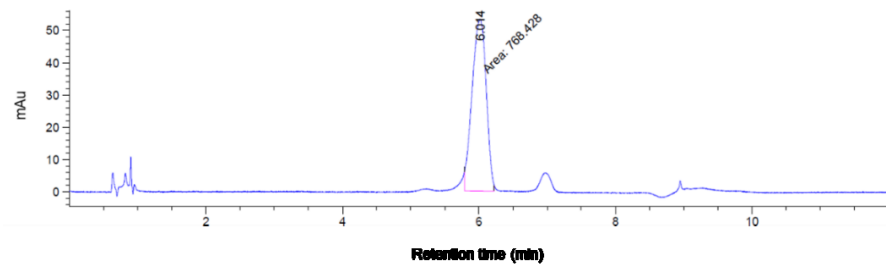
Gene	Primer	Sequence (5' - 3')	Product size
<i>GlcAT14B</i>	pGlcAT14B_fw	CGATAGAGGCCTTTGGTTTG	723 bp
	pGlcAT14B_rv	ATTCGAAGCTTCATCAATTTTAAAC	
<i>GlcAT14E</i>	pGlcAT14E_fw	AGTCCGAAAACATAGCGACA	2466 bp
	pGlcAT14E_rv	TGCTGAAAAC TAGTGATCATGAGAC	
<i>At4g27480</i>	pAt4g27480_fw	CGGAGAGAGTGGAAGAACGA	1489 bp
	pAt4g27480_rv	TAAAGAGCCTTCTCATATGTGATC	
<i>At1g71070</i>	pAt1g71070_fw	CAACCAGCTCCCAACAAC TTT	834 bp
	pAt1g71070_rv	TGTTCTTCACTTCTTCTTCTTTCT	
<i>At1g53100</i>	pAt1g53100_fw	TTGTTTCATGCGTTCACATTTT	1700 bp
	pAt1g53100_rv	GACACCAAAAAC TGAATCAAAATCC	
<i>AT1G03520</i>	pAt1g03520_fw	TGTTCTTCTCTTGTACCACC	941 bp
	pAt1g03520_rv	TCCCTAACCACTTATCCTATCCAAC	
	M13_rv	CCAGGAAACAGCTATGAC	
	MT486_fw	ATGGACGAACGGATAAACC	
	M13_fw	TGTAAAACGACGGCCAGT	

Supplementary table 5 - List of primers used in GT14 coding sequence cloning.

Gene	Primer	Sequence (5' - 3')	Product size
<i>GlcAT14B</i>	GlcAT14B_CDS_fw	ATGAAGAAATTGAAAAGCTATTACA	1632 bp
	GlcAT14B_CDS_rv	CCGGCATTGCCGTGGTCTGAAATT	
<i>GlcAT14E</i>	GlcAT14E_CDS_fw	ATGGGGTATGTAAACGTGGAGAAG	1630 bp
	GlcAT14E_CDS_rv	TTTACACTGACTAACTCCTGTATTA	
<i>At4g27480</i>	At4g27480_CDS_fw	ATGGCGTCCTCAAACCTCAGAG	1480 bp
	At4g27480_CDS_rv	TCTACATTGCCTTTGAGTCAATT	
<i>At1g71070</i>	At1g71070_CDS_fw	ATGGGAGCTGAAAAGAAATGGC	1710 bp
	At1g71070_CDS_rv	CTTGCAATTGATTGATTGTGAA	
<i>At1g53100</i>	At1g53100_CDS_fw	ATGCTTTTGTCTCAAGATCACAA	1512 bp
	At1g53100_CDS_rv	TACACATTGATTCTTACCTAATTT	
<i>AT1G03520</i>	At1g03520_CDS_fw	ATGAGAAAATTGTATAGGAAGCTGG	1603 bp
	At1g03520_CDS_rv	AGAACACTGTTTTTCGCCTAAATTCT	
	M13_rv	CCAGGAAACAGCTATGAC	
	pMDC83_seq	TCTAGAGGATCCCCGGGT	



Supplementary fig. 1 - Normalized levels of *FLA5* gene transcripts (arbitrary units) in a range of tissues. The final concentrations of mRNAs are expressed as arbitrary units that represent the numbers of copies per microliter of cDNA, normalized against the geometric means of three control genes that vary the least with respect to each other. Means of three biological replicates are shown and error bars indicate SDs for each mRNA.



Supplementary fig. 2 - Chromatogram of *FLA5* amplified fragment by qPCR and purified by HPLC. HPLC and chromatography were performed as described by Burton et al. (2004).

	1	10	20	30	40	50	60
AT4G31370 fragment sequence	ACACCAGAGGCTGATTCTCCAGCTCCTGCTCCAAGCGCCGATAATGAGAAGATTGAAGCT						
Sequencing result 1	-----CT						
Sequencing result 2	ACACCAGAGGCTGATTCTCCAGCTCCTGCTCCAAGCGCCGATAATGAGAAGATTGAAGCT						
AT4G31370 fragment sequence	GCTGATAAGGCTAAACCATCTTCTCTGCTTCTAAGGCTGGTTGGAGCTTTGATGTTATT						
Sequencing result 1	GCTGATAAGGCTAAACCATCTTCTCTGCTTCTAAGGCTGGTTGGAGCTTTGATGTTATT						
Sequencing result 2	GCTGATAAGGCTAAACCATCTTCTCTGCTTCTAAGGCTGGTTGGAGCTTTGATGTTATT						
AT4G31370 fragment sequence	CTCTTGTTAGCATTTTTGGCTAGCTTTGCTGGGTTCTAAGATTGCGGTGATGTGGGTTAA						
Sequencing result 1	CTCTTGTTAGCATTTTTGGCTAGCTTTGCTGGGTTCTAAGATTGCGGTGATGTGGGTTAA						
Sequencing result 2	CTCTTGTTAGCATTTTTGGCTAGCTTTGCTGGGTTCTAAGATTGCGGTGATGTGGGTTAA						
AT4G31370 fragment sequence	AGACAATAATGGTAGGTGAGATTCAAACAATTTTCTGTAGCTTTTAGCCAAGAAGAGA						
Sequencing result 1	AGACAATAATGGTAGGTGAGATTCAAACAATTTTCTGTAGCTTTTAGCCAAGAAGAGA						
Sequencing result 2	AGACAATAATGGTAGGTGAGAT-----						
AT4G31370 fragment sequence	GACACGGCCATTGGAAGT						
Sequencing result 1	GACACGGCCATTGGAAGT						
Sequencing result 2	-----						

Supplementary fig. 3 - Alignment of the sequencing results of *FLA5* amplified fragment by qPCR. The two sequencing results (1 and 2) with the primers qFLA5_fw and qFLA5_rv (Supplementary table 2) were aligned with the expected *FLA5* sequence. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

```

FLA5prom:GUS      GGCGGCGGCACTAGTGATATCACAAGTTTGTACAAAAAGCAGGCTTAAGGGTGGAAAGT
Sequencing result 1 -----TA
Sequencing result 2 -----

FLA5prom:GUS      GGAACAATAGATATGTTAAATCTATCTTCATATGCATAATCGATATTCAGGAAAAACTA
Sequencing result 1 TCTCTAATAGATATGTTAAATCTATCTTCATATGCATAATCGATATTCAGGAAAAACTA
Sequencing result 2 -----

FLA5prom:GUS      AAACATACATATATATATGTATCAAGATCGGAGAAGCAGAAGTAGCTATAGCTAGGAACC
Sequencing result 1 AAACATACATATATATGTATCAAGATCGGAGAAGCAGAAGTAGCTATAGCTAGGAACC
Sequencing result 2 -----

FLA5prom:GUS      AAATATTTCCCCCAAGTTGTAATGATCATGATGGAJATTTTAACTTGACTTCGCTA
Sequencing result 1 AAATATTTCCCCCAAGTTGTAATGATCATGATGGAJATTTTAACTTGACTTCGCTA
Sequencing result 2 -----

FLA5prom:GUS      TAATTGGGTGATTTATAATCAGACTCAAAACCGATTGTATTCTTTGCATATTATTTCAT
Sequencing result 1 TAATTGGGTGATTTATAATCAGACTCAAAACCGATTGTATTCTTTGCATATTATTTCAT
Sequencing result 2 -----

FLA5prom:GUS      TTCAGAAAGTTAATAAAATTTACACAATAACGTATAAGAAATAGGAAGATATATACGT
Sequencing result 1 TTCAGAAAGTTAATAAAATTTACACAATAACGTATAAGAAATAGGAAGATATATACGT
Sequencing result 2 -----

FLA5prom:GUS      TTAGTGGTATCTTATTGCGAAAGATATGCTCGAGTAGCCAAATATAATGAAGATCGATT
Sequencing result 1 TTAGTGGTATCTTATTGCGAAAGATATGCTCGAGTAGCCAAATATAATGAAGATCGATT
Sequencing result 2 -----

FLA5prom:GUS      GGTGCCAGGTCACCCCTATCTCTACAACCTTAGAATAATCTGTCAAGAAGATGACATC
Sequencing result 1 GGTGCCAGGTCACCCCTATCTCTACAACCTTAGAATAATCTGTCAAGAAGATGACATC
Sequencing result 2 -----

FLA5prom:GUS      TCGTCTAATTTTAGACCCCTAACATTACTAGTATTGGTAATTATATAAATATTTTATT
Sequencing result 1 TCGTCTAATTTTAGACCCCTAACATTACTAGTATTGGTAATTATATAAATATTTTATT
Sequencing result 2 -----

FLA5prom:GUS      ACATTATATAGCGAATTGATTCCAAATCACTATATGATCATGATCTGTTTCCTTTTA
Sequencing result 1 ACATTATATAGCGAATTGATTCCAAATCACTATATGATCATGATCTGTTTCCTTTTA
Sequencing result 2 -----

FLA5prom:GUS      TTCATTAGTTCATATATAAGTATACCTTCGTGTTGTTTGTACTGTTTATAATTCAA
Sequencing result 1 TTCATTAGTTCATATATAAGTATACCTTCGTGTTGTTTGTACTGTTTATAATTCAA
Sequencing result 2 -----

FLA5prom:GUS      AAGCATAATTACGTAGGCATGTAGCCATTTTGGGTCAATCTCGTTAGTGAACAT
Sequencing result 1 AAGCATAATTACGTAGGCATGTAGCCATTTTGGGTCAATCTCGTTAGTGAACAT
Sequencing result 2 -----

FLA5prom:GUS      TATTTTCGTCACTTTGTTCTAAAAGTATCTTAAGACTATAATATTACGATTAGTTGTT
Sequencing result 1 TATTTTCGTCACTTTGTTCTAAAAGTATCTTAAGACTATAATATTACGATTAGTTGTT
Sequencing result 2 -----

FLA5prom:GUS      TTAAAATAGAGATAAAGGTCAAATGAGCCGAACTCTTAATTATAGAAATGAAACGG
Sequencing result 1 TTAAAATAGAGATAAAGGTCAAATGAGCCGAACTCTTAATTATAGAAATGAAACGG
Sequencing result 2 -----

FLA5prom:GUS      TGTAAAAATATTATTAGTTGTTGTTATTCGTATTTCGTAGGTTAACAAAGAGATATGAA
Sequencing result 1 TGTAAAAATATTATTAGTTGTTGTTATTCGTATTTCGTAGGTTAACAAAGAGATATGAA
Sequencing result 2 -----

FLA5prom:GUS      CTCGAGTTGCCAATTCGAATAATTGGTATAAAGCTGAAACTCTTAAATATTGTTTCGTT
Sequencing result 1 CTCGAGTTGCCAATTCGAATAATTGGTATAAAGCTGAAACTCTTAAATATTGTTTCGTT
Sequencing result 2 -----

FLA5prom:GUS      TCTACATATTATTATACATATGAGTCATGAGTGCTTAAAGACCACTCGAACATATAA
Sequencing result 1 TCTACATATTATTATACATATGAGTCATGAGTGCTTAAAGACCACTCGAACATATAA
Sequencing result 2 -----

FLA5prom:GUS      TCGACAAATGAAACGTTGTTATACAAATTTAGATAGAAATATAATAGTAAGATATAATT
Sequencing result 1 TCGACAAATGAAACGTTGTTATACAAATTTAGATAGAAATATAATAGTAAGATATAATT
Sequencing result 2 TCGCCAAATGAAAC--CTTTTCCATTTGGGTGGAAAT--TTATAGTAAGAAATTTAATG

FLA5prom:GUS      GATTATACATTCTTTTTGGATGATAATTGATACCTAAGAAGCTTCTTATAAATGCGATA
Sequencing result 1 GATTATACATTCTTTTTGGATGATAATTGATACCTAAGAAGCTTCTTATAAATGCGATA
Sequencing result 2 GATTAAACCATTCTTTTTGGATGATAATTGATACCTAAGAAGCTTCTTATAAATGCGATA

FLA5prom:GUS      TCAACTGTGACAATCTCTTTGACGTAATCCGTCGTTACATGGTTTATATATGATATATA
Sequencing result 1 TCAACTGTGACAATCTCTTTGACGTAATCCGTCGTTACATGGTTTATATATGATATATA
Sequencing result 2 TCAACTGTGACAATCTCTTTGACGTAATCCGTCGTTACATGGTTTATATATGATATATA

FLA5prom:GUS      CTGAAATTATACAAATTGACGTTCACTACTTTTTTTTTTTTTTTTTTTTGTGATTCAC
Sequencing result 1 CTGAAATTATACAAATTGACGTTCACTACTTTTTTTTTTTTTTTTTTTTGTGATTCAC
Sequencing result 2 CTGAAATTATACAAATTGACGTTCACTACT--TTTTTTTTTTTTTTTTTTGTGATTCAC

FLA5prom:GUS      TCTTAAATTTTATTACCTTTTAAAGAACATTTATCATTATTGCAAACTCCAACTTAAT
Sequencing result 1 TCTTAAATTTTATTACCTTTTAAAGAACATTTATCATTATTGCAAACTCCAACTTAAT
Sequencing result 2 TCTTAAATTTTATTACCTTTTAAAGAACATTTATCATTATTGCAAACTCCAACTTAAT

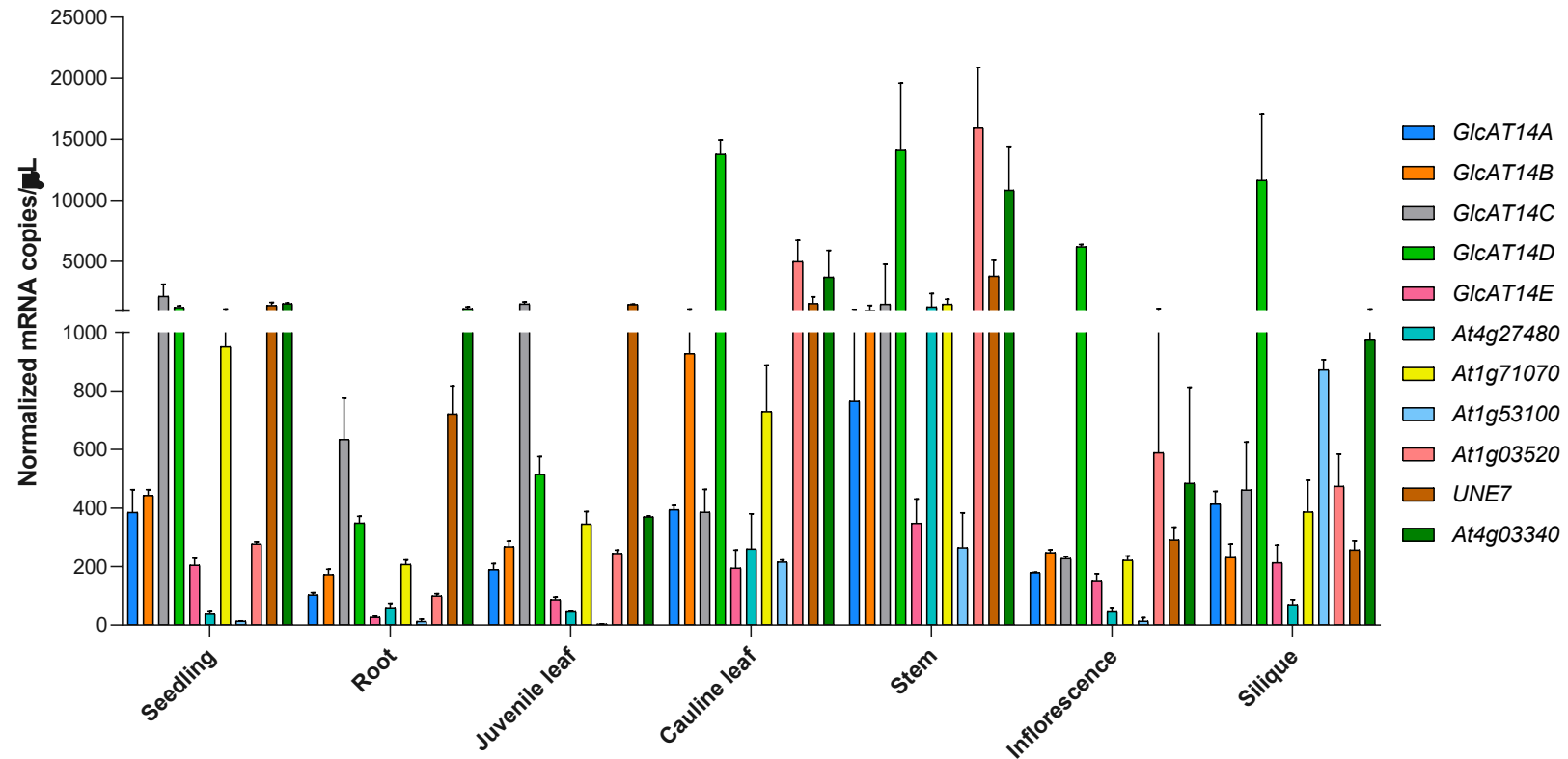
```

Supplementary fig. 4 - Alignment of the sequencing results of *FLA5prom:GUS* construct.

The two sequencing results (1 and 2) with the primers pFLA5_fw and pFLA5_rv (Supplementary table 3) were aligned with the expected *pBGWFS7:FLA5prom* sequence. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

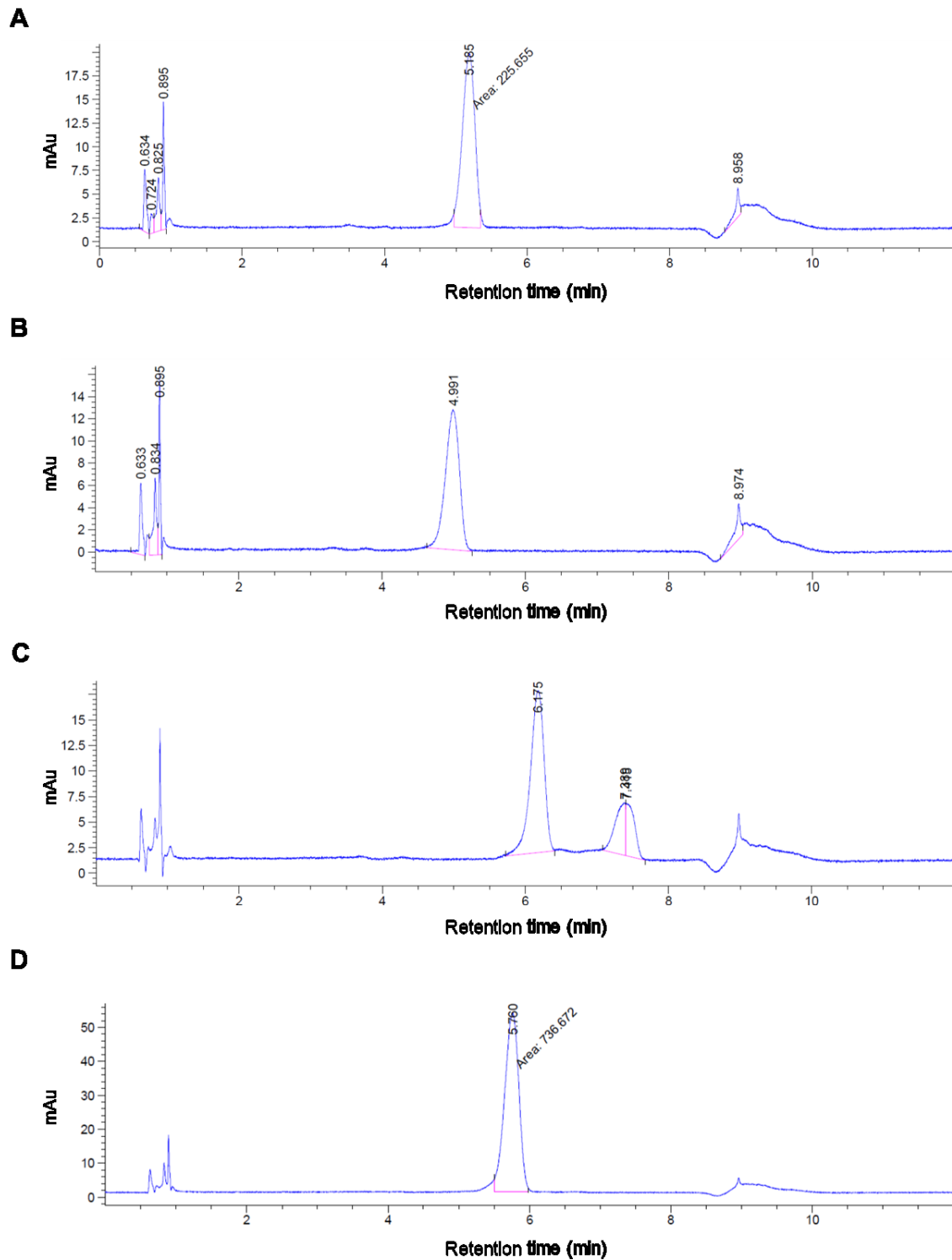
Supplementary fig. 5 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from *fla5-1* (A) and *fla5-2* (B) mutants.

The sequencing results with the primer LBb1.3 (Supplementary table 1) were aligned with the expected *FLA5* flanking sequence. The alignment was performed with Geneious Prime (<https://www.geneious.com>).



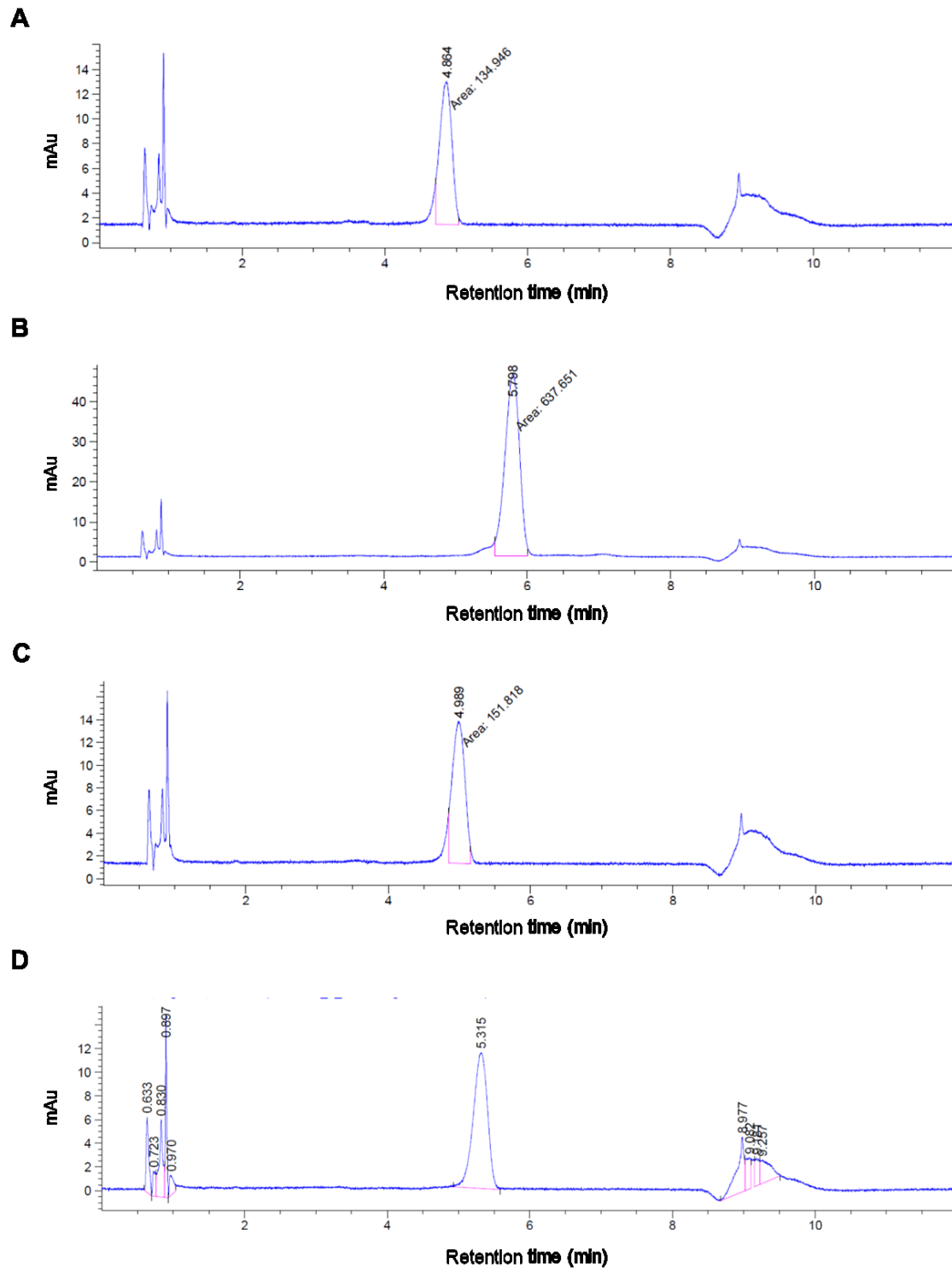
Supplementary fig. 6 - Tissue-specific absolute expression of GT14 genes in Arabidopsis obtained by qPCR.

The analysis includes *GlcAT14A*, *GlcAT14B*, *GlcAT14C*, *GlcAT14D*, *GlcAT14E*, *At4g27480*, *At1g71070*, *At1g53100*, *At1g03520*, *UNE7* and *At4g03340*. The final concentrations of mRNAs are expressed as arbitrary units that represent the numbers of copies per microliter of cDNA, normalized against the geometric means of three control genes that vary the least with respect to each other. Means of three biological replicates are shown and error bars indicate SDs for each mRNA.



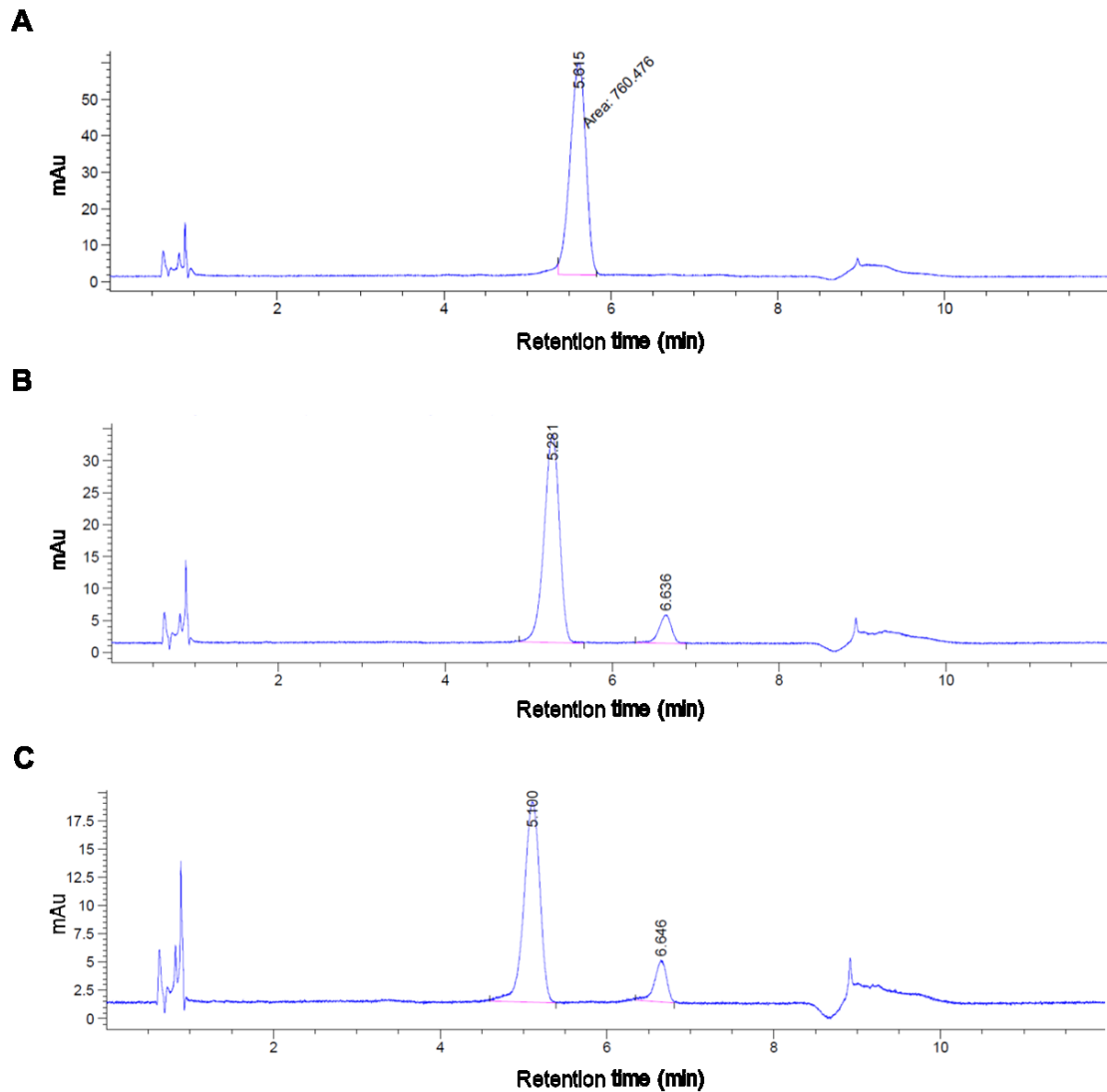
Supplementary fig. 7 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.

HPLC and chromatography of *GlcAT14A* (A), *GlcAT14B* (B), *GlcAT14C* (C) and *GlcAT14D* (D) were performed as described by Burton et al. (2004).



Supplementary fig. 8 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.

HPLC and chromatography of *GlcAT14E* (A), *At4g27480* (B), *At1g71070* (C) and *At1g53100* (D) were performed as described by Burton et al. (2004).



Supplementary fig. 9 - Chromatogram of GT14 amplified fragments by qPCR and purified by HPLC.

HPLC and chromatography of *At1g03520* (A), *UNE7* (B) and *At4g03340* (C) were performed as described by Burton et al. (2004).

A		1	10	20	30	40	50	60
AT5G39990	fragment sequence	GCTTGGATGGCGTTATCTCGGCCATTTCGTCGACTATTGTATAATGGGTTGGGATAATCTA						
	Sequencing result 1	GCTTGGATGGCGTTATCTCGGCCATTTCGTCGACTATTGTATAATGGGTTGGGATAATCTA						
	Sequencing result 2	-----CTA						
AT5G39990	fragment sequence	CCTCGTACGGTCTTAATGTATTACTCGAATTTTCTCTCTTCTCCAGAAGGATATTTTCAC						
	Sequencing result 1	CCTCGTACGGTCTTAATGTATTACTCGAATTTTCTCTCTTCTCCAGAAGGATATTTTCAC						
	Sequencing result 2	CCTCGTACGGTCTTAATGTATTACTCGAATTTTCTCTCTTCTCCAGAAGGATATTTTCAC						
AT5G39990	fragment sequence	ACTGTTCTCTGCAACGC						
	Sequencing result 1	-----						
	Sequencing result 2	ACTGTTCTCTGCAACGC						
B		1	10	20	30	40	50	60
AT5G15050	fragment sequence	GAGTCACGGGATGGTCACTCCGGGTGGTGGTGTATTGGTACGAGGGAAAAACGGGAGCGA						
	Sequencing result 1	-----AACGGGAGCGA						
	Sequencing result 2	GAGTCACGGGATGGTCACTCCGGGTGGTGGTGTATTGGTACGAGGGAAAAACGGGAGCGA						
AT5G15050	fragment sequence	TCCGTGTGCGGTGATTGGTGATACGTCGGTTATTAACCTGGTCTTGGAGCTAAACG						
	Sequencing result 1	TCCGTGTGCGGTGATTGGTGATACGTCGGTTATTAACCTGGTCTTGGAGCTAAACG						
	Sequencing result 2	TCCGTGTGCGGTGAT-----						
C		1	10	20	30	40	50	60
AT2G37585	fragment sequence	CTCGAGTTTTCGATTGGGGATGGGATAATCTCCCAAGAACGCTATTGATGTACTATACC						
	Sequencing result 1	-----CGCTATTGATGTACTATACC						
	Sequencing result 2	CTCGAGTTTTCGATTGGGGATGGGATAATCTCCCAAGAACGCTATTGATGTACTATACC						
AT2G37585	fragment sequence	AACTTCCTCTTGTCTATCTGAAGGCTACTTTCAGACAGTTGTGTGCAACAATAAAGATTAT						
	Sequencing result 1	AACTTCCTCTTGTCTATCTGAAGGCTACTTTCAGACAGTTGTGTGCAACAATAAAGATTAT						
	Sequencing result 2	AACTTCCTCTTGTCTATCTGAAGGCTACTTTCAGACAGTTGTGTGCAACAATAAAGATTAT						
AT2G37585	fragment sequence	CAGAACAACAACGGTTAACACGACCTGCATTACACAAAATGGGATCCGCTTCAACAGCGC						
	Sequencing result 1	CAGAACAACAACGGTTAACACGACCTGCATTACACAAAATGGGATCCGCTTCAACAGCGC						
	Sequencing result 2	CAGAACAACAACGGT-----						
AT2G37585	fragment sequence	ACCCTGAATGTAAC						
	Sequencing result 1	ACCCTGAATGTAAC						
	Sequencing result 2	-----						
D		1	10	20	30	40	50	60
AT3G24040	fragment sequence	CAACTATCCTCTGCAACACGGATACCTTCAAAAAGACCATCATGAACAACAACTTACTCT						
	Sequencing result 1	CAACTATCCTCTGCAACACGGATACCTTCAAAAAGACCATCATGAACAACAACTTACTCT						
	Sequencing result 2	-----CTTACTCT						
AT3G24040	fragment sequence	ACCTAGCTTCTAACGACACTTCTAAAGAAAGATACCATCAGCTTGATCACAAGAGTTCA						
	Sequencing result 1	ACCTAGCTTCTAACGACACTTCTAAAGAAAGATACCATCAGCTTGATCACAAGAGTTCA						
	Sequencing result 2	ACCTAGCTTCTAACGACACTTCTAAAGAAAGATACCATCAGCTTGATCACAAGAGTTCA						
AT3G24040	fragment sequence	CAGAAATGGTTGACTCCGGAGCAGCTTTTGCAAGAGGATTACAGTACGACGACTGTCC						
	Sequencing result 1	CAGAAA-----						
	Sequencing result 2	CAGAAATGGTTGACTCCGGAGCAGCTTTTGCAAGAGGATTACAGTACGACGACTGTCC						
AT3G24040	fragment sequence	TTGATCGCATCG						
	Sequencing result 1	-----						
	Sequencing result 2	TTGATCGCATCG						

Supplementary fig. 10 - Alignment of the sequencing results of *At5g39990* (A), *At5g15050* (B), *At2g37585* (C) and *At3g24040* (D) amplified fragments by qPCR.

The two sequencing results (1 and 2) with the primers qAtGlcAT14A_fw and qAtGlcAT14A_rv, qAtGlcAT14B_fw and qAtGlcAT14B_rv, qATGLCAT14C_fw and qATGLCAT14C_rv, qAtGlcAT14D_fw and qAtGlcAT14D_rv (Supplementary table 2) were aligned with the respective expected of *At5g39990* (A), *At5g15050* (B), *At2g37585* (C) and *At3g24040* (D) sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

A		1	10	20	30	40	50	60
AT3G15350 fragment sequence		GCTGGCCTCTCGGATTAATAACGATCCTATGTACTCTAAAACCGGGATGTATATATGATA						
Sequencing result 1		GCTGGCCTCTCGGATTAATAACGATCCTATGTACTCTAAAACCGGGATGTATATATGATA						
Sequencing result 2		-----TATATATGATA						
AT3G15350 fragment sequence		ACCAAGGCTAACCTTGTGACATATAAAGGACCAACGATGGTGGCAAATACTCTTCACGC						
Sequencing result 1		ACCAAGGCTAACCTTGT-----						
Sequencing result 2		ACCAAGGCTAACCTTGTGACATATAAAGGACCAACGATGGTGGCAAATACTCTTCACGC						
B		1	10	20	30	40	50	60
AT4G27480 fragment sequence		CATCCTCCTCCAGTACAGCCTAGTCTTCTCGCTTGGTTATCTAGTTTCAGGGTCAAGA						
Sequencing result 1		-ATCCTCCTCCAGTACAGCCTAGTCTTCTCGCTTGGTTATCTAGTTTCAGGGTCAAGA						
Sequencing result 2		-----TCAGA						
AT4G27480 fragment sequence		GGTGACTTAGAGAGTCTGTGGAGAGTGTACGAACATTATACCATCCTAGGAATCAGTAC						
Sequencing result 1		GGTGACTTAGAGAGTCTGTGGAGAGTGTACGAACATTATACCATCCTAGGAATCAGTAC						
Sequencing result 2		GGTGACTTAGAGAGTCTGTGGAGAGTGTACGAACATTATACCATCCTAGGAATCAGTAC						
AT4G27480 fragment sequence		GTTGTTTCATCTTGATCTCGAGTCTCCTGCTGAAGAGAGGCTTGAGCTGGCTAAACGTGTG						
Sequencing result 1		GTTGTTTCATCTTGATCTCGA-----						
Sequencing result 2		GTTGTTTCATCTTGATCTCGAGTCTCCTGCTGAAGAGAGGCTTGAGCTGGCTAAACGTGTG						
AT4G27480 fragment sequence		AG						
Sequencing result 1		--						
Sequencing result 2		AG						
C		1	10	20	30	40	50	60
AT1G71070 fragment sequence		CGCCTTGGATTGTACTAAGCAGACCTTCTAGAACTACTGCATCTTCGGTTGGGATAATT						
Sequencing result 1		CGCCTTGGATTGTACTAAGCAGACCTTCTAGAACTACTGCATCTTCGGTTGGGATAATT						
Sequencing result 2		-----CTTCGGTTGGGATAATT						
AT1G71070 fragment sequence		TACCAAGAATCCTTCTCATGTATTTTAACAATGTCATTCTCTCTGAAGAATGCTACTTCC						
Sequencing result 1		TACCAAGAAT-----						
Sequencing result 2		TACCAAGAATCCTTCTCATGTATTTTAACAATGTCATTCTCTCTGAAGAATGCTACTTCC						
AT1G71070 fragment sequence		ACAG						
Sequencing result 1		----						
Sequencing result 2		ACAG						
D		1	10	20	30	40	50	60
AT1G53100 fragment sequence		CACC GGTTGATGAGAGATCAGAGTTAGCTTCTCGCATTTCAAACGATCCTATGTATTCAA						
Sequencing result 1		-----AACGATCCTATGTATTCAA						
Sequencing result 2		CACC GGTTGATGAGAGATCAGAGTTAGCTTCTCGCATTTCAAACGATCCTATGTATTCAA						
AT1G53100 fragment sequence		AAGCAGGAAATGTATATATGATAACTAAAGCTAATCTTGTGACTTATACCGGACCAACAA						
Sequencing result 1		AAGCAGGAAATGTATATATGATAACTAAAGCTAATCTTGTGACTTATACCGGACCAACAA						
Sequencing result 2		AAGCAGGAAATGTATATATGATAACTAAAGCTAAT-----						
AT1G53100 fragment sequence		TG GTGGCGAATACG						
Sequencing result 1		TG GTGGCGAATACG						
Sequencing result 2		-----						

Supplementary fig. 11 - Alignment of the sequencing results of *At3g15350* (A), *At4g27480* (B), *At1g71070* (C) and *At1g53100* (D) amplified fragments by qPCR.

The two sequencing results (1 and 2) with the primers qAtGlcAT14E_fw and qAtGlcAT14E_rv, qAt4g27480_fw and qAt4g27480_rv, qAt1g71070_fw and qAt1g71070_rv, qAt1g53100_fw and qAt1g53100_rv (Supplementary table 2) were aligned with the respective expected of *At3g15350* (A), *At4g27480* (B), *At1g71070* (C) and *At1g53100* (D) sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

A		1	10	20	30	40	50	60
AT1G03520	fragment sequence	GTACTGCATTGGGGATGGGATAATTTCCCAAGAACAATACTAATGTACTACACAAATTT						
	Sequencing result 1	GTACTGCATTGGGGATGGGATAATTTCCCAAGAACAATACTAATGTACTACACAAATTT						
	Sequencing result 2	-----						
AT1G03520	fragment sequence	CGTATCTTCTCCTGAAGGATACTTCCACACAGTAATCTGCAACAGCAAGAGTTTATCAA						
	Sequencing result 1	CGTATCTTCTCCTGAAGGATACTTCCACACAGTAATCTGCAACAGCAAGAGTTTATCAA						
	Sequencing result 2	-----CTCCTGAAGGATACTTCCACACAGTAATCTGCAACAGCAAGAGTTTATCAA						
AT1G03520	fragment sequence	CACTGCAATCGGTCACGACCTTCACTACATTGCTTGGGACAGTCCTCCAAAGCAACACC						
	Sequencing result 1	CACTGCAATCG-----						
	Sequencing result 2	CACTGCAATCGGTCACGACCTTCACTACATTGCTTGGGACAGTCCTCCAAAGCAACACC						
B		1	10	20	30	40	50	60
AT3G03690	fragment sequence	GTGAACCACGACCTCCATTACATCGCGTGGGACAATCCGCCAAAGCAGCACCCCAAGATA						
	Sequencing result 1	-----AGCAGCACCCCAAGATA						
	Sequencing result 2	GTGAACCACGACCTCCATTACATCGCGTGGGACAATCCGCCAAAGCAGCACCCCAAGATA						
AT3G03690	fragment sequence	TTAGGGACAAGGGATTACAGAAAGATGGTAATGAGCAACCGACCATTCGCTAGGAAGTTC						
	Sequencing result 1	TTAGGGACAAGGGATTACAGAAAGATGGTAATGAGCAACCGACCATTCGCTAGGAAGTTC						
	Sequencing result 2	TTAGGGACAAGGGATTACAGAAAGATGGTAATGAGCAACCGACCATTCGCTAGGAAGTTC						
AT3G03690	fragment sequence	AAGAGCAATGACCCCGTTCTCAATAGGATAGACCGGGAGATTCTGAGAAGGAAACGGAAA						
	Sequencing result 1	AAGAGCAATGACCCCGTTCTCAATAGGATAGACCGGGAGATTCTGAGAAGGAAACGGAAA						
	Sequencing result 2	AAGAGCAATGACCCCGT-----						
AT3G03690	fragment sequence	CTTGGGTCG						
	Sequencing result 1	CTTGGGTCG						
	Sequencing result 2	-----						
C		1	10	20	30	40	50	60
AT4G03340	fragment sequence	CCAAAGCAACACCCAAACTCTTTATCGATGAAGGATTTGACAAAGATGGTCAAAAGCAAA						
	Sequencing result 1	-----AAAGCAAA						
	Sequencing result 2	CCAAAGCAACACCCAAACTCTTTATCGATGAAGGATTTGACAAAGATGGTCAAAAGCAAA						
AT4G03340	fragment sequence	GCCCCATTGCTCGAAAGTTCACAAGAACGATCCTGTATTAGACAAGATCGATAGAGAG						
	Sequencing result 1	GCCCCATTGCTCGAAAGTTCACAAGAACGATCCTGTATTAGACAAGATCGATAGAGAG						
	Sequencing result 2	GCCCCATTGCTCGAAAGTTCACAAGAACGATCCTGTATTAGACAAGATCGATAGAGAG						
AT4G03340	fragment sequence	CTCTTAGGCGCACTCATCGGTTTTCTCAGGAGCTTGGTGCAATGGTAGCTCG						
	Sequencing result 1	CTCTTAGGCGCACTCATCGGTTTTCTCAGGAGCTTGGTGCAATGGTAGCTCG						
	Sequencing result 2	CTCT-----						

Supplementary fig. 12 - Alignment of the sequencing results of *At1g03520* (A), *At3g03690* (B) and *At4g03340* (C) amplified fragments by qPCR.

The two sequencing results (1 and 2) with the primers qAt1g03520_fw and qAt1g03520_rv, qAt3g03690_fw and qAt3g03690_rv, qAt4g03340_fw and qAt4g03340_rv (Supplementary table 2) were aligned with the respective expected of *At1g03520* (A), *At3g03690* (B) and *At4g03340* (C) sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

A			B			C					
MT589:ipAtg1cAt14B	1381	GCACAGCTCTAGTAAATTAAGAAATTAACAAAGTTTGTACAAAAAGAGGCTCCGAAAT	1440	721	ATTATCACAAAGTTTGTACAAAAAGAGGCTCCGAAATTCGCCCTTAGTCCGAAACATA	780	pMT589:ipAtg27480	121	TCTAGTTAATTAAGAAATTAACAAAGTTTGTACAAAAAGAGGCTCCGAAATTCGCCCT	180	
Sequencing result	1	-----TTGTACAAAAAGAGGCTCCGAAATTCGCCCTTAGTCCGAAACATA	25	Sequencing result	1	-----TTGTACAAAAAGAGGCTCCGAAATTCGCCCTTAGTCCGAAACATA	47	Sequencing result	1	-----TTGTACAAAAAGAGGCTCCGAAATTCGCCCT	32
MT589:ipAtg1cAt14B	1441	TGCGCCCTCGATAGAGGCTTTGGTTGTGCGACATTTGACATCATCCCTTTAGT	1500	781	GGGACAGCTAAATTAATGAAGTTATACTTCAACCTAACACATTAACATATTGAA	840	pMT589:ipAtg27480	181	TGGGAGAGAGTGGAGAGCATATAACGCCCTTAAAGATTTTAGGGTTAGCAAAAAAT	240	
Sequencing result	26	TGCGCCCTCGATAGAGGCTTTGGTTGTGCGACATTTGACATCATCCCTTTAGT	85	Sequencing result	48	GGGACAGCTAAATTAATGAAGTTATACTTCAACCTAACACATTAACATATTGAA	107	Sequencing result	33	TGGGAGAGAGTGGAGAGCATATAACGCCCTTAAAGATTTTAGGGTTAGCAAAAAAT	92
MT589:ipAtg1cAt14B	1501	TGATATTTTGTATGTGGGCTTAAAGTGTTCATTTGGGCTTTCTAGATTTAAAGCCCA	1560	841	TCAGCCACTGAAATTAATGCTTTCTAAATGGAGTACTCTCCCTCCGATTAAGATA	900	pMT589:ipAtg27480	241	AATTAAGAAAAAGGGGAAATTTGTATTTAAGTAACTAAAGTAAATCAACATTTTATT	300	
Sequencing result	86	TGATATTTTGTATGTGGGCTTAAAGTGTTCATTTGGGCTTTCTAGATTTAAAGCCCA	145	Sequencing result	108	TCAGCCACTGAAATTAATGCTTTCTAAATGGAGTACTCTCCCTCCGATTAAGATA	167	Sequencing result	93	AATTAAGAAAAAGGGGAAATTTGTATTTAAGTAACTAAAGTAAATCAACATTTTATT	152
MT589:ipAtg1cAt14B	1561	TTAATCCTTGGCTTTTAAAGGATTAGCATTAAGTGTCTAATAGTACGCTTAATAGTCT	1620	901	AATTTTTAGGAAAAATATTTCGCCAAGAGATAGAGTTTATATTTCCAAATGCAAAAT	960	pMT589:ipAtg27480	301	ACGTAGTTTCAATTACTAGAGAGTTTACTTTGTATGATCAATTTAAATATTGGTATA	360	
Sequencing result	146	TTAATCCTTGGCTTTTAAAGGATTAGCATTAAGTGTCTAATAGTACGCTTAATAGTCT	205	Sequencing result	168	AATTTTTAGGAAAAATATTTCGCCAAGAGATAGAGTTTATATTTCCAAATGCAAAAT	227	Sequencing result	153	ACGTAGTTTCAATTACTAGAGAGTTTACTTTGTATGATCAATTTAAATATTGGTATA	212
MT589:ipAtg1cAt14B	1621	GACAAAGTCATTCACAGTCAATACTTAAATGGGAAAGATTAACCTTGGAGACCGG	1680	961	TTTAAAGCTGTTTCCGCTTTACCTCTAGTAAATATTATATTTTAAAGATA	1020	pMT589:ipAtg27480	361	AATCTTACATTTAAAGAAAAATTTATGATTAATCAATAAAATAGAGTTTGTATCAAA	420	
Sequencing result	206	GACAAAGTCATTCACAGTCAATACTTAAATGGGAAAGATTAACCTTGGAGACCGG	265	Sequencing result	228	TTTAAAGCTGTTTCCGCTTTACCTCTAGTAAATATTATATTTTAAAGATA	287	Sequencing result	213	AATCTTACATTTAAAGAAAAATTTATGATTAATCAATAAAATAGAGTTTGTATCAAA	272
MT589:ipAtg1cAt14B	1681	TTAAATGCTCTTCAATTAATTAAGCCCTTCCATCTACTCTCTAAAGAAAGTGTATGCTTT	1740	1021	TTAAGAGGTTAAATGCTTTTATACAAATTTCTTAAGAAATGCGAAATTTAAAGAAAA	1080	pMT589:ipAtg27480	421	ATAAAAACTACTAGCTTAATAGCTCTAAATGTTTGCACATTAATGAAATTTAGTGAAAT	480	
Sequencing result	266	TTAAATGCTCTTCAATTAATTAAGCCCTTCCATCTACTCTCTAAAGAAAGTGTATGCTTT	325	Sequencing result	288	TTAAGAGGTTAAATGCTTTTATACAAATTTCTTAAGAAATGCGAAATTTAAAGAAAA	347	Sequencing result	273	ATAAAAACTACTAGCTTAATAGCTCTAAATGTTTGCACATTAATGAAATTTAGTGAAAT	332
MT589:ipAtg1cAt14B	1741	CTTAAGCACCATTAATACTTACTTAAAGAAATTAGAGAGATTGAATGACCATTTACT	1800	1081	TTCACTTTAGTGGAGCGAGAGAGATTAATTTATAGAGTCCGAACTTTTCACTAAA	1140	pMT589:ipAtg27480	481	GGATATAGAAAAATATATTTTCCATCAAACTCTTTTACATATTCATACATATAATAT	540	
Sequencing result	326	CTTAAGCACCATTAATACTTACTTAAAGAAATTAGAGAGATTGAATGACCATTTACT	385	Sequencing result	348	TTCACTTTAGTGGAGCGAGAGAGATTAATTTATAGAGTCCGAACTTTTCACTAAA	407	Sequencing result	333	GGATATAGAAAAATATATTTTCCATCAAACTCTTTTACATATTCATACATATAATAT	392
MT589:ipAtg1cAt14B	1801	ATTATTCAACCAATATATTTCTTCAAGAAATGAACCTTTGAGTGGGATAAAATGTACTTT	1860	1141	AATTAATTTAGAGCAAAAAAGGCGACATTTCTCTTTTGTATACATACATATCTTT	1200	pMT589:ipAtg27480	541	AAATCCATACCATTAAGGAAACATATTTACATATTAATCAATAACAAATTAAGT	600	
Sequencing result	386	ATTATTCAACCAATATATTTCTTCAAGAAATGAACCTTTGAGTGGGATAAAATGTACTTT	445	Sequencing result	408	AATTAATTTAGAGCAAAAAAGGCGACATTTCTCTTTTGTATACATACATATCTTT	467	Sequencing result	393	AAATCCATACCATTAAGGAAACATATTTACATATTAATCAATAACAAATTAAGT	452
MT589:ipAtg1cAt14B	1861	AAACTATCAGATAAATACACATGATCTGTATTAATGAATTAATCAATGTCCTCAAGCA	1920	1201	TTGACTTTGAAGAGATTTTTTTTCCACAGATTTTGAAGAGATTAATGACCTCACTTAA	1260	pMT589:ipAtg27480	601	ATAACCAAGATTTGAATGAGGACATACCACTTAAGCTAAATCATCAATATGCCACC	660	
Sequencing result	446	AAACTATCAGATAAATACACATGATCTGTATTAATGAATTAATCAATGTCCTCAAGCA	505	Sequencing result	468	TTGACTTTGAAGAGATTTTTTTTCCACAGATTTTGAAGAGATTAATGACCTCACTTAA	527	Sequencing result	453	ATAACCAAGATTTGAATGAGGACATACCACTTAAGCTAAATCATCAATATGCCACC	512
MT589:ipAtg1cAt14B	1921	TACATAAATCGGATAAAATTTAGAGAAAAAGTAAAAATACAAAAATACCGTGAACCA	1980	1261	GCATGTATATGTCGCCGCACACTGCTTTTGAAGAAAGAGGATGAAAGCTAGTTTG	1320	pMT589:ipAtg27480	661	TCTTAAAAACATATATAGGGTTTCATATGCGCTGATACAGATTTTGGCATTTTC	720	
Sequencing result	506	TACATAAATCGGATAAAATTTAGAGAAAAAGTAAAAATACAAAAATACCGTGAACCA	565	Sequencing result	528	GCATGTATATGTCGCCGCACACTGCTTTTGAAGAAAGAGGATGAAAGCTAGTTTG	587	Sequencing result	513	TCTTAAAAACATATATAGGGTTTCATATGCGCTGATACAGATTTTGGCATTTTC	572
MT589:ipAtg1cAt14B	1981	TTTTTTTTTGTATATAGATATCGACATCGAAGAGAGACAGACCAACCAACAGCA	2040	1321	TTGGGCTTCAGTGCAGCTCGAAGATCACTGTGTAAAGCCATCAATCTCAACAGTGAC	1380	pMT589:ipAtg27480	721	GCATTACTAGGCCACTTTCTCCGATTTTCATTTTATTCGAAATGCAATCATCAAGCTGT	780	
Sequencing result	566	TTTTTTTTTGTATATAGATATCGACATCGAAGAGAGACAGACCAACCAACAGCA	625	Sequencing result	588	TTGGGCTTCAGTGCAGCTCGAAGATCACTGTGTAAAGCCATCAATCTCAACAGTGAC	619	Sequencing result	573	GCATTACTAGGCCACTTTCTCCGATTTTCATTTTATTCGAAATGCAATCATCAAGCTGT	632
MT589:ipAtg1cAt14B	2041	AAACACAGCAATCTCTGATTCATCTCTGATTTGTATTTGTAGTCTCACTCTGAA	2100				pMT589:ipAtg27480	781	CGTGCATATACACAGTTTGTCCACACAGTATACAGTAAATGATCTCCATATATGA	840	
Sequencing result	626	AAACACAGCAATCTCTGATTCATCTCTGATTTGTATTTGTAGTCTCACTCTGAA	685				Sequencing result	633	CGTGCATATACACAGTTTGTCCACACAGTATACAGTAAATGATCTCCATATATGA	692	
MT589:ipAtg1cAt14B	2101	GTCTCTTAAGCCACAACCACTGGAAGAGATCCATTAAGAGCTGTATAAAATGATG	2160				pMT589:ipAtg27480	841	ATAAGAACTACTCTATGTTTTTTAGCTGTATATGAAATTAATGCAATTTAATAGC	900	
Sequencing result	686	GTCTCTTAAGCCACAACCACTGGAAGAGATCCATTAAGAGCTGTATAAAATGATG	732				Sequencing result	693	ATAAGAACTACTCTATGTTTTTTAGCTGTATATGAAATTAATGCAATTTAATAGC	752	

Supplementary fig. 13 - Alignment of the sequencing results of *GlcAT14Bprom:YFP* (A), *GlcAT14Eprom:YFP* (B) and *At4g27480prom:YFP* (C) constructs.

The sequencing results with the primer M13_fw (Supplementary table 4 - List of primers used in GT14 promoter cloning.) were aligned with the respective expected *pMT589:GlcAT14Bprom*, *pMT589:GlcAT14Eprom* and *pMT589:At4g27480prom*, sequences. The alignment was performed using Geneious Prime (<https://www.geneious.com>).

A	MT589: ipAt1g71070	721	ATTATCAACAGGTTTGTACAAAAAGAGGCTCGGAATTCGCCCTTCAACAGCTCCCAA	780
	Sequencing result	1	-----TTTGTACAAAAAGAGGCTCGGAATTCGCCCTTCAACAGCTCCCAA	48
	MT589: ipAt1g71070	781	CAACTTTGTCAACACATATTCATATTTGAGCACTTTGTCTGAATATTTGGTGAAG	840
	Sequencing result	49	CAACTTTGTCAACACATATTCATATTTGAGCACTTTGTCTGAATATTTGGTGAAG	100
	MT589: ipAt1g71070	841	AATCCAAAAGGGTAAAAATTTATGTTTTTACTAGAAAGTAGTAAATCAAAATGATATAA	908
	Sequencing result	109	AATCCAAAAGGGTAAAAATTTATGTTTTTACTAGAAAGTAGTAAATCAAAATGATATAA	168
	MT589: ipAt1g71070	901	GATCGACTACTATTCGAGAGTAGCAATTTTCATTTTAAATCGAGAATCTCGTGT	960
	Sequencing result	169	GATCGACTACTATTCGAGAGTAGCAATTTTCATTTTAAATCGAGAATCTCGTGT	228
	MT589: ipAt1g71070	1021	AAGTAATCTCTCTCTCGACGTGTAAACACGCGACACACACACCTTCACITTTATTT	1080
	Sequencing result	289	AAGTAATCTCTCTCTCGACGTGTAAACACGCGACACACACACCTTCACITTTATTT	1408
B	MT589: ipAt1g53100	601	ACGGCAGATGCCAGCTGTAGTAAATTAAGAAATATCAACAAGTTTGTACAAAAAGCAG	660
	Sequencing result	1	-----TTTGTACAAAAAGCAGCTGTAGTAAATTAAGAAATATCAACAAGTTTGTACAAAAAGCAG	15
	MT589: ipAt1g53100	661	GCTCGGAATCGCCCTTTTGTATGCGTGCACATTCAAATTTATGTTTCTTCTCTC	720
	Sequencing result	16	GCTCGGAATCGCCCTTTTGTATGCGTGCACATTCAAATTTATGTTTCTTCTCTC	75
	MT589: ipAt1g53100	721	TGGAAGTTTTGTTTTTTTTTGGTATGGAATATTTATTAATTTATTTCTAAACATAC	780
	Sequencing result	76	TGGAAGTTTTGTTTTTTTTTGGTATGGAATATTTATTAATTTATTTCTAAACATAC	135
	MT589: ipAt1g53100	781	TATATATACACTTATTAAGTCTAGCTTTGGTTGGTATGCTCTTAAGGTCAACAGT	840
	Sequencing result	136	TATATATACACTTATTAAGTCTAGCTTTGGTTGGTATGCTCTTAAGGTCAACAGT	195
	MT589: ipAt1g53100	841	CAAACTCTATGTATTAATCTGAACATCAAACTACTAATTCGCATTTGGTATGGTAT	900
	Sequencing result	196	CAAACTCTATGTATTAATCTGAACATCAAACTACTAATTCGCATTTGGTATGGTAT	255
C	pMT589: ipAt1g03520	1	GTATTAAGAATATTAACAAGTTTGTACAAAAAGAGGCTCGGAATTCGCCCTT	60
	Sequencing result	1	-----TTTGTACAAAAAGAGGCTCGGAATTCGCCCTT	38
	pMT589: ipAt1g03520	61	TGCTTCTCTCTGTACACCAAAAGTCAAACTGGCTATAGATAGATAGACTATACACT	120
	Sequencing result	59	TGCTTCTCTCTGTACACCAAAAGTCAAACTGGCTATAGATAGATAGACTATACACT	98
	pMT589: ipAt1g03520	121	AGCAGTGAATTAGGAGCAACATAGACAGAGTGTGTTTTTCTCGCGCTTAATTCGT	180
	Sequencing result	99	AGCAGTGAATTAGGAGCAACATAGACAGAGTGTGTTTTTCTCGCGCTTAATTCGT	158
	pMT589: ipAt1g03520	181	TACTTATCAAGCTATTATATAGTGTATATGAGTGTATATGAGTGTATATGAGT	240
	Sequencing result	159	TACTTATCAAGCTATTATATAGTGTATATGAGTGTATATGAGTGTATATGAGT	218
	pMT589: ipAt1g03520	241	GTTTGTAGAGTGTATGATGATCAATTAATCTCTACTAGTATGTTTATTTGTTGCAACA	300
	Sequencing result	219	GTTTGTAGAGTGTATGATGATCAATTAATCTCTACTAGTATGTTTATTTGTTGCAACA	278

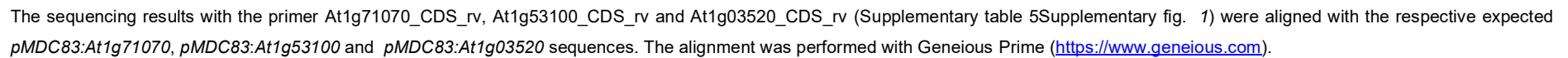
Supplementary fig. 14 - Alignment of the sequencing results of *At1g71070prom:YFP* (A), *At1g53100prom:YFP* (B) and *At1g03520prom:YFP* (C) constructs.

The sequencing results with the primer M13_fw (Supplementary table 4 - List of primers used in GT14 promoter cloning.) were aligned with the respective equivalent *pMT589:At1g71070prom*, *pMT589:At1g53100prom* and *pMT589:At1g03520prom*, sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

A pMDC83::AtGlcAT14B 1441 ACGAAATTTAGATCTTATAATCTCAGACTCTGTTTATGTTTAAATGAATCAITTTGAGG 1500 Sequencing result 1 ACGAAATTTAGATCTTATAATCTCAGACTCTGTTTATGTTTAAATGAATCAITTTGAGG 60 pMDC83::AtGlcAT14B 1501 TTTATTATGTGATGATTTTTCAGATCTGTTGACATGTTTTCGTATCTTCTCGAGAT 1560 Sequencing result 61 TTTATTATGTGATGATTTTTCAGATCTGTTGACATGTTTTCGTATCTTCTCGAGAT 120 pMDC83::AtGlcAT14B 1561 CTCAAATTCATGATCATACAGCAACATTTGCGTGGAAAGGTGATTTAGCTAGCTACA 1620 Sequencing result 121 CTCAAATTCATGATCATACAGCAACATTTGCGTGGAAAGGTGATTTAGCTAGCTACA 180 pMDC83::AtGlcAT14B 1621 TTACAATCTACAGAGCTTTGTTTGGCACTTCTCTGATAATTTTGGTAAAACTTGT 1680 Sequencing result 181 TTACAATCTACAGAGCTTTGTTTGGCACTTCTCTGATAATTTTGGTAAAACTTGT 240 pMDC83::AtGlcAT14B 1681 ATAGGTTCTCATAGCAAGCTCTATATATGATCTGTTTGTATATCTCAAGAAAG 1740 Sequencing result 241 ATAGGTTCTCATAGCAAGCTCTATATATGATCTGTTTGTATATCTCAAGAAAG 300 pMDC83::AtGlcAT14B 1741 CTGATGTTTTCTGGGTTTTCAGAGAGAGGATGCTCAGACCTTTCAAACTTTCAGAG 1800 Sequencing result 301 CTGATGTTTTCTGGGTTTTCAGAGAGAGGATGCTCAGACCTTTCAAACTTTCAGAG 360 pMDC83::AtGlcAT14B 1801 GTGCTTCATTTTTCATATCTTCTGTTTGGTAACTCAACATAAGGGAATGTCTAA 1860 Sequencing result 361 GTGCTTCATTTTTCATATCTTCTGTTTGGTAACTCAACATAAGGGAATGTCTAA 420 pMDC83::AtGlcAT14B 1861 ATGTAGAGTCTTCTTAAATGTAGAGCTCTTTAGACATCAATGTATGATATACAAAT 1920 Sequencing result 421 ATGTAGAGTCTTCTTAAATGTAGAGCTCTTTAGACATCAATGTATGATATACAAAT 480 pMDC83::AtGlcAT14B 1921 CTTAGGCTCTAAATGTTTTTATGATGTTTGTGTAATAGATCTGTTGATGATGAT 1980 Sequencing result 481 CTTAGGCTCTAAATGTTTTTATGATGTTTGTGTAATAGATCTGTTGATGATGAT 540 pMDC83::AtGlcAT14B 1981 CTAGACCTTTTGTGATATTTCATATGGGAGTGGGATACCTTCTCGAATTTGTGTA 2040 Sequencing result 541 CTAGACCTTTTGTGATATTTCATATGGGAGTGGGATACCTTCTCGAATTTGTGTA 600 pMDC83::AtGlcAT14B 2041 TGTATTACGCAAACTTCTTTCATGCGCAGAGGTTACTTCCACATGTTGATTCGAAT 2100 Sequencing result 601 TGTATTACGCAAACTTCTTTCATGCGCAGAGGTTACTTCCACATGTTGATTCGAAT 660 pMDC83::AtGlcAT14B 2101 CTCGAGATTTACACACACAGGTTACAGAGTACTTACATTCATCTCATGAGCAATC 2160 Sequencing result 661 CTCGAGATTTACACACACAGGTTACAGAGTACTTACATTCATCTCATGAGCAATC 720 pMDC83::AtGlcAT14B 2161 CTCGAAGCAACATCTCCACATCTTACATAGATGATTCAGAGAGTGGTGGACAT 2220 Sequencing result 721 CTCGAAGCAACATCTCCACATCTTACATAGATGATTCAGAGAGTGGTGGACAT 780 pMDC83::AtGlcAT14B 2221 ACCTGCTCTTTGCTAGGAAATTCCTAGAGATGAGCTGTTCTCGATAGATTGATCG 2280 Sequencing result 781 ACCTGCTCTTTGCTAGGAAATTCCTAGAGATGAGCTGTTCTCGATAGATTGATCG 840 pMDC83::AtGlcAT14B 2281 AACTATTGTTTCGAGATCAGGAGATGCTACTCGGTTGGTGGTATTGATGATGAGG 2340 Sequencing result 841 AACTATTGTTTCGAGATCAGGAGATGCTACTCGGTTGGTGGTATTGATGATGAGG 900 pMDC83::AtGlcAT14B 2341 AAAACGGAGGATCGCTGCTGCTGCTGATTTGATGATGCTGATTAACCTGCTCTG 2400 Sequencing result 901 AAAACGGAGGATCGCTGCTGCTGCTGATTTGATGATGCTGATTAACCTGCTCTG 960 pMDC83::AtGlcAT14B 2401 GACTTAACGAGATGAGAACTCTAATCTATTGTTATCTAGAGATTTTCAGACAC 2460 Sequencing result 961 GACTTAACGAGATGAGAACTCTAATCTATTGTTATCTAGAGATTTTCAGACAC 984	B pMDC83::AtGlcAT14E 961 TGAGCGGAGACTCACTTATGAGAACACAACTGAGTGGATGAGAGAGTGAAGCAT 1020 Sequencing result 1 -----CACACAACTCACTGAGTGGATGAGAGAGTGAAGCAT 34 pMDC83::AtGlcAT14E 1021 ATTCAGCTTCTGAGTTTCCATTTTGGTTAGTCTATGTTAGTGGATGAGTGGTGG 1080 Sequencing result 35 ATTCAGCTTCTGAGTTTCCATTTTGGTTAGTCTATGTTAGTGGATGAGTGGTGG 94 pMDC83::AtGlcAT14E 1081 TTTCTGTTGAGGAGAGAGAGAGACAGCATTAATGATGATGATGATGATGATGAT 1140 Sequencing result 95 TTTCTGTTGAGGAGAGAGAGAGACAGCATTAATGATGATGATGATGATGATGATGAT 154 pMDC83::AtGlcAT14E 1141 TGACCAATCTGATATTATGCTGGTGCACCTGCGGAGGTTTGGCAAGTGGCTTAAAT 1200 Sequencing result 155 TGACCAATCTGATATTATGCTGGTGCACCTGCGGAGGTTTGGCAAGTGGCTTAAAT 214 pMDC83::AtGlcAT14E 1201 TATTCAGTGGTCACTCAATCTTAACTGATGATGATGATGATGATGATGATGATGAT 1260 Sequencing result 215 TATTCAGTGGTCACTCAATCTTAACTGATGATGATGATGATGATGATGATGATGATGAT 274 pMDC83::AtGlcAT14E 1261 CTCTCTCTCAGCATCAAACTTGAAGCATTTATCTGGTGGCTGTTGATATCTAAGA 1320 Sequencing result 275 CTCTCTCTCAGCATCAAACTTGAAGCATTTATCTGGTGGCTGTTGATATCTAAGA 334 pMDC83::AtGlcAT14E 1321 CCAAACTCTAGTCAATGATCAAGCAAGTGGAAATTTTTTCTGATCTGTTGTTGAC 1380 Sequencing result 335 CCAAACTCTAGTCAATGATCAAGCAAGTGGAAATTTTTTCTGATCTGTTGTTGAC 394 pMDC83::AtGlcAT14E 1381 AGGATCACTGTTGATGCTCTATCAAGCGCTTTGATAGATGATGATGATGATGATGAT 1440 Sequencing result 395 AGGATCACTGTTGATGCTCTATCAAGCGCTTTGATAGATGATGATGATGATGATGAT 454 pMDC83::AtGlcAT14E 1441 CAATTACACAGCACTCTCTAATGATATACCAACTCTGTTTCACTGCGGAGGTTA 1500 Sequencing result 455 CAATTACACAGCACTCTCTAATGATATACCAACTCTGTTTCACTGCGGAGGTTA 514 pMDC83::AtGlcAT14E 1501 CTTCAGACCGTGAATTCGACAGCTGCTGAGTTTGCAGAACTGCAATCAACATGACT 1560 Sequencing result 515 CTTCAGACCGTGAATTCGACAGCTGCTGAGTTTGCAGAACTGCAATCAACATGACT 574 pMDC83::AtGlcAT14E 1561 ACATACATTTCTGGGAGATTCGCGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1620 Sequencing result 575 ACATACATTTCTGGGAGATTCGCGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 634 pMDC83::AtGlcAT14E 1621 CACAATCCGATGATATGAGAGCTGCTGCTGTTTCTAGGAAATTCAGAGAGATGATGA 1680 Sequencing result 635 CACAATCCGATGATATGAGAGCTGCTGCTGTTTCTAGGAAATTCAGAGAGATGATGA 694 pMDC83::AtGlcAT14E 1681 AGTACTTAATAGATGATAGAGATTAATTAACGAGAGATGATAGAGATGATTTTAC 1740 Sequencing result 695 AGTACTTAATAGATGATAGAGATTAATTAACGAGAGATGATAGAGATGATTTTAC 754 pMDC83::AtGlcAT14E 1741 TCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800 Sequencing result 755 TCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 814 pMDC83::AtGlcAT14E 1801 TGTCTCTAGCTTTGAGACTCAGAGGCTCAAGACTTGTGACAGGATGATGACAGAG 1860 Sequencing result 815 TGTCTCTAGCTTTGAGACTCAGAGGCTCAAGACTTGTGACAGGATGATGACAGAG 839	C pMDC83::At4g27480 721 TTCTCTTGAAGCAGAGTAAAGATGGATGGTTCATCAATCTCAGTGGCTGGATATC 780 Sequencing result 1 -----ATGCTCTGGATATC 16 pMDC83::At4g27480 781 CTCTAGTCACTCAAGATGTAATATAGTATGATATGATCTCTCTCTCTCTCTCTCTCT 840 Sequencing result 17 CTCTAGTCACTCAAGATGTAATATAGTATGATATGATCTCTCTCTCTCTCTCTCTCTCT 76 pMDC83::At4g27480 841 TAATCTCTGCGGAGTATATCAATGCGCTTCTGCTGAGATGATGATGATGATGATGAT 900 Sequencing result 77 TAATCTCTGCGGAGTATATCAATGCGCTTCTGCTGAGATGATGATGATGATGATGATGAT 136 pMDC83::At4g27480 901 ACTGAGCGGAGCTCACTTTATGACCAAGTACCAATGAGCTGAGAGAGATGATG 960 Sequencing result 137 ACTGAGCGGAGCTCACTTTATGACCAAGTACCAATGAGCTGAGAGAGATGATG 196 pMDC83::At4g27480 961 ATTATTCTCTGATAGATGATTTAGAGCATTTTATGCTGTTTCTGCTCAAAATGCT 1020 Sequencing result 197 ATTATTCTCTGATAGATGATTTAGAGCATTTTATGCTGTTTCTGCTCAAAATGCT 256 pMDC83::At4g27480 1021 TTTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080 Sequencing result 257 TTTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 316 pMDC83::At4g27480 1081 AGAAGTCAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140 Sequencing result 317 AGAAGTCAGATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 376 pMDC83::At4g27480 1141 TTACCGTAAATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200 Sequencing result 377 TTACCGTAAATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 436 pMDC83::At4g27480 1201 ACTGAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260 Sequencing result 437 ACTGAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496 pMDC83::At4g27480 1261 TGGGCAATCTTCCAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320 Sequencing result 497 TGGGCAATCTTCCAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 556 pMDC83::At4g27480 1321 GGAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380 Sequencing result 557 GGAATTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 616 pMDC83::At4g27480 1381 GACCTGATTTCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440 Sequencing result 617 GACCTGATTTCTCTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 676 pMDC83::At4g27480 1441 AACGATACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500 Sequencing result 677 AACGATACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 736 pMDC83::At4g27480 1501 GATCTCTCTCTGATAGATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560 Sequencing result 737 GATCTCTCTCTGATAGATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796 pMDC83::At4g27480 1561 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620 Sequencing result 797 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856 pMDC83::At4g27480 1621 AAGCCCGCTCTGAGCTAAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680 Sequencing result 857 AAGCCCGCTCTGAGCTAAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 890
---	--	---

Supplementary fig. 15 - Alignment of the sequencing results of *GlcAT14B::GFP* (A), *GlcAT14E::GFP* (B) and *At4g27480::GFP* (C) constructs.

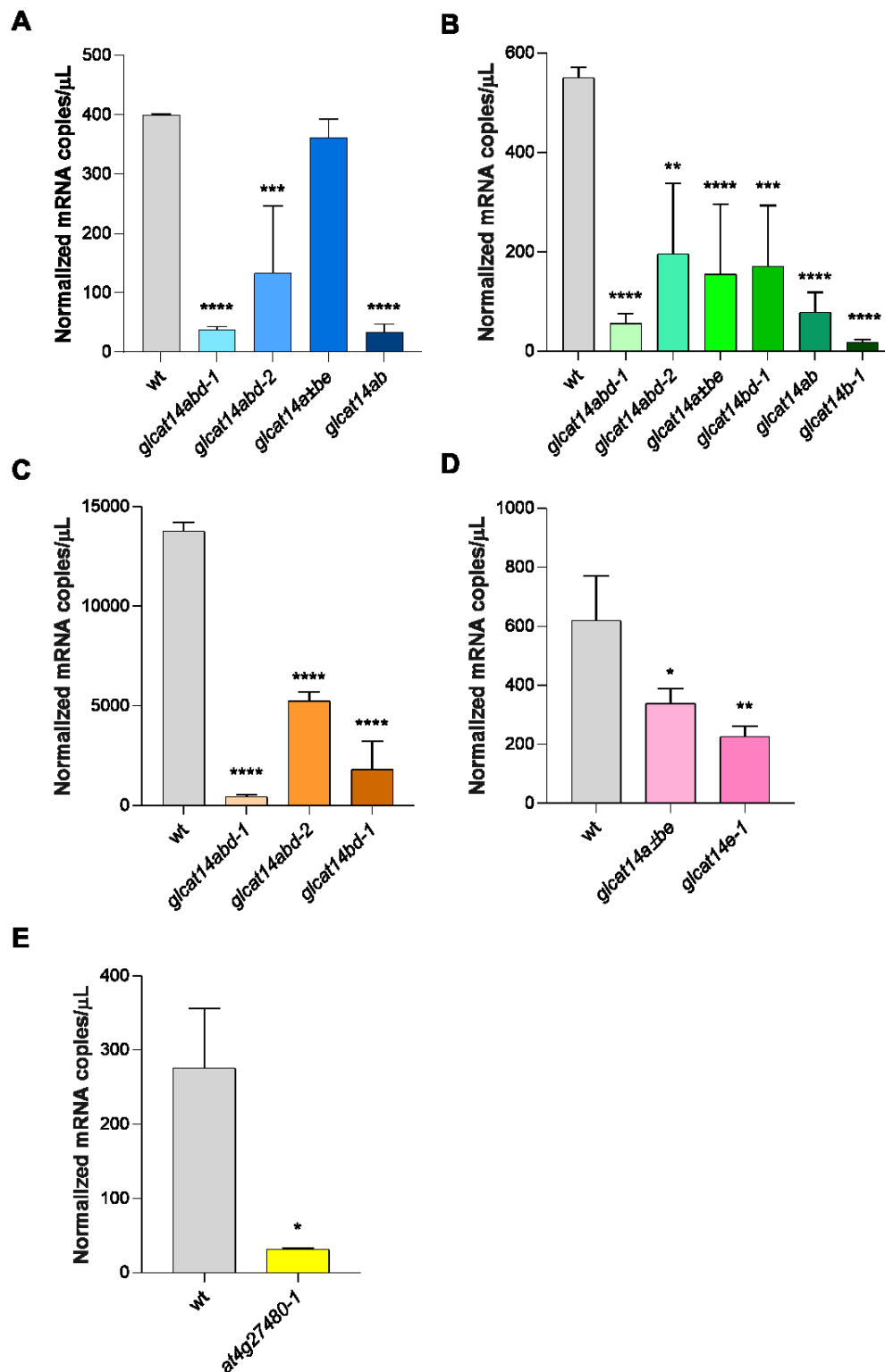
The sequencing results with the primers AtGlcAT14B_CDS_rv, AtGlcAT14E_CDS_rv and At4g27480_CDS_rv (Supplementary table 5) were aligned with the respective expected *pMDC83::GlcAT14B*, *pMDC83::GlcAT14E* and *pMDC83::At4g27480* sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).



A					B				
AtGlcAT14A	1441	GCCGCGGGATTGAACTTCATTGATCATAGTAACATTGGATGGAAGCGTAAGAAGA	1500		AtGlcAT14B	2401	GGTCTTGGAGCTAAACGGATCGAGAACTCATAACTTATTTGTTATCTACTGAGAATTC	2460	
		TTCATTGATCATAGTAACATTGGATGGAAGCGTAAGAAGA					ATTGTTATCTACTGAGAATTC		
Sequencing result (SALK_043905)	1	-----TTCATTGATCATAGTAACATTGGATGGAAGCGTAAGAAGA	44		Sequencing result (SALK_080923)	1	-----ATTGTTATCTACTGAGAATTC	23	
AtGlcAT14A	1501	GTTTATAGAGTTTTGTTTCTGTGTCCTTTGTTTCGTATAGTAGAAGGAACCTGATTTT	1560		AtGlcAT14B	2461	AGACCACGGCAATGCCGGTAATGGCAAATTAGAGTTTGTGCATAGAGATGGTTACGTGGT	2520	
		GTTTATAGAGTTTTGTTTCTGTGTCCTTTGTTTCGTATAGTAGAAGGAACCTGATTTT					AGACCACGGCAATGCCGGTAATGGCAAATTAGAGTTTGTGCATAGAGATGGTTACGTGGT		
Sequencing result (SALK_043905)	45	GTTTATAGAGTTTTGTTTCTGTGTCCTTTGTTTCGTATAGTAGAAGGAACCTGATTTT	104		Sequencing result (SALK_080923)	24	AGACCACGGCAATGCCGGTAATGGCAAATTAGAGTTTGTGCATAGAGATGGTTACGTGGT	83	
AtGlcAT14A	1561	GTTGTTGTTTGTATCTGGTTCTTTTAGATCACAAAGAGCAAAACCGGTGATTATAGATC	1620		AtGlcAT14B	2521	TTTTCACAAATGAACGTGATTGATTATCCGATACTTTGATATGGTTGTATAGAGATCT	2580	
		GTTGTTGTTTGTATCTGGTTCTTTTAGATCACAAAGAGCAAAACCGGTGATTATAGATC					TTTTCACAAATGAACGTGATTGATTATCCGATACTTTGATATGGTTGTATAGAGATCT		
Sequencing result (SALK_043905)	105	GTTGTTGTTTGTATCTGGTTCTTTTAGATCACAAAGAGCAAAACCGGTGATTATAGATC	164		Sequencing result (SALK_080923)	84	TTTTCACAAATGAACGTGATTGATTATCCGATACTTTGATATGGTTGTATAGAGATCT	143	
AtGlcAT14A	1621	CTGGACTGTATTGAAACAAGAAATCTGATGTTTTTGGGTTACGCAAGACGGAGCATCC	1680		AtGlcAT14B	2581	TTTGGACTTTCCTCATTGAGAAAGTCAGAAATTCGGGAAAGATCATTGATTATTGG	2640	
		CTGGACTGTATTGAAACAAGAAATCTGATGTTTTTGGGTTACGCAAGACGGAGCATCC					TTTGGACTTTCCTCATTGAGAAAGTCAGAAATTCGGGAAAGATCATTGATTATTGG		
Sequencing result (SALK_043905)	165	CTGGACTGTATTGAAACAAGAAATCTGATGTTTTTGGGTTACGCAAGACGGAGCATCC	224		Sequencing result (SALK_080923)	144	TTTGGACTTTCCTCATTGAGAAAGTCAGAAATTCGGGAAAGATCATTGATTATTGG	203	
AtGlcAT14A	1681	CAACAGCATTCAAGCTTTTCACAGGTGCTTGATCCCTCTAAAGCCTACTCGGTTTGTAC	1740		AtGlcAT14B	2641	AATTTGCTCTTATTGCCGACCTAAAGTTTGGCCTTTTCCAACCTGATACCAATTTATA	2700	
		CAACAGCATTCAAGCTTTTCACAGGTGCTTGATCCCTCTAAAGCCTACTCGGTTTGTAC					AATTTGCTCTTATTGCCGACCTAAAGTTTGGCCTTTTCCAACCTGATACCAATTTATA		
Sequencing result (SALK_043905)	225	CAACAGCATTCAAGCTTTTCACAGGTGCTTGATCCCTCTAAAGCCTACTCGGTTTGTAC	284		Sequencing result (SALK_080923)	204	AATTTGCTCTTATTGCCGACCTAAAGTTTGGCCTTTTCCAACCTGATACCAATTTATA	263	
AtGlcAT14A	1741	ACTATAATCATCTCTGAAATGCTATGATTGATTTCATGTTTAGTTGGATGAAACACTGA	1800		AtGlcAT14B	2701	GGTTTGTACATATTAGTCAATGCTCTTTTATTGTTTGCATGATGAAGTTACCTGACTAT	2760	
		ACTATAATCATCTCTGAAATGCTATGATTGATTTCATGTTTAGTTGGATGAAACACTGA					GGTTTGTACATATTAGTCAATGCTCTTTTATTGTTTGCATGATGAAGTTACCTGACTAT		
Sequencing result (SALK_043905)	285	ACTATAATCATCTCTGAAATGCTATGATTGATTTCATGTTTAGTTGGATGAAACACTGA	344		Sequencing result (SALK_080923)	264	GGTTTGTACATATTAGTCAATGCTCTTTTATTGTTTGCATGATGAAGTTACCTGACTAT	323	
AtGlcAT14A	1801	CTCATCATACTTAGATTGCTAGGAAAATCTTTATAATCTCTGAGCAATTCACAGACTGT	1860		AtGlcAT14B	2761	TGTGACTAACATATAGGTGTTTCATTGATATTATTGCTTTTGAATTAATTGATACATGA	2820	
		CTCATCATACTTAGATTGCTAGGAAAATCTTTATAATCTCTGAGCAATTCACAGACTGT					TGTGACTAACATATAGGTGTTTCATTGATATTATTGCTTTTGAATTAATTGATACATGA		
Sequencing result (SALK_043905)	345	CTCATCATACTTAGATTGCTAGGAAAATCTTTATAATCTCTGAGCAATTCACAGACTGT	404		Sequencing result (SALK_080923)	324	TGTGACTAACATATAGGTGTTTCATTGATATTATTGCTTTTGAATTAATTGATACATGA	383	
AtGlcAT14A	1861	TTTGATTGGTTGCTCTATGTAGCTGTTTAGTTTCGGTCGGTGCTTCTCAGATTGCTCCA	1920		AtGlcAT14B	2821	TTCAATAACTCACATTGATATTGTTGTTTTAAATTTATAGATATACACCCATGTGACTAA	2880	
		TTTGATTGGTTGCTCTATGTAGCTGTTTAGTTTCGGTCGGTGCTTCTCAGATTGCTCCA					TTCAATAACTCACATTGATATTGTTGTTTTAAATTTATAGATATACACCCATGTGACTAA		
Sequencing result (SALK_043905)	405	TTTGATTGGTTGCTCTATGTAGCTGTTTAGTTTCGGTCGGTGCTTCTCAGATTGCTCCA	464		Sequencing result (SALK_080923)	384	TTCAATAACTCACATTGATATTGTTGTTTTAAATTTATAGATATACACCCATGTGACTAA	443	
AtGlcAT14A	1921	TTTCTCATTGGTTGAACAACCGTGAATCATTATGCTTAGATTAGAATTTGATAAGATC	1980		AtGlcAT14B	2881	GTGCAAAATTACATGAGGCGCTATTGATGGTCACTACTCATTCCATGTGACTGCTGGTGG	2940	
		TTT					GTGCAAAATTACATGAGGCGCTATTGA		
Sequencing result (SALK_043905)	465	TTT-----	467		Sequencing result (SALK_080923)	444	GTGCAAAATTACATGAGGCGCTATTGA-----	469	

Supplementary fig. 17 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from *atg1cat14a-2* (A) and *atg1cat14b-1* (B) mutants.

The sequencing result with the primer LBB1.3 (Supplementary table 1Supplementary fig. 1) were aligned with the expected GlcAT14A and GlcAT14B flanking sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).



Supplementary fig. 18 - qPCR analysis to detect the presence of *GlcAT14A* (A), *GlcAT14B* (B), *GlcAT14D* (C), *GlcAT14E* (D) and *At4g27480* (E) transcripts in the inflorescences of wt and mutants. The final concentrations of mRNAs are expressed as arbitrary units that represent the numbers of copies per microliter of cDNA, normalized against the geometric means of three control genes that vary the least with respect to each other. Means of three biological replicates are shown and error bars indicate SDs for each mRNA.

A

AtGlcAT14D	3181	GATCAGTGTAACATTTTCTCTGGTTCTTGTAGGTCAGGAAGTTGAAGCCAGTTATTG	3240
Sequencing result (GK363F05)	1	-----TCTCTTGGTTCTTGTAGGTCAGGAAGTTGAAGCCAGTTATTG	44
AtGlcAT14D	3241	TTGATCCAGGGCTCTATCTTGTGGAGAAAACCGATATGTTTTTCGCTTCGAGAAACGAG	3300
Sequencing result (GK363F05)	45	TTGATCCAGGGCTCTATCTTGTGGAGAAAACCGATATGTTTTTCGCTTCGAGAAACGAG	104
AtGlcAT14D	3301	AGTTGCCTAAGGCTTCAAGTTATTCTCAGGTATGCTTAGCTACGTTACATTACGTTTTC	3360
Sequencing result (GK363F05)	105	AGTTGCCTAAGGCTTCAAGTTATTCTCAGGTATGCTTAGCTACGTTACATTACGTTTTC	164
AtGlcAT14D	3361	ATTTTCCCTTTTCTGTTTTCTGCAATTATCTGTGGATCATTACTAATTGACCACCTTGT	3420
Sequencing result (GK363F05)	165	ATTTTCCCTTTTCTGTTTTCTGCAATTATCTGTGGATCATTACTAATTGACCACCTTGT	224
AtGlcAT14D	3421	TGCTATTTAGCATATTCAGTGTGTGCAATCATTGGATTGATATTATTATGGAAGATTA	3480
Sequencing result (GK363F05)	225	TGCTATTTAGCATATTCAGTGTGTGCAATCATTGGATTGATATTATTATGGAAGATTA	284
AtGlcAT14D	3481	CTTTTCCCTAATTATGCTGTTTTGTAGAACTTAATCCCTTTTCTTTTACATGGTTAGC	3540
Sequencing result (GK363F05)	285	CTTTTCCCTAATTATGCTGTTTTGTAGAACTTAATCCCTTTTCTTTTACATGGTTAGC	344
AtGlcAT14D	3541	AGTTTCTTGGCACCATTTCATAATTTGGTAGGTAGATACGTGATAGTAGAAGTGGATA	3600
Sequencing result (GK363F05)	345	AGTTTCTTGGCACCATTTCATAATTTGGTAGGTAGATACGTGATAGTAGAAGTGGATA	404
AtGlcAT14D	3601	AGTTTAGGTGAAAGCTCTATCAGAATATAGACTACATGACGCTCTGTTTATCAGTAATGC	3660
Sequencing result (GK363F05)	405	AGTTTAGGTGAAAGCTCTATCAGAATATAGACTACATGACGCTCTGTTTATCAGTAATGC	464
AtGlcAT14D	3661	TATTAGCTGCTTTAA-----GCAGATAAGTGAGGAAAGGAGCCAGTGCTAGAAATTG	3715
Sequencing result (GK363F05)	465	TATTAGCTGCTTT-----	477

B

AtGlcAT14D	3181	GATCAGTGTAACATTTTCTCTGGTTCTTGTAGGTCAGGAAGTTGAAGCCAGTTATTG	3240
Sequencing result (GK508D01)	1	-----TCTCTTGGTTCTTGTAGGTCAGGAAGTTGAAGCCAGTTATTG	45
AtGlcAT14D	3241	TTGATCCAGGGCTCTATCTTGTGGAGAAAACCGATATGTTTTTCGCTTCGAGAAACGAG	3300
Sequencing result (GK508D01)	46	TTGATCCAGGGCTCTATCTTGTGGAGAAAACCGATATGTTTTTCGCTTCGAGAAACGAG	105
AtGlcAT14D	3301	AGTTGCCTAAGGCTTCAAGTTATTCTCAGGTATGCTTAGCTACGTTACATTACGTTTTC	3360
Sequencing result (GK508D01)	106	AGTTGCCTAAGGCTTCAAGTTATTCTCAGGTATGCTTAGCTACGTTACATTACGTTTTC	165
AtGlcAT14D	3361	ATTTTCCCTTTTCTGTTTTCTGCAATTATCTGTGGATCATTACTAATTGACCACCTTGT	3420
Sequencing result (GK508D01)	166	ATTTTCCCTTTTCTGTTTTCTGCAATTATCTGTGGATCATTACTAATTGACCACCTTGT	225
AtGlcAT14D	3421	TGCTATTTAGCATATTCAGTGTGTGCAATCATTGGATTGATATTATTATGGAAGATTA	3480
Sequencing result (GK508D01)	226	TGCTATTTAGCATATTCAGTGTGTGCAATCATTGGATTGATATTATTATGGAAGATTA	285
AtGlcAT14D	3481	CTTTTCCCTAATTATGCTGTTTTGTAGAACTTAATCCCTTTTCTTTTACATGGTTAGC	3540
Sequencing result (GK508D01)	286	CTTTTCCCTAATTATGCTGTTTTGTAGAACTTAATCCCTTTTCTTTTACATGGTTAGC	345
AtGlcAT14D	3541	AGTTTCTTGGCACCATTTCATAATTTGGTAGGTAGATACGTGATAGTAGAAGTGGATA	3600
Sequencing result (GK508D01)	346	AGTTTCTTGGCACCATTTCATAATTTGGTAGGTAGATACGTGATAGTAGAAGTGGATA	405
AtGlcAT14D	3601	AGTTTAGGTGAAAGCTCTATCAGAATATAGACTACATGACGCTCTGTTTATCAGTAATGC	3660
Sequencing result (GK508D01)	406	AGTTTAGGTGAAAGCTCTATCAGAATATAGACTACATGACGCTCTGTTTATCAGTAATGC	465
AtGlcAT14D	3661	TATTAGCTGCTTTAAGCAGATAAGTGAGGAAAGGAGCCAGTGCTAGAAATTGCTTAC	3720
Sequencing result (GK508D01)	466	TATTAGCTGCTTTAAGCAGATAAGTGAGGAAAGGAGCCAGTGCTAGAAATTGCTTAC	525
AtGlcAT14D	3721	AATGGTTCTCATTACTGTTATATTTTGTGGTTGTTTAAACAACAAAGTTCTCCTATTGAA	3780
Sequencing result (GK508D01)	526	AATGGTTCTCATTACTGTTATATTTTGTGGTTGTTTAAACAACAAAGTTCTCCTATTGAA	585
AtGlcAT14D	3781	TAGCATAATTATATTTAGTTCCTACTGAATATAGCATATTAGGAAAAGAAATCTTTGG	3840
Sequencing result (GK508D01)	586	TAGCATAATTATATTTAGTTCCTACTGAATATAGCATA-----	623

Supplementary fig. 19 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from *atglcat14d-1* (A) and *atglcat14d-2* (B) mutants.

The sequencing result with the primer o8409 (Supplementary table 1) were aligned with the expected GlcAT14D flanking sequence. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

A

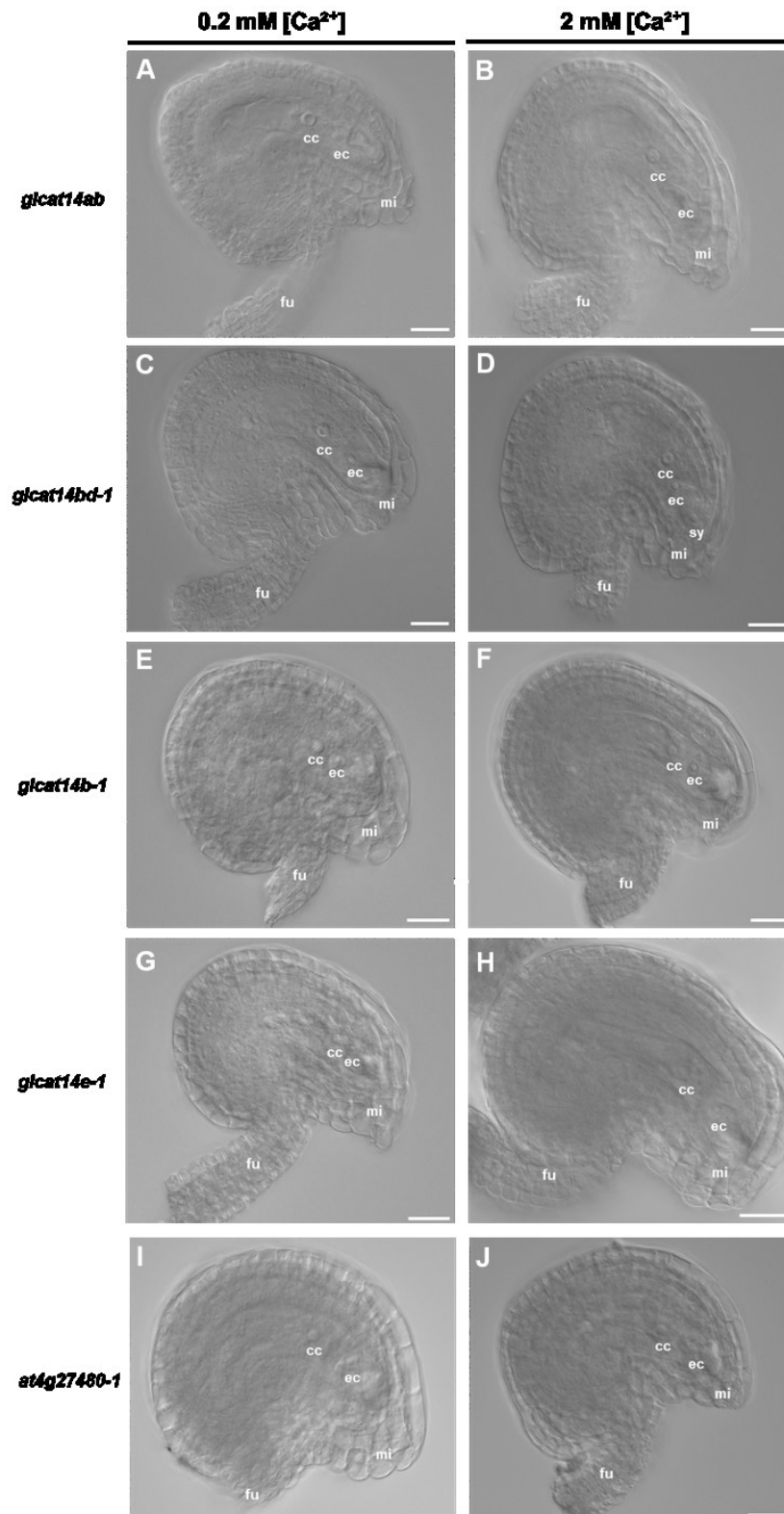
AtGlcAT14E	2221	GTTCCTGAGTTTCCTATTGGTTTAGTCTTATGTTCTAGTGAGTCTAGTTGTTGTTCT	2280
		GTTCCT	
Sequencing result (SALK_022820)	1	-----GTTCCT	7
AtGlcAT14E	2281	TGTGAAGGAGAGCGAGCAGCATTATGATTGATCTGGACTCTACTTGTGAACA	2340
		TGTGAAGGAGAGCGAGCAGCATTATGATTGATCTGGACTCTACTTGTGAACA	
Sequencing result (SALK_022820)	8	TGTGAAGGAGAGCGAGCAGCATTATGATTGATCTGGACTCTACTTGTGAACA	67
AtGlcAT14E	2341	AATCTGATATTACTGGGTACACCTCGCCGAAGTTGCCAAGTGCCTTTAAGTTATTCA	2400
		AATCTGATATTACTGGGTACACCTCGCCGAAGTTGCCAAGTGCCTTTAAGTTATTCA	
Sequencing result (SALK_022820)	68	AATCTGATATTACTGGGTACACCTCGCCGAAGTTGCCAAGTGCCTTTAAGTTATTCA	127
AtGlcAT14E	2401	CTGGTCAGTCATCTAATCCCTCTAACTGCATTGGAATGATACATTGGATCAAACTCTT	2460
		CTGGTCAGTCATCTAATCCCTCTAACTGCATTGGAATGATACATTGGATCAAACTCTT	
Sequencing result (SALK_022820)	128	CTGGTCAGTCATCTAATCCCTCTAACTGCATTGGAATGATACATTGGATCAAACTCTT	187
AtGlcAT14E	2461	CTCACCATTCAAACTCTGAAGCATTATTCTCGGTGCTGTGTATATCTATAAGACAAAA	2520
		CTCACCATTCAAACTCTGAAGCATTATTCTCGGTGCTGTGTATATCTATAAGACAAAA	
Sequencing result (SALK_022820)	188	CTCACCATTCAAACTCTGAAGCATTATTCTCGGTGCTGTGTATATCTATAAGACAAAA	247
AtGlcAT14E	2521	CTCTAGTCATATGCATCAAGCTTGAAATATTTTTTCGTACTGTTGTGTACAGGATC	2580
		CTCTAGTCATATGCATCAAGCTTGAAATATTTTTTCGTACTGTTGTGTACAGGATC	
Sequencing result (SALK_022820)	248	CTCTAGTCATATGCATCAAGCTTGAAATATTTTTTCGTACTGTTGTGTACAGGATC	307
AtGlcAT14E	2581	AGCTTGGATGGCTCTATCACGCCGTTTGTAGAGTATTGCATATGGGATGGGACAAATT	2640
		AGCTTGGATGGCTCTATCACGCCGTTTGTAGAGTATTGCATATGGGATGGGACAAATT	
Sequencing result (SALK_022820)	308	AGCTTGGATGGCTCTATCACGCCGTTTGTAGAGTATTGCATATGGGATGGGACAAATT	367
AtGlcAT14E	2641	ACCACGAACCTCTCTAATGTATTACACCAACTTCGTTTCATCGCCGGAAGTTACTTCCA	2700
		ACCACGAACCTCTCTAATGTATTACACCAACTTCGTTTCATCGCCGGAAGTTACTTCCA	
Sequencing result (SALK_022820)	368	ACCACGAACCTCTCTAATGTATTACACCAACTTCGTTTCATCGCCGGAAGTTACTTCCA	427
AtGlcAT14E	2701	GACCGTGATTGCAACGTGCTGAGTTTGCAAAACTGCAGTAAACCATGACCTACACTA	2760
		GACCGTGATTGCAACGTGCTGAGTTTGCAAAACTGCAGTAAACCATGACCTACACTA	
Sequencing result (SALK_022820)	428	GACCGTGATTGCAACGTGCTGAGTTTGCAAAACTGCAGTAAACCATGACCTACACTA	487
AtGlcAT14E	2761	CATTTCCTGGGACAAATCCGCCGAGCAGCAGCCACACGTTGTGCTCAACGACACAAT	2820
		CATTTCCTGGGACAAATCCGCCGAGCAGCAGCCACACGTTGTGCTCAACGACACAAT	
Sequencing result (SALK_022820)	488	CATTTCCTGGGACAAATCCGCCGAGCAGCAGCCACACGTTGTGCTCAACGACACAAT	547
AtGlcAT14E	2821	GCCGATGATATGGAGCGGTGCTGCGTTTGTAGGAAATCAGAAGAGATGATGAAGTACT	2880
		GCCGATGATATGGAGCGGTGCTGCGTTTGTAGGAAATCAGAAGAGATGATGAAGTACT	
Sequencing result (SALK_022820)	548	GCCGATGATATGGAGCGGTGCTGCGTTTGTAGGAAATCAGAAGAGATGATGAAGTACT	607
AtGlcAT14E	2881	TAATAAGATCGATAAGGAATTACTTAAACGAGGAGATGATAAGATAGTTTACTCTCTGG	2940
		TAATAAGATCGATAAGGAATTACTTAAACGAGGAGATGATAAGATAGTTTACTCTCTGG	
Sequencing result (SALK_022820)	608	TAATAAGATCGATAAGGAATTACTTAAACG-----	637

B

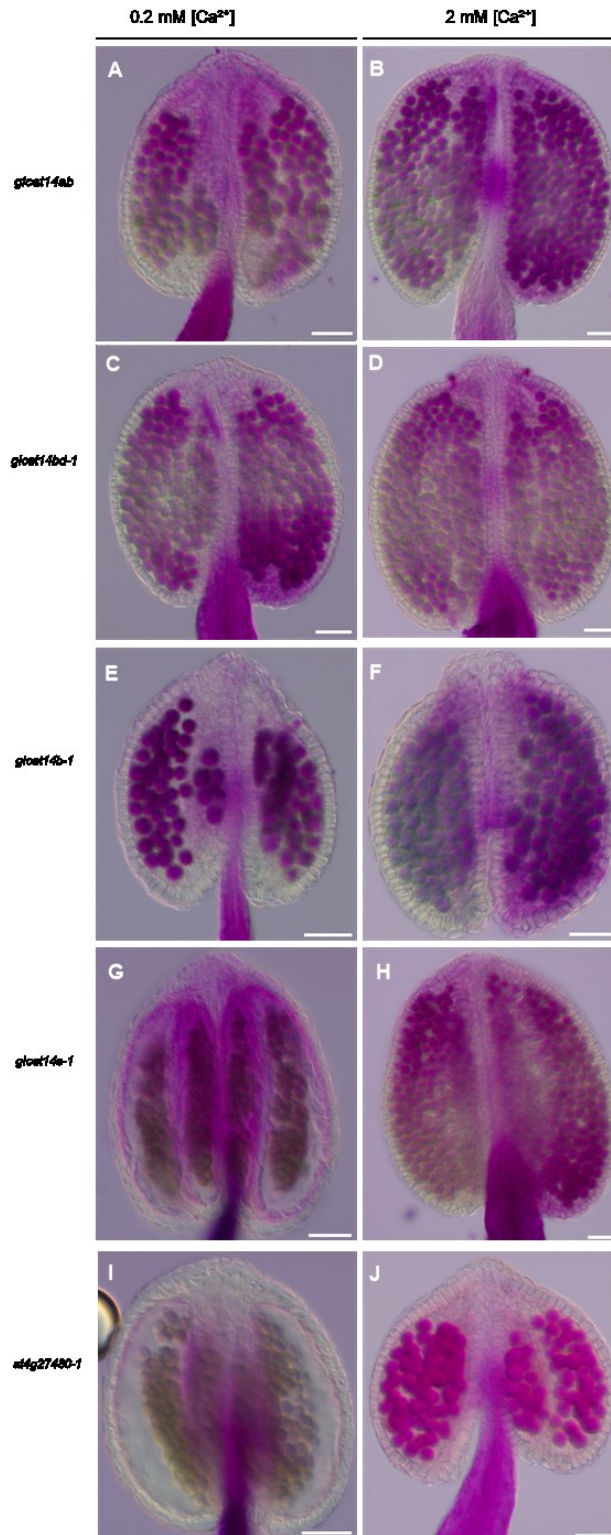
At4g27480	1381	TCCTCTCTCTGTCGGTCTTAACTCACTCATCTCTCTACAATCTATCGACCACCA	1440
		ACCA	
Sequencing result (SALK_105434C)	1	-----ACCA	4
At4g27480	1441	ACGAGACACGAGTCGAATTGACAGATCAAGATCAATCAGTCTCTCGCATCTCTCCAG	1500
		ACGAGACACGAGTCGAATTGACAGATCAAGATCAATCAGTCTCTCGCATCTCTCCAG	
Sequencing result (SALK_105434C)	5	ACGAGACACGAGTCGAATTGACAGATCAAGATCAATCAGTCTCTCGCATCTCTCCAG	64
At4g27480	1501	TACAGCCTAGTCTCTCTCGCTTTGGTTATCTAGTTTCAGGGTCAAGAGTGACTTAGAGA	1560
		TACAGCCTAGTCTCTCTCGCTTTGGTTATCTAGTTTCAGGGTCAAGAGTGACTTAGAGA	
Sequencing result (SALK_105434C)	65	TACAGCCTAGTCTCTCTCGCTTTGGTTATCTAGTTTCAGGGTCAAGAGTGACTTAGAGA	124
At4g27480	1561	GTCTGTGGAGAGTGTACGAACATTATACCATCTAGGAATCAGTACGTTGTTCATCTTG	1620
		GTCTGTGGAGAGTGTACGAACATTATACCATCTAGGAATCAGTACGTTGTTCATCTTG	
Sequencing result (SALK_105434C)	125	GTCTGTGGAGAGTGTACGAACATTATACCATCTAGGAATCAGTACGTTGTTCATCTTG	184
At4g27480	1621	ATCTCGAGTCTCTCTGCTGAAGAGAGGCTTGAGTGGCTAAACGTGTGAGCCAAAGTCTTG	1680
		ATCTCGAGTCTCTCTGCTGAAGAGAGGCTTGAGTGGCTAAACGTGTGAGCCAAAGTCTTG	
Sequencing result (SALK_105434C)	185	ATCTCGAGTCTCTCTGCTGAAGAGAGGCTTGAGTGGCTAAACGTGTGAGCCAAAGTCTTG	244
At4g27480	1681	TTTTTAGCGATGTTGAAATGTTACATGATCACAAGGCTAACCTTGTACGTATCGAG	1740
		TTTTTAGCGATGTTGAAATGTTACATGATCACAAGGCTAACCTTGTACGTATCGAG	
Sequencing result (SALK_105434C)	245	TTTTTAGCGATGTTGAAATGTTACATGATCACAAGGCTAACCTTGTACGTATCGAG	304
At4g27480	1741	GACCCACAATGGTGGCTAACCGCTCCAGCTTGGCTATTCTCTTGAAGCAGAGTAAAG	1800
		GACCCACAATGGTGGCTAACCGCTCCAGCTTGGCTATTCTCTTGAAGCAGAGTAAAG	
Sequencing result (SALK_105434C)	305	GACCCACAATGGTGGCTAACCGCTCCAGCTTGGCTATTCTCTTGAAGCAGAGTAAAG	364
At4g27480	1801	AATGGGATTGGTTCATCAATCTCAGTGCCTCGGATTATCTCTAGTCACTCAAGATGGTA	1860
		AATGGGATTGGTTCATCAATCTCAGTGCCTCGGATTATCTCTAGTCACTCAAGATGGTA	
Sequencing result (SALK_105434C)	365	AATGGGATTGGTTCATCAATCTCAGTGCCTCGGATTATCTCTAGTCACTCAAGATGGTA	424
At4g27480	1861	ATATAGTTATGATCTAGTCTCTCTTTCTTCAAAATCTCTGCGCCGAGATTATCA	1920
		ATATAGTTATGATCTAGTCTCTCTTTCTTCAAAATCTCTGCGCCGAGATTATCA	
Sequencing result (SALK_105434C)	425	ATATAGTTATGATCTAGTCTCTCTTTCTTCAAAATCTCTGCGCCGAGATTATCA	453

Supplementary fig. 20 - Alignment of the sequencing results of flanking sequences of T-DNA insertions from *atglc14e-1* (A) and *at4g27480* (B) mutants.

The sequencing result with the primer LBb1.3 (Supplementary table 1) were aligned with the expected GlcAT14E and At4g27480 flanking sequences. The alignment was performed with Geneious Prime (<https://www.geneious.com>).

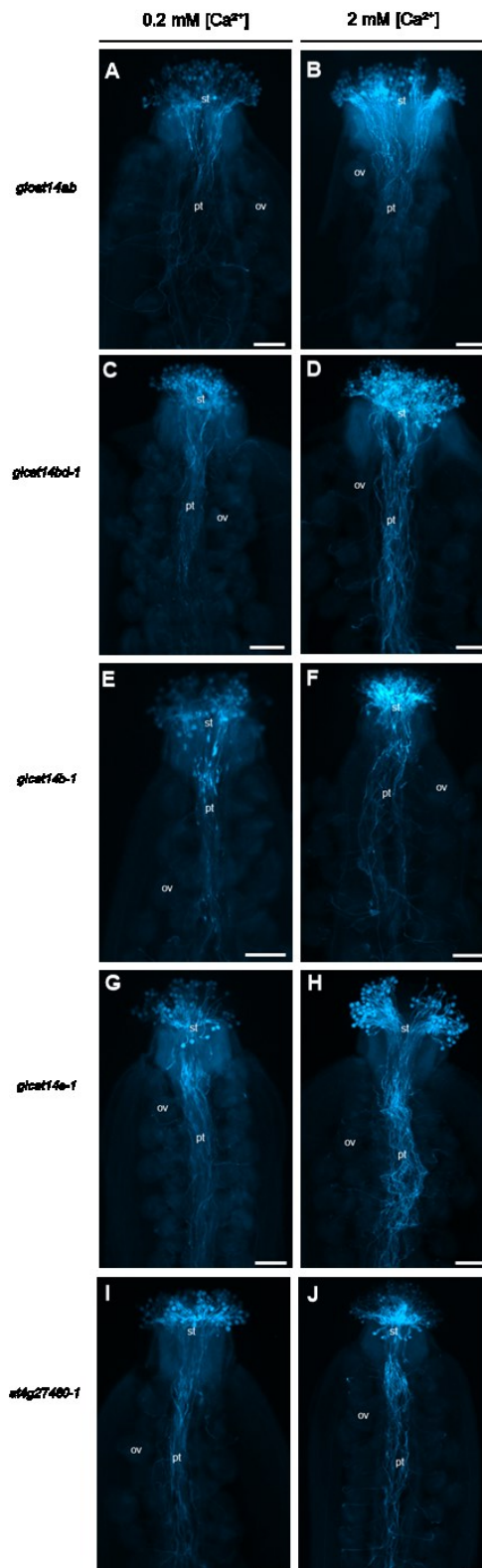


Supplementary fig. 21 - Cleared ovules of Arabidopsis mutant and wt plants grown using hydroponic solutions. Cleared ovules of *atglcat14ab* (A-B), *atglcat14bd-1* (C-D), *atglcat14b-1* (E-F), *atglcat14e-1* (G-H) and *at4g27480-1* (I-J) plants grown with LCNS (left column) and BNS (right column). Abbreviations: cc, central cell; ec, egg cell; fu, funiculus; mi, micropyle. Size bar, 20 μ m.



Supplementary fig. 22 - Pollen from Arabidopsis mutant and wt plants grown using hydroponic solutions stained with Alexander's staining.

Anthers with pollen grains of *glcat14ab* (A-B), *glcat14bd-1* (C-D), *glcat14b-1* (E-F), *glcat14e-1* (G-H) and *at4g27480-1* (I-J) plants grown with LCNS (left column) and BNS (right column). Pink-red stained pollen grains are viable. White arrows indicate green-blue stained non-viable pollen. Size bar, 50 μ m.



Supplementary fig. 23 - Aniline blue staining of Arabidopsis mutant and wt pistils of plants grown using hydroponic solutions.

Pistils of *glcat14ab* (A-B), *glcat14bd-1* (C-D), *glcat14b-1* (E-F), *glcat14e-1* (G-H) and *at4g27480-1* (I-J) plants grown with LCNS (left column) and BNS (right column). Abbreviations: ov, ovule; pt, pollen tube; st, stigma. White arrows indicate stigmas with absence of germinated pollen grains. Size bar, 100 μm.