

**Regulation of mucilage production in
the *Arabidopsis* seed coat through
MADS-box TF family members**

Debra Ruth David

ORCID: 0000-0001-7942-1390

Submitted in total fulfillment of the requirements of the degree
of

Doctor of Philosophy

December 2018

School of BioSciences
The University of Melbourne

Abstract

During Arabidopsis seed development, the epidermal cells of the seed coat produce large quantities of pectinaceous mucilage polysaccharides. Following imbibition, mucilage is released from these cells and forms a thick protective layer around the germinating seed that can be visualised with Ruthenium Red staining. The seed coat epidermis is therefore an attractive system to investigate the production of cell wall polysaccharides, particularly pectin. Previously, mutations in the transcriptional regulator LEUNIG_HOMOLOG (LUH) have been shown to be associated with a mucilage extrusion defect that is caused by reduced expression of the β -galactosidase MUCILAGE MODIFIED2 (MUM2). To better understand the role of LUH in regulating MUM2, RNA-Seq analysis was performed on whole seeds and identified genes that were differentially expressed in *luh* mutants. This transcriptomic analysis not only revealed elevated expression of transcription factors with a known role in epidermal cell differentiation, but also several MADS-box transcription factors that perform roles in floral development and silique shattering such SHATTERPROOF1 (SHP1) and SHP2.

This thesis presents evidence that SHP1, SHP2 and SEPALLATA3 (SEP3) are a new class of regulators involved in mucilage polysaccharides modification, as loss of their activity result in mucilage extrusion defects. These mucilage defects are enhanced in *shp1 shp2* double mutants and correlate with reduced MUM2 expression. While carbohydrate analysis data failed to show an expected increase in galactose residues attached to pectin in *shp1 shp2* and *sep3* mutants, the decrease of MUM2 in these mutants is shown to contribute to their mucilage defects as partial rescue of mucilage defects in *shp1 shp2* and *sep3* was achieved with rescue of MUM2 expression. This is further supported by the use of SRDX repressor motif fusions to SHP2 that not only produced mucilage extrusion defects when introduced into Col-0, but also were correlated with the level of reduced MUM2 levels.

Finally, the positive regulation of *MUM2* by MADS-box TFs was established through a combination of CArG-box motif mutations in the regulatory sequences of *MUM2*, demonstrating the requirement of MADS-box TFs for proper *MUM2* seed coat expression. This was further supported by luciferase-based transactivation assays, which produced increased *MUM2* promoter activity on the addition of SHP1, SHP2 and SEP3. In addition, this positive regulation of *MUM2* by SHP1, SHP2 and SEP3 was shown to occur in a complex with LUH, mediated by the co-regulator SEUSS (SEU), in protein-protein interaction assays such as yeast 2-hybrid (Y2H), Y3H, and bimolecular fluorescence complementation (BiFC).

Overall, this study provides novel, clear evidence that MADS-box TFs can directly bind to *MUM2* and mediate its activation with LUH. Future work includes expanding on and integrating MADS-box TFs into a working model that describes the regulatory pathways controlling mucilage polysaccharides production in the *Arabidopsis* seed coat.

Declaration

This is to certify that:

- i) The thesis comprises only my original work towards this PhD except where indicated in the Preface,
- ii) Due acknowledgement has been made in the text to all other material used,
- iii) The thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices

Debra Ruth David

December 2018

Preface

I acknowledge the hard work of my collaborators who have provided both technical and intellectual contributions throughout this study.

Chapter 3: Identifying and validating differentially expressed genes in *luh* seeds

Momeneh Foroutan (The University of Melbourne)

- Assisted with the RNA-seq analysis for Col and *mum1* at 3, 7, 10 DPA.

Prof. Martina Doblin (University of Technology, Sydney)

- Performed the statistical analysis between the RNA-seq datasets for *luh-4* and *luh-5* at 7 DPA.

Chapter 4: Characterisation of identified MADS-box mutants

Momeneh Foroutan (The University of Melbourne)

- Assisted with the RNA-seq analysis for Col and *shp1 shp2* at 7 DPA.

Cherie Beahan (The University of Melbourne)

- Assisted with the processing and analysis of carbohydrate linkage on Col, *shp1 shp2*, *sep3* and *mum2* mutants

Dr John Golz (The University of Melbourne)

- Supplied constructs used for assessing complementation of *shp1 shp2* mutants

Chapter 5: Defining the mechanism by which MADS-box transcription factors regulate *MUM2*

Dr John Golz (The University of Melbourne)

- Created deletion and mutation constructs used for assessing utility of CARG sites in the *MUM2* intron, used in microscopy and transactivation assays.

- Created SHP2 fused to an SRDX motif that was used for assessing SHP2 as a *MUM2* activator.
- Created BiFC fusion constructs for accessing interactions of MADS-box TFs with components of the LUH regulatory complex.

Joanna Kaptur (Golz Lab,,The University of Melbourne)

- Performed Y2H & Y3H assays.

Overall proofreading of thesis:

John Golz, Monika Doblin and Tony Bacic

Acknowledgements

After years of browsing through completed thesis, my favourite part has always been reading the acknowledgements. Somehow, the gratitude and relief in those pages always inspired me. Now that it is my turn, I must admit that I never understood how hard it is to put down thanks in words. While I was tempted to cop-out and fill this page up with odes to beer instead, B. Kavanaugh has taught me that this may come across as buffoonish and may cost me at future interviews, so I apologise in advanced for this bland rendition.

I would like to acknowledge the efforts of my supervisory team of John, Monika and Tony in guiding me for these four arduous years. While I have made a clear decision to leave academia, I am thankful to especially Mon and Tony as their help have proven invaluable during the last stages of thesis writing. I have said time and again that I cannot express my gratitude enough, and I am sure at this point, you both probably agree with this sentiment and are simultaneously sick of hearing it. I also cannot thank enough my advisory committee of Alex Andrianopoulos and Staffan Persson, whose help I could have never foreseen needing to such a great extent and am so grateful to have. Alex, thank you for intervening when I needed you the most and for sticking your neck out for me. When I say this was a life-saver for me, I mean it literally, as my world, for the better (or worse) part of 2018 was crumbling around me. I truly cannot say how I would have managed to trudge along without your help. Staffan, you are a unicorn – a great PhD supervisor (limited resource), a great advisory member, and just a lovely person overall. I know your students are lucky to have you, and your support to me has been invaluable more than you realise. I would also like to thank the School of BioSciences, in particular Anne Moran and Devi Stuart-Fox, for their listening ear, understanding and support – please continue the good work, and I hope more students will no longer be voiceless and have the strength to speak up and seek your help when needed.

Although some of us have gone off in different continents, I am thankful to Joanne, David, Nadee, Mahmud, Francis, Arca, Joanna, Lucas and Qi Wei during my time in the Golz lab. What can I say, commiseration breeds camaraderie, and there was A LOT of commiseration, wasn't there? Joanna, thank you for watering my plants while I was on holidays and I hope I bribed you well enough for that! I will miss your cakes and if you end up staying in Australia I promise to visit often with munchkins (the boardgame)! I am also very glad to have met Harsh (yes you too David, despite your resemblance to *Harry Potter*), Brianna and Trish who have in some way or another, spoilt me and I do hope to continue having all of you in my life - your friendship has been invaluable, and I cannot stress this enough! I would also like to thank the Genetics/Botany/BioSciences postgrads – too many to list out, but all the same, am grateful for your friendship and gossip sessions. Beyond friends from university, I would like to thank Bashar, aka “rich hobo” and Moody for their spontaneous visits and Bloomberg and GE fuelled meals that add colour (and weight) to my life. I would also like to thank my immediate family for their lack of interest in my thesis, and for not asking me about my thesis submission deadline, as nebulous it can be.

Last, but definitely not the least, I would like to thank my husband, Michael Berger, for his support (both financial and emotional) and for always believing in me even when I make clearly bad decisions, like certain regrettable commitments. Thank you for handling all the numerous questions as to when I will hand in/when will I have a baby, as we both know, I absolutely loathe answering both those questions .

Dedicated to my sanity, may I regain it soon.

Table of contents

Chapter 1

Introduction..... 1

1.1 Arabidopsis seed coat as a model for plant cell wall polysaccharide

biosynthesis and regulation 1

1.1.1 Myxospermy 3

1.1.2 Seed coat development in *Arabidopsis* 4

1.1.3 Composition of mucilage in *Arabidopsis* 6

1.2 Genes involved in mucilage polysaccharides regulation 8

1.2.1 Transcriptional regulation of mucilage polysaccharides: tier 3 TFs 9

1.2.2 Transcriptional regulation of mucilage polysaccharides: tier 2 TFs 10

1.2.3 Transcriptional regulation of mucilage polysaccharides: tier 1 TFs 11

1.2.4 Enzyme effectors of mucilage polysaccharides biosynthesis/modification/secretion.....13

1.3 Groucho (Gro)/Tup1 transcriptional regulators LUH and LUG 17

1.3.1 LUH regulation of *MUM2* in mucilage polysaccharides maturation 19

1.4 The MADS-box TF family 22

1.4.1 MADS-box TF family members in *Arabidopsis* 23

1.4.2 This study.....26

Chapter 2

Materials and methods 46

2.1 Plant materials & methods 47

2.1.1 Plant growth conditions	47
2.1.2 Floral dip transformation	48
2.1.3 Generation of crosses	49
2.1.4 Extraction of genomic DNA from <i>Arabidopsis</i>	50
2.2 General molecular materials & methods	51
2.2.1 Growth media.....	51
2.2.2 Chemicals & reagents	52
2.2.3 Plasmid vectors	52
2.2.4 Bacterial strains and growth conditions	53
2.2.5 Polymerase chain reaction (PCR)	53
2.2.6 Plasmid DNA extraction	54
2.2.7 Restriction digestion	54
2.2.8 Vector dephosphorylation	55
2.2.9 DNA ligation.....	55
2.2.10 DNA sequencing	55
2.2.11 Construct generation	56
2.2.12 Transformation of bacteria.....	57
2.3 Quantitative reverse transcription PCR (qRT-PCR)	58
2.3.1 Tissue isolation for RNA extraction	58
2.3.2 RNA extraction	59
2.3.3 RNA quality control.....	59
2.3.4 qRT-PCR primer design	60
2.3.5 qRT-PCR experimental design & parameters.....	60
2.4 Microscopy	62
2.4.1 Light microscopy	62
2.4.2 Compound fluorescence microscopy	62
2.4.3 Scanning electron microscopy (SEM)	63
2.5 Ruthenium Red staining of mucilage	64

2.6 Yeast-based assays	65
2.6.1 Preparation of competent yeast	65
2.6.2 Co-transformation of yeast Y2HGold (Yeast 2-Hybrid)	65
2.6.3 Sequential co-transformation of yeast Y2HGold (Yeast 3-Hybrid)	66
2.6.4 Protein interaction testing	66
 2.7 Transactivation assay	 67
2.7.1 Infiltration of <i>Nicotiana benthamiana</i>	67
2.7.2 Dual luciferase assay.....	67
 2.8 Bimolecular fluorescence complementation (BiFC)	 69
 2.9 Carbohydrate linkage analysis	 70
2.9.1 Extraction of mucilage from mature seeds	70
2.9.2 Preparation and gas chromatography–mass spectrometry (GC-MS) of extracted mucilage	71
 2.10 RNA-seq	 72
2.10.1 Tissue collection and RNA extraction	72
2.10.2 Mapping of RNA-seq reads	72
2.10.3 Identification of differentially expressed (DE) genes.....	73
2.10.4 Statistical analysis.....	74
 Chapter 3	
Identifying and validating differentially expressed genes in <i>luh</i> seeds.....	76
 3.1 Introduction	 77
 3.2 Results	 81

3.2.1 RNA-seq on <i>luh-5</i> compared to <i>luh-4</i>	81
3.2.1.1 Characterising gene expression changes in <i>luh</i> mutants at different stages of seed development	85
3.2.2 Validating selected TFs using qRT-PCR analysis	88
3.2.3 Investigating expression of selected TFs using promoter-reporter analysis	91
3.2.4 Examining mucilage defects in mutants of candidate <i>MUM2</i> repressors.....	95
3.2.4.1 Genetic interactions	97
3.2.5 Identification of seed coat-specific promoters	98
3.2.5.1 Characterisation of seed coat-specific promoter activity	99
3.2.5.2 Utility of seed coat specific promoters driving artificial micro-RNAs..	101
3.2.6 Analysis of <i>AG</i> knockdown in seed coat	102
3.3 Discussion	104
3.3.1 Experimental discrepancies between <i>luh-4</i> and <i>luh-5</i>	104
3.3.2 The role of <i>LUH</i> in the developing seed coat	105
3.3.3 The roles of MADS-box TFs in the seed coat	108
3.3.4 The utility of <i>GILTpro</i> as a seed coat-specific tool	110
3.3.5 <i>AG</i> has roles in controlling seed size in <i>Arabidopsis</i>	112
3.3.6 Conclusion	114
 Chapter 4	
Characterisation of identified MADS-box TF mutants	150
 4.1 Introduction	151
 4.2 Results	103
4.2.1 Complementation of MADS-box TF mutants	154
4.2.2 Examination of the epidermal layer in MADS-box TF mutants.....	156
4.2.3 Expression of <i>MUM2</i> in MADS-box TF mutants.....	157
4.2.3.1 Heterologous expression of <i>MUM2</i> in MADS-box TF mutants.....	158

4.2.4 Expression of <i>LUH</i> in MADS-box TF mutants	159
4.2.4.1 Heterologous expression of <i>LUH</i> in MADS-box TF mutants	160
4.2.5 Carbohydrate analysis of Col-0, <i>shp1 shp2</i> and <i>sep3</i> mucilage.....	162
4.2.6 Differentially expressed genes in <i>shp1 shp2</i> and Col-0 whole seeds	163
4.3 Discussion	167
4.3.1 Mucilage defects in <i>shp1 shp2</i> and <i>sep3</i> are due to downstream factors.....	167
4.3.2 <i>MUM2</i> , but not <i>LUH</i> expression levels correlate well across RNA-seq and qRT-PCR.....	168
4.3.3 MADS-box TFs SHP1, SHP2 and SEP3 may act as <i>MUM2</i> activators	170
4.3.4 Issues associated with carbohydrate analysis of <i>shp1 shp2</i> and <i>sep3</i> mutants	172
4.3.5 Other factors involved in mucilage polysaccharides biosynthesis/modification that could be contributing to mucilage defects seen in <i>shp1 shp2</i> mutants	173
4.3.6 Conclusion	175

Chapter 5

Defining the mechanism by which MADS-box transcription factors regulate *MUM2*..... 194

5.1 Introduction 195

5.2 Results 199

5.2.1 Analysis of CArG box and CL motifs within the regulatory regions of the <i>MUM2</i> intron	199
5.2.1.1 Analysis of CArG box, CL and MYB binding site deletions	200
5.2.1.2 Analysis of CArG box and CL motif mutations	202
5.2.2 Candidate MADS-box TFs binding partners to identified CArG box and CL motifs in the <i>MUM2</i> intron	205
5.2.3 SHP2 regulates <i>MUM2</i> via CArG box/CL motifs in <i>Il</i>	207

5.2.3.1 SHP2-SRDX fusions induce mucilage extrusion defects in transgenic lines	207
5.2.3.2 MADS-box TFs activate <i>MUM2</i> in transactivation assay	209
5.2.4 Interactions of SHP1, SHP2 with LUH/SEU	212
5.2.4.1 Physical interactions between MADS-box TFs and components of the LUG/LUH regulatory complex	213
5.2.4.2 Physical interactions between MADS-box TFs and the LUG/LUH regulatory complex <i>in planta</i>	215
5.3 Discussion	219
5.3.1 Mutations of functional CArG box and CL motifs support activation of <i>MUM2</i> by MADS-box TFs	219
5.3.2 MADS-box TFs are <i>MUM2</i> activators	221
5.3.3 LUH directly activates <i>MUM2</i> via MADS-box TFs	224
5.3.4 Potential role of MYBs in <i>MUM2</i> regulation	227
5.3.5 Conclusion	228
 Chapter 6	
Final discussion and conclusions	260
 6.1 The roles of MADS-box TFs in <i>Arabidopsis</i> seeds	261
6.2 Future directions	264
 Chapter 7	
References	268
 Appendices	298

List of tables

Table 1.1	Summary of effector enzymes involved in mucilage polysaccharides production, modification and secretion.	28
Table 1.2	Summary of TF families elevated in <i>luh</i> compared to Col-0 in 7 DPA seeds.	31
Table 3.1	List of TFs elevated in <i>luh-4</i> and <i>luh-5</i> whole seeds at the linear cotyledon stage of development.	116
Table 3.2	Expression pattern of seed coat expressed TFs that are elevated in <i>luh</i> mutants.	118
Table 3.3	List of seed coat expressed TFs that are elevated in <i>luh-5</i> across early to late seed developmental stages.	121
Table 3.4	Expression of LUH-regulated TFs in <i>luh</i> seeds at early and late developmental stages.	125
Table 3.5	qRT-PCR analysis of seed coat-specific TFs in <i>luh-4</i> mutants.	126
Table 3.6	Mucilage extrusion phenotypes of lines transformed with amiRNAs targeting <i>LUH</i> or <i>MUM2</i> and controlled by various seed coat-specific promoters.	127
Table 3.7	Phenotypes of Col-0 lines transformed with the <i>35S_{pro}::amiR-AG</i> construct.	128
Table 3.8	Phenotypes of seeds derived from transgenic <i>GILT_{pro}::amiR-AG</i> lines.	129
Table 4.1	Phenotypes of transgenic <i>shp1 shp2</i> and <i>sep3</i> mutant lines.	177
Table 4.2	Phenotypic rescue of MADS-box TF mutants by <i>MUM2</i> or <i>LUH</i>	178
Table 4.3	Sugar linkage composition of extracted mucilage from Col-0, <i>mum2</i> , <i>shp1 shp2</i> and <i>sep3</i> seeds.	179
Table 4.4	Calculated polysaccharide composition of extracted mucilage from Col-0, <i>mum2</i> , <i>shp1 shp2</i> and <i>sep3</i> seeds	181

Table 4.5	List of differentially expressed genes encoding cell wall-related enzymes in <i>shp1 shp2</i> seeds at the linear cotyledon stage.	182
Table 5.1	Analysis of <i>MUM2</i> promoter/intron-reporter constructs with various deletions.	230
Table 5.2	Summary of CArG box, CL motifs and MYB binding sites deletions and mutations.	232
Table 5.3	Summary of CArG box motifs mutations.	234

List of figures

Figure 1.1	Double fertilisation in <i>Arabidopsis</i>	32
Figure 1.2	Development of the seed coat layers in <i>Arabidopsis</i>	33
Figure 1.3	Stages of MSCs development in <i>Arabidopsis</i>	34
Figure 1.4	Composition of <i>Arabidopsis thaliana</i> seed coat mucilage.	35
Figure 1.5	Different types of pectin found in <i>Arabidopsis</i>	36
Figure 1.6	Proposed transcriptional network regulating MSC differentiation and mucilage polysaccharides production in <i>Arabidopsis</i>	37
Figure 1.7	Structure of Gro/Tup1 family members in <i>Arabidopsis</i>	38
Figure 1.8	The LUG/LUH regulatory complex in <i>Arabidopsis</i>	39
Figure 1.9	Proposed models of <i>MUM2</i> regulation by LUH.	40
Figure 1.10	Proposed model of <i>MUM2</i> regulation in the seed in seeds.	41
Figure 1.11	Structures of MADS Type I & II TFs.	42
Figure 1.12	Roles of MADS-box TF complexes during plant development. ..	43
Figure 1.13	ABCDE model of floral development.	44
Figure 3.1	<i>MUM2</i> promoter (+11) activity and predicted activity of a <i>MUM2</i> repressor.	130
Figure 3.2	Comparison of DEGs present in <i>luh-4</i> , <i>luh-5</i> and <i>mum2-1</i> whole seeds at the linear cotyledon stage of development.	132
Figure 3.3	Comparison of DEGs present in <i>luh-5</i> at different stages of seed development.	133
Figure 3.4	Overview of the selection process of TFs elevated in <i>luh-4</i> for further examination.	134
Figure 3.5	Promoter-reporter activity of MADS-box TFs in the developing seed coat.	136
Figure 3.6	Promoter-reporter activity of early differentiation TFs in the developing seed coat.	138

Figure 3.7	Promoter-reporter activity of candidate <i>MUM2</i> repressors in the developing seed coat.	139
Figure 3.8	Ruthenium Red staining of select MADS-box TF mutant seeds following various treatments.	141
Figure 3.9	Ruthenium Red staining of select MADS-box TF mutant seeds.	142
Figure 3.10	Heatmap showing the expression of putative seed coat specific genes identified in the Belmonte dataset.	143
Figure 3.11	Relative expression of putative seed coat specific genes in Col-0 and <i>luh-4</i> mutant seed coats at the linear cotyledon stage.	144
Figure 3.12	Promoter-reporter activity of putative seed coat-specific genes at various stages of seed coat development.	145
Figure 3.13	Magnified view of promoter-reporter activity in the seed coat at the linear cotyledon stage.	146
Figure 3.14	Activity of the <i>GILT</i> promoter in other plant tissues.	147
Figure 3.15	Average weight of seeds derived from plants expressing the <i>GILT_{pro}::amiR-AG</i> construct.	148
Figure 4.1	Categorisation of T ₂ seeds for complementation analysis.	185
Figure 4.2	Seed coats of mature MADS-box TF mutant seeds as seen by SEM.	186
Figure 4.3	Expression of <i>MUM2</i> in MADS-box TF mutants.	187
Figure 4.4	<i>MUM2</i> promoter I1- <i>VENUSn</i> reporter activity in seed coats of phenotypically wildtype, <i>shp1 shp2</i> and <i>sep3</i> mutants.	188
Figure 4.5	Mucilage release phenotypes of seeds derived from transgenic MADS-box TF mutants with elevated levels of <i>MUM2</i> or <i>LUH</i>	190
Figure 4.6	Expression of <i>LUH</i> in MADS-box TF mutants.	191
Figure 4.7	Expression of <i>MUM2</i> in <i>shp1 shp2</i> lines carrying <i>35S_{pro}::LUH</i> . .	192
Figure 5.1	Potential TF binding sites in region A of the <i>MUM2</i> intron 1.	236
Figure 5.2	Activity of <i>MUM2</i> promoter/intron-reporter deletions constructs within the developing seed coat of T ₁ transgenic lines.	237

Figure 5.3	Enlarged view of <i>MUM2</i> promoter/intron-reporter deletions constructs within the epidermal layer of the seed coat.	239
Figure 5.4	Activity of <i>MUM2</i> promoter/intron-reporter mutation constructs within the developing seed coat of T ₁ transgenic lines.	240
Figure 5.5	Enlarged view of <i>MUM2</i> promoter/intron-reporter mutations constructs within the epidermal layer of the seed coat.	242
Figure 5.6	Expression pattern of MADS-box TFs in the seed coat and embryos at different developmental stages.	243
Figure 5.7	Mucilage extrusion phenotypes of <i>35S_{pro}::SHP2::SRDX</i> lines. ...	244
Figure 5.8	Floral defects of <i>35S_{pro}::SHP2::SRDX</i> lines.	245
Figure 5.9	Expression of <i>MUM2</i> in <i>35S_{pro}::SHP2::SRDX</i> lines.	246
Figure 5.10	Constructs generated for transactivation assays.	247
Figure 5.11	Transactivation assays on the <i>MUM2</i> promoter/intron reporter. .	248
Figure 5.12	Transactivation assay performed on the <i>MUM2</i> promoter/intron using different combinations of MADS-box TFs.	249
Figure 5.13	Promoter-reporter activity of LUG/LUH regulatory complex components in the developing seed coat.	250
Figure 5.14	Visualising maximal reporter activity in seeds of lines transformed with the LUH and SEU promoter-reporters.	251
Figure 5.15	Y2H assays between MADS-box TFs and components of the LUG/LUH regulatory complex.	252
Figure 5.16	Y3H assays to assess the capacity of SEP3 to function as a ternary factor.	253
Figure 5.17	BiFC interactions between MADS-box TFs and LUH.	254
Figure 5.18	BiFC interactions between MADS-box TFs and SEU.	256
Figure 5.19	BiFC interactions between SEP3, SHP1 and SHP2.	257
Figure 5.20	BiFC interactions between SEP3, SHP1, SHP2 and components of the LUH regulatory complex.	258

Abbreviations

3-AT	3-aminotriazole
ABF3	ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR3
ABS	ARABIDOPSIS BSISTER
AD	GAL4 activation domain
AG	AGAMOUS
AGL	AGAMOUS-LIKE
ALC	ALCATRAZ
amiRNA	Artificial microRNA
ANT	AINTEGUMENTA
AP1-3	APETALA1-3
ARF2	AUXIN RESPONSE FACTOR2
ASD1/2	ALPHA-L-ARABINOFURANOSIDASE1/2
ASG4	ALTERED SEED GERMINATION4
ATHB1	ARABIDOPSIS THALIANA HOMEODOMAIN1
BAN	BANYULS
BD	GAL4 DNA binding domain
BGAL6	β -GALACTOSIDASE6
BGLU15/45	BETA GLUCOSIDASE15/45
bHLH	basic helix-loop-helix
BiFC	Biomolecular fluorescence complementation
BXL1	β -XYLOSIDASE1
bZIP44	BASIC LEUCINE-ZIPPER44
CaMV	Cauliflower mosaic virus
CESA5-19	CELLULOSE SYNTHASE5-19
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation sequencing

CO1	CONSTANS
COBL2	COBRA-LIKE2
CPC	CAPRICE
CSLA2	CELLULOSE SYNTHASE-LIKE A2
CZF2	COLD INDUCED ZINC FINGER PROTEIN2
DEF	DEFICIENS
DF1	DE1 BINDING FACTOR1
DNA	Deoxyribonucleic acid
DPA	Days post anthesis
DPI	Days post infection
EAR	Ethylene-responsive element binding factor-associated amphiphilic repression
ECH	ECHIDNA
EDA29	EMBRYO SAC DEVELOPMENT ARREST29
EMS	Ethyl methanesulfonate
EDTA	Ethylenediaminetetraacetic acid
EGL3	ENHANCER OF GLABRA3
EMSA	Electrophoretic mobility shift assay
EXO70A1	EXOCYST SUBUNIT EXO70 FAMILY PROTEIN A1
FLA7/11/15	FASCICLIN-LIKE ARABINOGALACTAN PROTEINS7/11/15
FLC	FLOWERING LOCUS C
FLY1	FLYING SAUCER1
FRA8	FRAGILE FIBER8
GATL5	GALACTURONOSYLTRANSFERASE-LIKE5
GAUT11	GALACTURONOSYLTRANSFERASE11
GL2	GLABRA2
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
HAT	Histone acetylases
HDAC	Histone deacetylases

HEC3	HECATE3
HG	Homogalacturonan
IBH1	IL11 BINDING bHLH1
IND	INDEHISCENT
IRX9/10/14	IRREGULAR XYLEM9/10/14
KNAT7	KNOTTED ARABIDOPSIS THALIANA7
LFY	LEAFY
LUH	LEUNIG_HOMOLOG
MAF1	MADS AFFECTING FLOWERING1
MCM1	MINICHROMOSOME MAINTENANCE1
MEF2	MYOCYTE ENHANCER FACTOR2
miR172	microRNA172
MSC	Mucilage secreting cells
MSR1	MANNAN SYNTHESIS RELATED1
MUCI10/21/71	MUCILAGE-RELATED10/21/71
MUM1-5	MUCILAGE MODIFIED1-5
MYB5/23/52/61/75	MYB DOMAIN PROTEIN5/23/52/61/75
NAC019	NAC DOMAIN CONTAINING PROTEIN19
NARS1-2	NAC-REGULATED SEED MORPHOLOGY1-2
NF-YA7	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT A-7
PAE1/2	PECTIN ACETYLESTERASE1/2
PCR	Polymerase chain reaction
PG2	POLYGALACTURONASE2
PI	PISTILLATA
PIF1	PHYTOCHROME-INTERACTING FACTOR1
PME	PECTIN METHYLESTERASE
PMEI6/14	PECTIN METHYLESTERASE INHIBITOR6/14
PRX36/56	PEROXIDASE36/56
qRT-PCR	quantitative reverse transcription PCR
RD20	RESPONSIVE TO DESICCATION20
RG1	Rhamnogalacturonan I

RHM2/3	RHAMNOSE BIOSYNTHESIS2/3
RNA	Ribonucleic acid
rpm	Revolutions per minute
SBT1.7	SUBTILISIN-LIKE SERINE PROTEASE1.7
SC	Seed coat
SCL3/8	SCARECROW-LIKE3/8
SDS	Sodium dodecyl sulphate
SEC8	SUBUNIT OF EXOCYST COMPLEX8
SEM	Standard error of the mean
SEP1-4	SEPALLATA1-4
SEU	SEUSS
SHP1/2	SHATTERPROOF1/2
SOC1	SUPPRESSOR OF OVEREXPRESSION OF CO1
SOS5	SALT OVERLY SENSITIVE5
SPT	SPATULA
STK	SEEDSTICK
SVP	SHORT-VEGETATIVE PHASE
TBL38	TRICHOME BIREFRINGENCE-LIKE38
T-DNA	Transfer DNA
TE	Tris, EDTA buffer
TF	Transcription factor
TPL	TOPLESS
TPR1-4	TOPLESS-RELATED1-4
TRAP-seq	Translating ribosome affinity purification
TREAT	t-tests Relative to a Threshold
TT2/8/10	TRANSPARENT TESTA2/8/10
TTG1-2	TRANSPARENT TESTA GLABRA1-2
URP3	UAS-TAGGED ROOT PATTERNING3
VDD	VERDANDI
VNI2	VND-INTERACTING2
WRKY47	WRKY-PROTEIN46

Y2H	Yeast two-hybrid
Y3H	Yeast three-hybrid
YIB4A/B	YPT/RAB GTPASE INTERACTING PROTEIN 4A/B

Chapter 1

Introduction

Introduction

1.1 Arabidopsis seed coat as a model for plant cell wall polysaccharide biosynthesis and regulation

Plant cell walls, which are rich in polysaccharides, are important for plant growth and development (Caffall & Mohnen, 2009; Doblin et al., 2010). The cell wall not only provides structural support, but is vital for withstanding environmental stress and for protection against pathogens and wounding (Caffall & Mohnen, 2009). Moreover, plant cell walls are abundant in polysaccharides, such as cellulose and pectin, which are extensively used in agro-industrial applications. For example, cellulose is used in bioenergy, agriculture and manufacturing for producing commercially viable products such as biofuels and spinnable fibres such as cotton (Doblin et al., 2014; Gomez et al., 2008), while pectin is used in industries such as pharmaceuticals and food production as dietary fat substitutes, edible papers and gelling agents (Thakur et al., 1997; Voragen et al., 2009). In addition, pectin such as homogalacturonan (HG) and rhamnogalacturonan 1 (RG1) are important sources of soluble dietary fibres and have been associated with improved human health benefits such as increased satiety, lower blood glucose and cholesterol (Adam et al., 2015; Brouns et al., 2012; Tiwary et al., 1997; Voragen et al., 2009). Thus, the growing demand for plant-based polysaccharide products makes the production of plants with designer cell walls an attractive prospect (Doblin et al., 2014). However, this process requires an in-depth understanding of plant cell wall regulation which is largely lacking; currently, our understanding of the biosynthesis and regulation of plant cell wall polysaccharides remains the major bottleneck to the manipulation of walls for biotechnological applications (Doblin et al., 2014; Harholt et al., 2010; Li et al., 2014; Mohnen, 2008).

Studying plant cell wall polysaccharide regulation to understand processes involved in its synthesis, modification and secretion is complicated by a high degree of redundancy in gene regulatory pathways, the differences in composition at various

growth stages and between tissues and the presence of many polysaccharides and glycoproteins coexisting in walls at any one time (Arsovski et al., 2010; Haughn & Western, 2012). Some of these complexities have been overcome by studying the production, modification and secretion of seed mucilage polysaccharides in the model plant *Arabidopsis thaliana*. The mucilage in this species is non-essential under ideal laboratory growth conditions (Haughn & Western, 2012), making it highly amenable to genetic analysis as mutants either lacking or with altered mucilage are viable. The availability of a full genome sequence, many mutant collections and the ease of detection of mucilage mutants has provided unparalleled insight into the production of pectins (HG and RG1), as well as hemicelluloses (galactoglucomannans, xyloglucans, heteroxylans) and cellulose (Arsovski et al., 2010; Western et al., 2000). Collectively, these advantages make *Arabidopsis* an attractive system for further investigation of the biosynthesis, modification and secretion of cell wall polysaccharides, particularly pectin.

1.1.1 Myxospermy

Seed bearing plants, spermatophytes, are divided into angiosperms that produce seeds encased within fruits, and gymnosperms that produce seeds that are exposed to the external environment. Angiosperms encompass both dicotyledons, producing two embryonic leaves, which include plant families such as *Brassicaceae*, *Asteraceae*, *Cistaceae*, *Solanaceae*, *Linaceae* and *Plantaginaceae*, and monocotyledons, producing one embryonic leaf, which include plant families such as *Poaceae*, *Arecaceae*, *Musaceae*, *Zingiberaceae* and *Amaryllidaceae* (Jordan, 2006). A trait shared by some angiosperms is the ability to extrude a polysaccharide-rich mucilage from the testa (epidermal layer of the seed coat) to surround the seed upon imbibition in water, termed myxospermy (Western et al., 2001). This involves the accumulation and desiccation of mucilage polysaccharides synthesised in the testa of the mature seed, that on contact with water, expands and breaks through the cell wall for extrusion to occur (Macquet et al., 2007b). The role of mucilage is not well

understood but is suggested to aid in germination, pathogen defence, dispersal and control the length of dormancy (Macquet et al., 2007b). Variation in mucilage composition of *A. thaliana* accessions suggests that myxospermy is an adaptive trait that is responsive to different environments. Although its adaptive value is unknown, it is thought to play a role in water-based seed dispersal (Macquet et al., 2007b; Saez-Aguayo et al., 2014). Regardless of the precise role of mucilage, myxospermous plants provide a good source of readily available and extractable polysaccharides and glycoproteins in mucilage that is amenable to genetic manipulations without deleterious effects on the seed.

1.1.2 Seed coat development in *Arabidopsis*

In angiosperms such as *Arabidopsis*, seed development begins with double fertilisation of the ovule. This process is triggered by the pollination of the carpel, resulting in pollen germination and pollen tube growth towards the ovules, culminating in the release of two sperm cells from the pollen tube tip following contact with the ovule (Faure & Dumas, 2001; Sprunck & Dresselhaus, 2015). Within the ovule, one of these sperm cells fertilises the egg cell to produce a diploid zygote that develops into an embryo, while the other fertilises the central cell to form the triploid endosperm (Fig. 1.1) (Faure & Dumas, 2001; Sprunck & Dresselhaus, 2015). Unlike the embryo and endosperm, the seed coat is entirely maternally derived and is formed by the inner and outer integuments of the ovule (Fig. 1.1), once auxin signals from the endosperm release the epigenetic repression of seed coat development (Figueiredo et al., 2016).

The formation of the maternally derived seed coat is a highly dynamic process. In the early stages of seed coat development (heart stage; 3 days post anthesis [DPA]), the seed coat is formed from the inner and outer integument cells of the ovule that give rise to five distinct layers - three layers derived from the inner integument (ii1-3) and two layers derived from the outer integument (oi1-2) (Fig. 1.2) (Nakaune et al., 2005).

The inner integument consists of the tegmen (ii1), as well as two other layers (ii2-3) that undergo cell death in mid-seed coat development stages (linear cotyledon stage; 7-8 DPA) and eventually are crushed into a single layer called the tegmen by late seed coat development stages (mature green stage; 10-13 DPA), producing a single pigmented layer where proanthocyanidins are synthesised to produce tannins (Fig. 1.2) (Dean et al., 2011; Nakaune et al., 2005). Similarly, the outer integument forms two layers; the water impermeable palisade layer (oi1) and the epidermis (oi2) that eventually forms a single layer called the testa (oi) composed of mucilage secreting cells (MSCs) (Fig. 1.2) (Dean et al., 2011; Nakaune et al., 2005).

MSCs are specialised epidermal cells that undergo a dramatic differentiation process leading to the production and secretion pectinaceous mucilage. Morphological changes at different stages of seed development can be observed by histological sections and are frequently staged according to embryonic development (Arsovski et al., 2010; Dean et al., 2011; Francoz et al., 2015). The early stages of seed development (pre-globular to late heart stage; 0-4 DPA) are characterised by a large and rapid increase in cell size up to the torpedo stage (5-6 DPA) that is also associated with the accumulation of starch granules in the cytoplasm (Fig. 1.3) (Western et al., 2000). This is followed by the accumulation of mucilage polysaccharides in the apoplastic space located adjacent to the radial cell walls at the linear cotyledon stage (7-8 DPA). As mucilage polysaccharides accumulates, it eventually forces the cytoplasm into a column located in the centre of the cell (Fig. 1.3) (Western et al., 2000). Secondary cell wall material is subsequently deposited into this column from the mature green stage (10-13 DPA) onwards. Eventually, the cytoplasm is completely replaced with a cell wall, resulting in the formation of a volcano-shaped structure called the columella and marks death of the MSC (Fig. 1.3) (Arsovski et al., 2010; Western et al., 2000). During seed maturation (beyond 13 DPA), the seed begins to desiccate, resulting in the dehydration of mucilage polysaccharides in the radial pockets surrounding the columella. On seed imbibition, rehydration of mucilage polysaccharides results in a rapid increase in its volume due to swelling. The swollen mucilage polysaccharides generates sufficient force to rupture the radial cell walls,

allowing mucilage release and the formation of a gelatinous mucilage layer encapsulating the seed (Kunieda et al., 2013; Western et al., 2000).

1.1.3 Composition of mucilage in *Arabidopsis*

Extruded mucilage encapsulating the seed can be easily visualised with Ruthenium Red, a dye that binds to acidic polysaccharides, such as pectin (Western et al., 2001; Western et al., 2000; Western et al., 2004). In wildtype *Arabidopsis* seeds, Ruthenium Red staining reveals two distinct layers of mucilage that can be recognised following extrusion; the outer soluble layer and the inner adherent layer (Fig. 1.4) (Macquet et al., 2007a). The outer soluble layer represents approximately 65% of total mucilage and is characterised by its loose attachment to the seed which is readily dispersed on agitation in water, while the inner adherent layer represents approximately 35% of total mucilage that stains strongly with Ruthenium Red and remains tightly tethered to the seed (Fig. 1.4) (Macquet et al., 2007a). Due to their different composition and macromolecular properties, the outer soluble and inner adherent layers require extraction using different methods to determine their carbohydrate composition. Agitation in water is sufficient for extraction of the outer soluble layer, while a combination of heat, strong alkali, agitation and possibly sonication is necessary for the extraction of the inner adherent layer that is attached to MSC walls (Macquet et al., 2007a; Walker et al., 2011; Zhao et al., 2017).

Various carbohydrate analyses reveal that the outer soluble layer mostly consists of unbranched RG1, a disaccharide backbone of repeating units of alternating α -1, 4-linked galacturonic acid (4-GalA) and β -1, 2-linked rhamnose (2-Rha) residues (Dean et al., 2007; Macquet et al., 2007a; Walker et al., 2011) (Fig. 1.4 & 1.5). The presence of RG1 is also corroborated by extensive digestion by rhamnogalacturonan hydrolases, but not endo-polygalacturonases (Macquet et al., 2007b) (Fig. 1.4).

In contrast, the inner adherent layer is more diverse and contains both unbranched and branched RG1, as well as other polysaccharides in minor amounts such as HG, cellulose and hemicelluloses (Arsovski et al., 2009a; Dean et al., 2007; Doblin et al., 2010; Macquet et al., 2007b; Voiniciuc et al., 2015c). RG1 can be branched at the Rha residues at the C-4 position by attachment of either Ara, Gal, or arabinogalactan (AG) side chains while HG, a homopolymer of galacturonic acid (4-GalA), can be methyl-esterified at the C-6 position to which disrupts Ca^{2+} crosslinking and acetylated at the O-2 and O-3 positions to increase its solubility (Fig. 1.5) (Doblin et al., 2010; Harholt et al., 2010; Saez-Aguayo et al., 2013; Wolf et al., 2009). Immunolabelling with monoclonal antibodies reveals two distinct domains within the inner adherent layer - the inner domain (region closest to the seed) that is strongly labelled with JIM5, indicating the presence of moderately methyl-esterified HG and the outer domain (region furthest from the seed), which is labelled by JIM7 and indicates the presence of highly methyl-esterified HG (Fig. 1.4) (Macquet et al., 2007a). In addition, Ara and Gal, in the form of arabinans/galactans/AGs, are detected in the outer domain of the inner adherent layer, indicative that branched RG1 is primarily localised in the outer domain (Fig. 1.4 & 1.5). In terms of minor mucilage polysaccharides, calcofluor, a fluorescent probe for β -glucans, reveals that cellulose microfibrils are either found diffused within the inner mucilage, or arranged as cellulosic rays in the inner domain of the inner adherent layer and likely form a matrix to anchor RG1 to the seed coat and provide structural support for mucilage (Griffiths & North, 2017; Haughn & Western, 2012; Macquet et al., 2007a; Voiniciuc et al., 2015c). Additionally, interacting with cellulose microfibrils are hemicelluloses such as galactoglucomannans, xyloglucan and xylan, that have recently been detected in the inner domain of the inner adherent layer and are thought to help anchor mucilage to the seed coat (Arsovski et al., 2009a; Ralet et al., 2016; Voiniciuc et al., 2015a; Voiniciuc et al., 2015b; Yu et al., 2014).

1.2 Genes involved in mucilage polysaccharides regulation

A short incubation of seeds with Ruthenium Red allows for rapid screening of *Arabidopsis* lines displaying altered patterns of mucilage staining and/or mucilage release (Western et al., 2000). This has led to the identification of many genes involved in the regulation, biosynthesis, modification and secretion of polysaccharides observed in seed mucilage. For example, an ethyl methanesulfonate (EMS) mutagenesis screen identified several lines displaying a range of mucilage extrusion defects following Ruthenium Red staining (Western et al., 2001). Termed *mucilage modified* (*mum*), two lines (*mum1* and *mum2*) produced seeds that failed to release mucilage, whereas *mum4* seeds displayed reduced mucilage extrusion, while *mum3* and *mum5* seeds release mucilage with reduced adherence (Western et al., 2001). Mapping some of these mutations led to the identification of genes encoding transcriptional regulators, such as MUM1/LEUNIG_HOMOLOG (LUH) (Huang et al., 2011), effector enzymes involved in mucilage modification, such as MUM2/ β -GALACTOSIDASE6 (BGAL6) (Dean et al., 2007) and MUM4/RHAMNOSE BIOSYNTHESIS2 (RHM2) (Western et al., 2004), effector enzymes involved in mucilage attachment, such as MUM3/CELLULOSE SYNTHASE5 (CESA5) (Sullivan et al., 2011), and effector enzymes involved in the production of minor mucilage components, such as MUM5/MUCILAGE-RELATED21 (MUCI21) (Ralet et al., 2016). As more genes associated with mucilage polysaccharides biosynthesis, secretion and modification have been identified, they have been assigned to two broad categories: upstream transcription factors (TFs) that control seed coat development and the downstream effector enzymes that have direct impact on mucilage extrusion, and are collectively called “MSC tool box genes” (Francoz et al., 2015). A recent review proposes reclassifying the upstream TFs into a hierarchical model with three regulatory levels (tiers 1-3); tier 1 TFs directly regulate effector enzymes involved in the biosynthesis/modification/secretion of mucilage polysaccharides, tier 2 TFs regulate both effector enzymes and tier 1 TFs, and tier 3 TFs modulate lower tier TFs and effector enzymes (Golz et al., 2018) (Fig. 1.6). These tiers are described in the following section.

1.2.1 Transcriptional regulation of mucilage polysaccharides: tier 3 TFs

TFs involved in the development of the epidermal layer of the seed coat (MSCs) lie in the highest tier of transcriptional regulation in the three-tier model (Fig. 1.6) (Golz et al., 2018). This includes TFs that are expressed in early seed coat development such as NAC-REGULATED SEED MORPHOLOGY1 (*NARS1*), *NARS2*, *APETALA2* (*AP2*), *TRANSPARENT TESTA GLABRA1* (*TTG1*), *ENHANCER OF GLABRA3* (*EGL3*), *TRANSPARENT TESTA8* (*TT8*), *TT2*, *MYB DOMAIN PROTEIN5* (*MYB5*) and *MYB23* (Baudry et al., 2004; Gonzalez et al., 2009; Kunieda et al., 2008; Li et al., 2009; Ohto et al., 2009; Ranocha et al., 2014) (Fig. 1.6). As these TFs regulate early stages of MSC differentiation, their loss of activity results in the MSCs lacking the characteristic hexagonal shape, columella and mucilage.

For example, *NARS1* and *NARS2* are expressed in the outer integuments in early seed coat development and act redundantly to regulate differentiation of MSCs (Kunieda et al., 2008). Double mutant *nars1 nars2*, but not single mutants, exhibit abnormal seed shape, deformed MSCs with no columellae and thickened radial cell walls, and are unable to produce mucilage polysaccharides (Kunieda et al., 2008). *NARS1* and *NARS2* have also been shown to activate *PEROXIDASE36* (*PER36/PRX36*), a gene encoding a downstream effector enzyme that is proposed to enable mucilage extrusion through the modification of the MSC cell wall (Kunieda et al., 2013). However, reduced *PER36* expression does not explain the absence of mucilage polysaccharides production in the *nars1 nars2* double mutant, indicating that additional genes involved in mucilage polysaccharides production must be regulated by these TFs and therefore *NARS1* and *NARS2* are classified as tier 3 TFs (Voiniciuc et al., 2015c). *AP2*, first identified as a floral mutant, directs differentiation of the outer integument cells to form the testa (Western et al., 2001). Thus, *ap2* mutant seeds display developmentally defective MSCs lacking columella formation that fail to fully differentiate and therefore lack normal mucilage polysaccharides production and columellae structures (Jofuku et al., 1994; Western et al., 2001). Lack of mucilage polysaccharides production in *ap2* mutants may be explained by the loss of *MUM4*,

an UDP-L-Rha synthase that is involved in Rha production used in RG1 synthesis, which is indirectly upregulated by AP2 and TTG1 through activation of tier 2 TF GLABRA2 (GL2) (Western et al., 2004) (Fig. 1.6). Besides activating GL2, TTG1, a WD-40 repeat protein, forms a multimeric WD40-bHLH-MYB complex with basic helix-loop-helix (bHLH) proteins EGL3 and TT8 and promotes differentiation of the testa, as well as the production of proanthocyanins in the inner endothelial layer through the activation of *BANYULS* (*BAN*) (Baudry et al., 2006; Baudry et al., 2004). This WD40-bHLH-MYB complex also activates the downstream tier 2 TF *TTG2* to modulate the downstream effector enzyme *GALACTURONOSYLTRANSFERASE-LIKE5* (*GATL5*), which is necessary for proper mucilage polysaccharides production and adherence to the seed coat (Kong et al., 2013) (Fig. 1.6). In addition to forming WD40-bHLH-MYB complexes, TTG1 is also able to promote proper development of MSCs by associating with partially redundant TFs MYB5 and TT2 (Gonzalez et al., 2009). This partial redundancy in the seed coat is evident as *myb5* mutants, *myb5 tt2* double mutants, but not *tt2* mutants display defective MSC formation (Gonzalez et al., 2009). Besides the double mutant *myb5 tt2*, double mutants *myb5 myb23* also possess irregular seed shape and MSCs with flattened columellae and decreased mucilage compared to wildtype, indicating that MYB23 may have roles in epidermal differentiation (Gonzalez et al., 2009; Li et al., 2009; Matsui et al., 2005). As MYB5 is known to function in a partially redundant manner with MYB23 in the development and branching of trichomes, MYB23 is regarded as a tier 3 TF and may form complexes with MYB5 and TT2 in the seed coat (Gonzalez et al., 2009).

1.2.2 Transcriptional regulation of mucilage polysaccharides: tier 2 TFs

The next layer of transcriptional regulation in the seed coat and mucilage polysaccharides development are TFs in tier 2, with associated mutants usually producing less severe epidermal defects compared to mutants of tier 3 TFs. For example, mutants of tier 2 TFs such as *gl2* mentioned above (**Section 1.2.1**) have defective columella formation due to a loss of mucilage polysaccharides production,

which is less severe than the MSC defects observed in *ap2* (Western et al., 2001). The important distinction between tier 3 and tier 2 TFs is that tier 2 TFs are modulated, or proposed to be modulated, by tier 3 TFs and are able to modulate tier 1 TFs and downstream effector enzymes (Golz et al., 2018). This includes but is not limited to, TTG2, GL2, MYB61 and MYB52 (Golz et al., 2018) (Fig. 1.6). Mutants of *TTG2* are the only mutants of a tier 2 TF currently identified that have MSC developmental defects similar to the tier 3 TF *TTG1*, with a completely absent columella, but lies in tier 2 on the basis of being activated by *TTG1*-containing complexes (Johnson et al., 2002). As described previously, *TTG2* is activated jointly by *TTG1*, *EGL3* and *TT8* for the activation of effector enzyme *GATL5* while *GL2* is activated by *AP2* and *TTG1* for downstream activation of *MUM4*, *GATL5* and *PECTIN METHYLESTERASE INHIBITOR6 (PMEI6)* (Kong et al., 2013; Saez-Aguayo et al., 2013; Western et al., 2004) (Fig. 1.6). In addition to activating effector enzymes, both *TTG2* and *GL2* are also thought to regulate tier 1 TFs *KNOTTED ARABIDOPSIS THALIANA7 (KNAT7)* and *MYB75*, that form a complex to regulate the genes encoding effector enzymes such as *IRREGULAR XYLEM9 (IRX9)* (Fig. 1.6) (Bhargava et al., 2013; Golz et al., 2018; He et al., 2018; North et al., 2014; Romano et al., 2012). While *MYB61* does not appear to be regulated by any known tier 3 TF, its less severe MSC defects (reduced columella size) compared to mutants of tier 3 TFs such *ap2* and *tgt1* indicate that *MYB61* is likely to be downstream of TFs directly involved in MSC patterning (Western et al., 2004). Therefore, given the ability of *MYB61* to regulate tier 1 TF *KNAT7*, like *TTG2* and *GL2*, it is regarded as a tier 2 TF (Fig. 1.6) (Li et al., 2012; Penfield et al., 2001; Romano et al., 2012). Lastly, the recent identification of *MYB52* as a direct positive regulator of effector enzymes *PMEI6*, *PMEI14*, *SUBTILISIN-LIKE SERINE PROTEASE1.7 (SBT1.7)*, and its ability to activate tier 1 TF *SEEDSTICK (STK)*, which is also a direct activator of *PMEI6*, is consistent with *MYB52* being a tier 2 TF (Fig. 1.6) (Ezquer et al., 2016; Shi et al., 2018).

1.2.3 Transcriptional regulation of mucilage polysaccharides: tier 1 TFs

At the bottom of the hierarchy of mucilage polysaccharides transcriptional regulation lie tier 1 TFs, which are characterised by their ability to directly activate effector enzymes associated with the production, modification and secretion of mucilage polysaccharides (Fig. 1.6). These TFs do not have roles in epidermal differentiation and are typically modulated by upstream tier 2/3 TFs in order to target downstream enzyme effectors. Proposed TFs in this tier include TFs STK, SHATTERPROOF1 (SHP1), SHP2, KNAT7 and DE1 BINDING FACTOR1 (DF1). Mutants of these TFs have normal or near-normal MSC development but fail to extrude wildtype mucilage and have been shown to either directly regulate downstream effector enzymes or are associated with the mis-regulation of downstream effectors. For example, STK, and KNAT7 are positive regulators of *PMEI6* and *IRX9*, respectively, while *shp1 shp2* double and *df1* mutants are associated with normal MSCs development but defective mucilage extrusion, indicating disruptions to currently unidentified downstream effectors (Ezquer et al., 2016; He et al., 2018; Mizzotti et al., 2012; Voiniciuc et al., 2015c). STK, which can be activated by tier 2 TF MYB52, is able to bind the *PMEI6* promoter directly, thereby mediating its activation (Shi et al., 2018) (Fig. 1.6). However, heterologous expression of *PMEI6* is only able to partially complement mucilage defects in *stk* mutants, suggesting that STK may regulate other effector enzymes such as *MUM2* and *SBT1.7*, which are mis-regulated in *stk* mutants (Ezquer et al., 2016). Recently, KNAT7 has been found to activate the xylan biosynthetic genes *IRX9*, *IRX10*, *IRX14L* and *FRAGILE FIBER8 (FRA8)* using an electrophoretic mobility shift assay (EMSA; *IRX9*), and a luciferase based assay (*IRX9*, *IRX10*, *IRX14L* and *FRA8*) (He et al., 2018). In *shp1 shp2*, less adherent mucilage with uneven pectin distribution is observed, while *df1* mutants are reported to produce around half the normal amount of mucilage (Ehlers et al., 2016; Voiniciuc et al., 2015c). While TFs SHP1, SHP2 and DF1 have not been directly linked to downstream effector enzymes, mucilage defects in *shp1 shp2* are likely due to disruptions to the maturation of mucilage polysaccharides. Meanwhile, the partial loss of mucilage polysaccharides production in *df1* mutants can be explained by

disruptions in unidentified effectors involved in mucilage polysaccharides biosynthesis (Voiniciuc et al., 2015c).

Although not a TF, MUM1/LUH is a tier 1 transcriptional regulator whose mutants possess attributes similar to tier 1 TFs mutants. This is probably as LUH requires TFs to bind to target gene *MUM2*, that encodes a downstream effector enzyme (β -galactosidase) necessary for the removal of terminal galactose (*t*-Gal) residues from the RG1 backbone following secretion into the apoplast (Dean et al., 2007; Huang et al., 2011; Walker et al., 2011). MUM1/LUH is also implicated in the positive regulation of *SBT1.7* and β -XYLOSIDASE1 (*BXL1*) (Huang et al., 2011; Saez-Aguayo et al., 2013).

1.2.4 Enzyme effectors of mucilage polysaccharides biosynthesis/modification/secretion

Downstream targets of transcriptional regulators of mucilage polysaccharides biosynthesis, modification and secretion are referred to as enzyme effectors and are summarised in Table 1.1. As mucilage is mostly composed of RG1, many identified enzyme effectors are involved in pectin biosynthesis, secretion and modification. For example, MUM4 is a UDP-L-Rha synthase that converts UDP-D-Glc to UDP-L-Rha, which is used in the synthesis of the RG1 backbone in mucilage and is indirectly activated by tier 3 TFs, AP2 and TTG1, through presumably unidentified lower tier TFs (Oka et al., 2007; Western et al., 2004). Other genes involved in mucilage polysaccharides biosynthesis include: 1) *GALACTURONOSYLTRANSFERASE11* (*GAUT11*), a HG α -GalA transferase that is a member of the glycosyltransferase (GT) 8 family and catalyses the elongation of HG present in the inner adherent layer of mucilage (Caffall & Mohnen, 2009); 2) *MUCI70*, which together with *GAUT11* is implicated in the production of RG1 and HG, as *muci70* mutants have a 60% reduction in pectin found in mucilage, while double *muci70 gaut11* mutants produce almost no RG1 (Voiniciuc et al., 2018); 3) *GATL5*, another member of the GT8

family, that has been proposed to regulate the final size of pectin polymers in mucilage (Kong et al., 2013) and 4) RG1:rhamnosyltransferases1-4 (RRT1-4), members of the GT106 family, which transfers the rhamnose residue from UDP- β -l-rhamnose to RG1 oligosaccharides (Takenaka et al., 2018). Transcriptional regulation of *GAUT11* and *MUCI70* is currently unknown, while *GATL5* is activated by the TTG1, EGL3, TT8, MYB5 and TT2 complex through the tier 2 TF TTG2 (Kong et al., 2013). Besides genes encoding pectin biosynthetic enzymes, many effector proteins have been identified as involved in either RG1 or HG modification, such as *MUM2*, which is involved in pectin modification (see **Section 1.2.3**) and *PMEI6*, a pectin methylesterase inhibitor that regulates PECTIN METHYLESTERASES (PMEs) to control the degree of methyl-esterification of HG polymers in the inner adherent layer (Saez-Aguayo et al., 2013). Similar to *PMEI6*, *SBT1.7* is a protease proposed to regulate different PMEs from *PMEI6* and therefore produces an additive mucilage extrusion phenotype with patchy mucilage release in double *pmei6 sbt1.7* mutants (Rautengarten et al., 2008; Saez-Aguayo et al., 2013). Regulators of these enzymes include *LUH*, which positively regulates *MUM2*, *PMEI6* and *SBT1.7* and *STK*, which directly activates *PMEI6* (Ezquer et al., 2016; Ranocha et al., 2014). Other enzymes involved in mucilage polysaccharides modification include *PMEI14*, *FLYING SAUCER1* (*FLY1*) and *BXL1* and are summarised in Table 1.1, with details of their regulation and role in mucilage polysaccharides production described.

Although *Arabidopsis* mucilage is primarily composed of pectin, it has been useful in identifying genes involved in the synthesis and modification of cellulose and hemicelluloses (Griffiths & North, 2017) (Table 1.1). In terms of cellulose, *CESA5* is required for cellulose synthesis for the cellulosic rays in mucilage polysaccharides, with *cesa5* mutants displaying a 30% loss of crystalline cellulose in mucilage, resulting in decreased mucilage adherence (Harpaz-Saad et al., 2011; Mendu et al., 2011; Sullivan et al., 2011). Similarly, *CESA3* has been implicated in cellulose biosynthesis in mucilage polysaccharides, with *cesa3* mutants displaying altered distribution of cellulose as detected by Pontamine Fast Scarlet (S4B) staining (Griffiths et al., 2015). Mutant *cobra-like2* (*cobl2*) seeds have also been shown to

have similar mucilage defects to *cesa5* and produce less cellulose in the inner adherent layer, indicating that *COBL2* is likely involved in cellulose biosynthesis (Ben-Tov et al., 2015). Collectively, *CESA5*, *CESA3* and *COBL2* are essential for the proper formation of the cellulosic rays seen in *Arabidopsis* mucilage, which anchor mucilage to the seed. To date, no transcriptional regulator for these enzymes has been identified in mucilage.

In terms of hemicelluloses, several enzymes have been identified including *MUCI10*, *MUCI21/MUM5*, CELLULOSE SYNTHASE-LIKE A2 (*CSLA2*), *IRX9*, *IRX10* and *IRX14* that synthesis/modify either glucomannans and xylans (Voiniciuc et al., 2015b; Yu et al., 2014). *CSLA2* encodes a glucomannan synthase and is responsible for the synthesis of glucomannan, with *csla2* mutants producing 30% less Man and Glc content in mucilage, resulting in less adherent mucilage (Yu et al., 2014). Galactosylation of glucomannan in mucilage is mediated by *MUCI10*, a GALACTOSYLTRANSFERASE-LIKE (GTL) enzyme, necessary for the proper distribution of cellulose and pectin in the inner adherent layer of mucilage (Voiniciuc et al., 2015b). Transcriptional regulation of both *CSLA2* and *MUCI10* remains unknown. Besides glucomannans, xylans are an important component of hemicellulose in mucilage and have been proposed to be synthesised by *MUCI21/MUM5*, *IRX9*, *IRX10* and *IRX14*, enzymes involved in xylan assembly and elongation (Voiniciuc et al., 2015a). For example, *muci21* and *irx14* mutants have a 60% and 90% reduction in Xyl content in mucilage respectively, supporting their role in xylan production (Voiniciuc et al., 2015a). While *irx9* and *irx10* mutants do not display mucilage extrusion defects, both are expressed at the linear cotyledon stage of seed development that coincides with the onset of mucilage polysaccharides production (Voiniciuc et al., 2015a). As *IRX9* functions without the conserved aspartic acid residues required for GT43 enzyme activity, it is likely this protein plays a structural role within the xylan synthase enzyme complex, while *IRX14* and *IRX10* are catalytic subunits (Voiniciuc et al., 2015a). Although transcriptional regulation of *MUCI21* remains unknown, many xylan biosynthetic genes such as *IRX9*, *IRX10* and *IRX14* are regulated by tier 1 TF *KNAT7* (He et al., 2018).

Lastly, enzymes involved in the secretion of mucilage polysaccharides for their appropriate deposition in the apoplast have been identified and include SUBUNIT OF EXOCYST COMPLEX8 (SEC8), EXOCYST SUBUNIT EXO70 FAMILY PROTEIN A1 (EXO70A1), ECHIDNA (ECH), ROH1, YPT/RAB GTPASE INTERACTING PROTEIN 4A (YIP4A) and YIP4B (Kulich et al., 2010; McFarlane et al., 2013; North et al., 2014). These proteins/enzymes are required after mucilage polysaccharides biosynthesis, but before mucilage polysaccharides modification that occurs in the apoplast of MSCs. YIP4A, YIP4B and ECH form a complex to export pectin in post-Golgi vesicles to the plasma membrane for deposition of mucilage polysaccharides to the apoplast (McFarlane et al., 2013). Thus, mutants lacking component(s) of this complex have decreased mucilage polysaccharides export efficiency and accumulate mucilage polysaccharides in intracellular compartments, resulting in less mucilage extrusion (McFarlane et al., 2013). Similarly, exocyst subunits SEC8 and EXO70A1 form a complex with ROH1 for the intracellular secretion of mucilage polysaccharides into the apoplast, where they participate in the formation of the columella (Kulich et al., 2010). It is therefore unsurprising that *sec8* and *exo70a1* mutants extrude less mucilage, with poor to no columella formation (Kulich et al., 2010). To date, few transcriptional regulators of enzyme effectors involved in mucilage polysaccharides secretion have been identified. Overall, transcriptional regulation of downstream effector enzymes is not well understood, with 18/29 enzymes in Table 1.1 listed as having unknown regulators. This represents a considerable gap in our knowledge of mucilage polysaccharides regulation and hence, their further study is warranted.

1.3 Groucho (Gro)/Tup1 transcriptional regulators LUH and LUG

MUM1/LUH, a tier 1 regulator that regulates downstream effector MUM2/BGAL6 involved in pectin modification (see **Section 1.2.3** for details), belongs to the Gro/Tup1 transcriptional co-repressor family which is evolutionarily conserved between animals, fungi and plants, with its members required for many developmentally important events (Bao et al., 2010; Lee & Golz, 2012). The Gro/Tup1 family share similar structure and functions, underpinning their importance in transcriptional regulation. Members of this family are unable to bind to DNA directly and therefore require DNA-binding cofactors to bind to regulatory sequences of their target genes (Lee & Golz, 2012). In *Arabidopsis*, members of the Gro/Tup1 co-repressor family consist of two main sub-classes: the closely related LEUNIG (LUG) and LUH, and the closely related TOPLESS (TPL) and TOPLESS-RELATED1-4 (TPR1-4) (Liu & Karmarkar, 2008). All members of the Gro/Tup1 co-repressor family possess conserved WD-40 repeats at the C-terminus and central glutamine (Q)-rich regions, but have different conserved N-terminal domains (Fig. 1.7) (Lee & Golz, 2012). While both LUG/LUH and TPL/TPR subclasses contain a lissencephaly (LiSH) domain at their respective N-termini that facilitates protein dimerisation, substantial differences between LUG/LUH and TPL/TPR are observed (Gerlitz et al., 2005). For example, at the N-terminal, LUG/LUH possess a LUFS domain, named after the four founding members LUG, LUH, yeast Flo8, and human SSDP, that includes the LiSH domain, whereas TPL/TPR instead possess a conserved C-terminal to LisH (CTLH) domain in close proximity to the LisH domain (Fig. 1.7) (Lee & Golz, 2012; Sridhar et al., 2004). Both LUFS and CTLH domains are involved in protein-protein interactions and may contribute to the functional diversification of these co-repressors (Lee & Golz, 2012).

Within *Arabidopsis*, LUG and LUH have a moderate degree of overall sequence similarity (44%), with the N-terminal LUFS domain sharing over 80% identity and are therefore expected to have similar functions (Conner & Liu, 2000; Sitaraman et al., 2008). The LUFS domain in LUG/LUH is required for physical interaction with co-adaptor proteins SEUSS (SEU) and SEUSS-LIKE1-3 (SLK1-3) which have been

shown to interact with LUG/LUH via yeast 2-hybrid (Y2H) assays (Sitaraman et al., 2008; Sridhar et al., 2004; Stahle et al., 2009). As co-adaptor proteins SEU and SLKs have a similar sequence to metazoan Lim domain binding (Ldb) transcriptional adaptors, which are unable to bind to DNA directly, SEU and SLKs require TFs to bind to target DNA (Bao et al., 2010). Thus, LUG and LUH associate with co-adaptors SEU/SLKs and TFs to form a multimeric complex for the regulation of target genes, collectively referred to as the LUG/LUH regulatory complex (Fig. 1.8) (Bao et al., 2010). In addition to the partial redundancy of LUG and LUH described below, SEU and SLKs have been shown to function in a partially redundant manner, with SEU being the main co-adaptor, as mutations in *SLKs* only produce embryonic and ovule mutant phenotypes in a *seu* background (Bao et al., 2010). The functional redundancy between SEU and SLKs, and between LUG and LUH therefore adds a degree of intricacy in studying the LUG/LUH regulatory complex (Sitaraman et al., 2008).

While the LUG/LUH regulatory complex has been primarily identified as a transcriptional repressor that recruits either histone deacetylases (HDACs) (Gonzalez et al., 2007; Sitaraman et al., 2008; Sridhar et al., 2004) or other proteins with repressor functions (Gonzalez et al., 2007) to target genes for repression, LUH has recently shown to have the capacity to act as a transcriptional activator *in vivo* (Lee et al., 2015). Specifically, LUH binds to *PHYTOCHROME-INTERACTING FACTOR1* (*PIF1*) to activate *SCARECROW-LIKE3* (*SCL3*) in *Arabidopsis* seeds for germination (Lee et al., 2015). To date, the LUG/LUH regulatory complex has been identified in diverse roles in *Arabidopsis*, particularly LUG, which regulates the development of floral organ identity (Grigorova et al., 2011; Liu & Meyerowitz, 1995; Sridhar et al., 2006), embryo patterning (with LUH) (Sitaraman et al., 2008) and the abaxial organ identity of leaves (Cnops et al., 2004; Stahle et al., 2009). For example, LUG-SEU, together with AP2 represses *microRNA172* (*miR172*) for proper outer whorl development (Grigorova et al., 2011). LUG also directly represses the MADS-box TF *AGAMOUS* (*AG*) in the outer whorls of developing floral organs, as part of a complex that includes the MADS-box TFs AP1, SEPALLATA3 (*SEP3*) and SEU (Fig. 1.8)

(Liu & Meyerowitz, 1995; Sridhar et al., 2006). Interestingly, LUG can both complex and target MADS-box TFs (Fig. 1.8) (Sridhar et al., 2004). Specifically, AP1 and SEP3 recruit LUG-SEU to two CArG box motifs present in the second intron of *AG*, and subsequently enables LUG to prevent *AG* expression in the outer whorls of the developing sepals and petals (Sridhar et al., 2006) (Fig. 1.8). Additionally, LUG-SEU associate with MADS-box TFs AP1 and SHORT-VEGETATIVE PHASE (SVP) heterodimers, as well as AP1 and AGAMOUS-LIKE24 (AGL24) heterodimers to inhibit AP3, PISTILLATA (PI) and SEP3 expression in the outer whorls of the developing flower (Gregis et al., 2009).

In contrast to LUG, LUH appears to play minor roles in floral development (loss of *LUH* enhances floral phenotype in *lug* mutants) (Sitaraman et al., 2008), and has roles in regulating abiotic stress response (Shrestha et al., 2014), embryogenesis (with LUG) (Sitaraman et al., 2008) and mucilage polysaccharides maturation (Bui et al., 2011; Huang et al., 2011; Walker et al., 2011). The differences in roles between LUG and LUH in *Arabidopsis* are likely due to their partial functional redundancy, as over-expression of *LUH* under the cauliflower mosaic virus (CaMV) 35S promoter is unable to rescue floral phenotypes in *lug* mutants (Sitaraman et al., 2008), while expression of LUG under the *LUH* promoter in *luh* mutants is able to complement mucilage extrusion (Walker et al., 2011). Additionally, the division of LUG and LUH roles in *Arabidopsis* may be explained to some extent by their location: for example, LUH but not LUG, is expressed in the seed coat and is therefore able to regulate mucilage polysaccharides (Walker et al., 2011).

1.3.1 LUH regulation of *MUM2* in mucilage polysaccharides maturation

LUH is the most well studied tier 1 transcriptional regulator and is known to be a positive regulator of downstream enzyme effectors *MUM2*, *PMEI6*, *SBT1.7* and *BXL1* (Huang et al., 2011; Saez-Aguayo et al., 2013; Walker et al., 2011) (Table 1.1). Of these enzyme effectors, LUH is clearly an important positive regulator of *MUM2* as

not only does *luh* phenocopy *mum2* mutants, but their mucilage composition is highly similar, characterised by an increase in *t*-Gal on RG1, that can be rescued by the heterologous expression of *MUM2* in *luh* mutants (Walker et al., 2011). Although how LUH activates *MUM2* remains unknown, there are several possible ways positive regulation may be achieved: either via A) direct activation of *MUM2*, B) indirect activation of an intermediate *MUM2* activator or C) de-repression of a *MUM2* repressor (Fig. 1.9). In the model proposing direct activation, the LUH regulatory complex (LUH, SEU/SLKs and TFs) binds to and activates *MUM2* directly, perhaps through recruitment of histone acetyltransferases (HATs) (Fig. 1.9A). In the indirect activation model, the LUH regulatory complex activates *MUM2* indirectly by activating a *MUM2* activator which may again be achieved through recruitment of HATs (Fig. 1.9B). Lastly, the LUH regulatory complex may bind to and inhibit a *MUM2* repressor through recruitment of HDACs, thereby releasing *MUM2* repression and indirectly activating *MUM2* (Fig. 1.9C). In all models, *MUM2* is activated and removes *t*-Gal from RG1 side chains, resulting in mucilage with appropriate hydration properties that extrudes properly.

Based on current literature, LUH, like LUG, has been predominantly shown to act as a transcriptional repressor *in vivo*, and is involved in the repression of stress response genes *RESPONSIVE TO DESICCATION20* (*RD20*), *MYB2* and *NAC DOMAIN CONTAINING PROTEIN19* (*NAC019*), evidenced by the increased histone acetylation in the exons of these genes in *luh* mutants (Shrestha et al., 2014). In contrast, only one instance of transcriptional activation by LUH has been reported *in vivo* by chromatin immunoprecipitation assay (ChIP), which demonstrates that LUH together with PIF1 binds to *SCL3* for activation, but also binds to and represses *At5g51550*, *At5g53290*, and *EXPANSIN10* (*EXP10*) (Lee et al., 2015). Additionally, RNA-seq data supports LUH as both a transcriptional repressor and activator, associating it with the activation of 11 and the repression of 15 PIF1-regulated genes (Lee et al., 2015). Therefore, the direct and indirect activation models of *MUM2* by LUH cannot be excluded from consideration (Fig. 1.9A & B). However, a key consideration is the ability of LUG to complement mucilage extrusion defects in *luh*

mutants (Walker et al., 2011), further suggesting that LUH, in the context of the seed coat, is most likely acting as a transcriptional repressor in the seed coat and activates *MUM2* through de-repression (Fig. 1.9C).

Having established de-repression as the favoured model, one way to identify a potential *MUM2* repressor in the developing seed coat is to look for TFs that are elevated in the absence of *LUH*. This can be achieved by examining the transcriptional activity of genes in a *luh* background by RNA-seq and mining for transcription factors elevated in the seed coat at the linear cotyledon stage (7 DPA), which coincides with the onset of *MUM2* expression (Jayawardana, 2016). A set of RNA-seq experiments by Dr. Nadeeka Jayawardana (Jayawardana, 2016), a previous member of the Golz lab, revealed that numerous genes belonging to various TF families, most notably MADS-box TFs and MYBs, are elevated in *luh*, but not *mum2* mutants at the linear cotyledon stage and were therefore proposed to be candidate *MUM2* repressors (Table 1.2) (Jayawardana, 2016). As LUG seems to be functionally interchangeable with LUH within the developing seed coat (Walker et al., 2011), that similar to LUG inhibition of *AG* to allow for proper flower development, it is possible that MADS-box TFs are also repressed in the seed coat by LUH to allow for appropriate *MUM2* expression and subsequent mucilage extrusion (Fig. 1.10).

1.4 The MADS-box TF family

All members of the MADS-box TF family contain a conserved motif, the MADS-box, approximately 56-60 amino acids in length (Lu et al., 2007; Ma et al., 2005; Nam et al., 2003). The MADS-box domain facilitates the binding of MADS-box TFs to their targets via conserved binding sites called CArG-box motifs, a canonical sequence of CC(A/T)₆GG, although several variations to CArG-box motifs have been described in *Arabidopsis* (Tang & Perry, 2003; Tilly et al., 1998). The name MADS-box is derived from the four original members identified in yeast (MINICHROMOSOME MAINTENANCE1; MCM1), plants (AG & DEFICIENS; DEF) and mammals (SERUM RESPONSE FACTOR; SRF) (Shore & Sharrocks, 1995), which are known to have diverse roles in development. For example, MCM1 in *Saccharomyces cerevisiae* controls cell type-specific expression of pheromone and receptor genes (Lydall et al., 1991) and mammalian SRF is essential for cell proliferation and differentiation, such as the regulation of muscle-specific genes for muscle development (Pollock & Treisman, 1991). In flowers, MADS-box TFs have developmental roles, such as AG in specifying floral organ identity of carpels and stamens in *Arabidopsis* (Theissen, 2001) and DEF in specifying floral identity in *Antirrhinum majus* (Sommer et al., 1990). Two classes of MADS-box TFs have been identified in fungi, animals and plants based on differences in their amino acid consensus sequence within the MADS-box domain. Type I are denoted as SRF-like, while Type II are denoted as MYOCYTE ENHANCER FACTOR2 (MEF2)-like (Gramzow & Theissen, 2010; Ng & Yanofsky, 2001; Shore & Sharrocks, 1995). Type I MADS-box TFs contain an additional conserved region next to the SRF-like MADS domain; an unknown conserved region in plants, and a conserved SAM domain (named after SRF, ARG80, MCM1) known to facilitate dimerisation (Fig. 1.11) (Shore & Sharrocks, 1995; West et al., 1997). Type II MADS-box TFs are also known as MIKC-type proteins in plants, due to the long stretch (~180-200 amino acids) of domains including a MEF2-like MADS domain (M), intervening domain (I), keratin-like coil-coil domain (K) and a highly variable C-terminal (C) (Fig. 1.11) (Kaufmann et al., 2009; Theissen & Saedler, 2001). There are two clades of MIKC-type proteins; classical MIKC (MIKC^C) and the more recently discovered MIKC* that frequently

possess a longer I domain (Fig. 1.11) (Henschel et al., 2002; Zobell et al., 2010). In both MIKC-type proteins, the I domain determines the specificity of DNA binding dimerisation (Gramzow & Theissen, 2010), while the K domain facilitates dimerisation/tetramerisation between MADS-box proteins to form quartet structures needed for function in *Arabidopsis* (Yang et al., 2003; Yang & Jack, 2004). In certain MIKC-type proteins, the C-terminal is necessary for transactivation and higher-order complex formation (Puranik et al., 2014; van Dijk et al., 2010). In animals/fungi, type II MADS-box TFs consist of a MEF2-like MADS domain and an adjacent conserved MEF2 domain that facilitates dimerisation (Fig. 1.11) (Shore & Sharrocks, 1995; West et al., 1997). According to phylogenetic studies, the majority of MADS-box TFs identified in plants, animals and fungi belong to the type II category (Alvarez-Buylla et al., 2000).

Although MADS-box TFs are found in a large range of eukaryotes including yeast, plants, insects, mammals and amphibians, genomes of both fungi and animals typically encode one to five MADS-box TFs, while angiosperm genomes encode over 100 MADS-box TFs (Becker & Theissen, 2003; Gramzow et al., 2010; Gramzow & Theissen, 2010; Immink et al., 2009). This large number of MADS-box TFs in flowering plants indicates that the MADS-box TF family has expanded significantly and likely have important functions in development.

1.4.1 MADS-box TF family members in *Arabidopsis*

There are over 100 MADS-box TFs in *Arabidopsis*, which have various roles throughout plant developmental stages (Immink et al., 2009; Parenicova et al., 2003). These roles include specifying floral induction, floral organ identity, ovule development, fruit formation, silique shattering and seed production (Immink et al., 2009) (Fig. 1.12). In *Arabidopsis*, MIKC^C type proteins have major roles in floral induction, floral organ identity, fruit formation and seed development (Saedler et al., 2001). Conversely, MIKC* type MADS-box TFs have a smaller role in *Arabidopsis*

and are associated with male gametophyte development (Verelst et al., 2007). Both MIKC type MADS-box TFs form quartet structures in *Arabidopsis*, by binding to at least two CArG-box motifs within 250 bp to fold target DNA into an appropriate conformation accessible for regulation (Immink et al., 2009; Muino et al., 2014; Theissen et al., 2016) (Fig 1.13). Type I MADS-box TFs such as AGL80, AGL61 and AGL62 also have roles in embryo and endosperm development (Masiero et al., 2011).

MADS-box TFs involved in flowering time include SUPPRESSOR OF OVEREXPRESSION OF CO1 (SOC1) and FLOWERING LOCUS C (FLC), with FLC as a direct repressor of SOC1 to modulate flowering time. The antagonistic control of *SOC1* expression by both FLC (repressor) and CONSTANS (CO1; activator) controls when *SOC1* is expressed within the shoot apex, where SOC1, AGL24 and SVP activate downstream genes to establish floral meristem identity such as *LEAFY* (*LFY*), *AP1* and *CAULIFLOWER* (*CAL*) (Gregis et al., 2009; Lee & Lee, 2010). Of these TFs, only CO1 and LFY are not MADS-box TFs. In addition to specifying the floral meristem, AP1, AGL21 and SVP associate with LUG-SEU to directly repress class B and C genes according the ABC model, thereby promoting correct floral organ identity (Gregis et al., 2009).

Floral development is most famously described by the ABC model for floral organ specification and mainly involves activity of MADS-box TFs (Coen & Meyerowitz, 1991). Recently, the ABC model has been updated to the ABCDE model where there are five classes of genes; Class A (active in whorl 1 & 2), class B (active in whorl 2 & 3), class C (active in whorl 3 & 4), class D (active in developing ovules within whorl 4) and class E (active in whorl 1-4) (Theissen et al., 2016) (Fig. 1.13). According to this model, tetramers (also called quartets) of MADS-box TFs regulate floral organ identity, with the combination of class A and E TFs needed for sepal identity, class A, B and E TFs for petal identity, class B, C and E TFs for stamen identity and class C and E TFs for carpel identity (Fig. 1.13). Class A and C TFs mutually inhibit each other to establish a gradient of expression of class A and C TFs from whorl 1-4, allowing for appropriate development of the outer and inner whorls (Ng & Yanofsky,

2001; Sablowski, 2007). Class A TFs include AP2 and AP1, which together with SEP3 (class E) forms a regulatory complex with LUG-SEU to inhibit AG (class C) expression in the outer whorls (Fig. 1.8) (Sridhar et al., 2006). Class B TFs consist of AP3 and PI, while only AG has been identified as a class C TF (Ng & Yanofsky, 2001). Of these TFs, only AP2 described in the ABCDE model does not belong to the MADS-box TF family.

MADS-box TFs are also involved in fruit formation and silique dehiscence (Bowman et al., 1999; Ferrandiz et al., 1999). Within the ovule primordia, class D TFs SHP1, SHP2 and STK are predicted to complex with AG (class C) and SEP3 (class E) to specify ovule formation, as Y3H supports the formation of a complex containing SHP1, SHP2, STK and SEP3, while *stk shp1 shp2* triple mutants produce ovule defects (Fig. 1.13) (Favaro et al., 2003; Pinyopich et al., 2003; Theissen et al., 2016). STK, SHP1 and SHP2 are also crucial for seed dispersal in *Arabidopsis*, as STK is necessary for normal funiculus development and seed abscission, while SHP1 and SHP2 redundantly control the development of the dehiscence zone in the valve margins in siliques (Liljegren et al., 2000). Specifically, STK controls the formation of a lignification ring in the seed abscission zone to allow for seed separation from the funiculus by directly regulating bHLH TF *HECATE3* (*HEC3*) (Balanza et al., 2016). Similarly, SHP1 and SHP2 are needed for the expression of bHLH TFs *INDEHISCENT* (*IND*), *ALCATRAZ* (*ALC*) and *SPATULA* (*SPT*), all downstream TFs mediating the formation of both the lignified and the separation layer of the valve margin for seed dehiscence (Heisler et al., 2001; Liljegren et al., 2000; Rajani & Sundaresan, 2001).

Lastly, MADS-box TFs have roles in seed production and development, ranging from the production of anthocyanins, formation of the endothelium, to mucilage polysaccharides maturation (Ehlers et al., 2016; Nesi et al., 2002). *ARABIDOPSIS* BSISTER (*ABS*)/TT16 not only binds to and activates *BANYLUS* for proanthocyanin production in the inner endothelium layer, but together with STK is necessary for the proper formation of the endothelial layer, evidenced by the absence

of the endothelium in double *abs stk* mutants (Mizzotti et al., 2014; Nesi et al., 2002). In terms of mucilage extrusion, *stk* mutants, *shp1 shp2* double mutants and *abs/tt16* mutants reportedly have mucilage extrusion defects (Ehlers et al., 2016). While STK has been shown to directly activate *PMEI6* (Ezquer et al., 2016), SHP1, SHP2 and ABS/TT16 have not been linked to any downstream enzyme effector.

Overall, MADS-box TFs are involved in the regulation of nearly every aspect in the *Arabidopsis* life cycle and are clearly important transcriptional regulators. Given that MADS-box TFs, dependent on context, are known to be either targets of LUG or otherwise associate with the LUG-SEU complex, this forms the basis for investigating MADS-box TFs as candidate *MUM2* regulators in the de-repression model (Fig. 1.9C).

1.4.2 This study

According to the proposed de-repression model, loss of the LUH regulatory complex results in increased expression of *MUM2* repressor(s), which in turn limit *MUM2* expression, leading to mucilage extrusion defects (Fig. 1.9C). An RNA-seq experiment performed on 7 DPA seeds from *luh* mutants identified 10 MADS-box TF genes whose expression was elevated in the seed coat in comparison to wildtype and might therefore be candidate *MUM2* repressors (Table 1.2) (Jayawardana, 2016). Additionally, deletions in the 5' 200 bp regulatory sequence of the large first intron of *MUM2* (denoted as region A) that contains CArG-box motifs provide indirect evidence that MADS-box TFs may function as *MUM2* repressors, as lines transformed with a *MUM2_{pro}+I1*-reporter construct lacking region A retained reporter activity in the MSC of *luh* mutants (Jayawardana, 2016). Interestingly, CArG-box motif deletions in regulatory sequences of *MUM2* have indicated that MADS-box TFs may also function as *MUM2* activators, as CArG-box motif deletions fused to *MUM2* promoter-reporters have decreased seed coat fluorescence compared to their full length promoter-reporter counterparts in Col-0 (Jayawardana, 2016). Furthermore,

recent identification of mucilage extrusion defects in *stk* and *shp1 shp2* mutants support the role of MADS-box TFs in regulating mucilage processes (Ehlers et al., 2016; Ezquer et al., 2016).

Therefore, the overall objective of this study is to demonstrate that the regulation of *MUM2* involves MADS-box TFs and to determine the molecular mechanism underlying this regulation. This includes verifying that MADS-box TFs are elevated in *luh* mutants and are candidate *MUM2* repressors (**Chapter 3**), characterising mucilage defects in identified MADS-box TFs to show that the regulation of *MUM2* involves MADS-box TFs (**Chapter 4**) and investigating the molecular mechanism that determines how MADS-box TFs regulate *MUM2* (**Chapter 5**). Lastly, data from the three experimental chapters are summarised and new experimental directions are proposed (**Chapter 6**).

Table 1.1. Summary of effector enzymes involved in mucilage polysaccharides production, modification and secretion.

Enzyme effectors	Function (proposed/known)	Transcriptional regulation (proposed/known)	Reference
Pectin biosynthesis			
GATL5 (AT1G02720)	Proposed: GalAT glycosyltransferase; regulates final size of pectin in mucilage (<i>gatl5</i> has mucilage with higher molecular weight)	Activated by the TTG1/EGL3/TT8/MYB5/TT2 complex via TTG2	(Kong et al., 2013)
GAUT11 (AT1G18580)	HG α -GalA transferase; catalyses HG elongation	unknown	(Caffall & Mohnen, 2009) (Voiniciuc et al., 2018)
MUCI70 (AT1G28240)	Proposed: Synthesis of RG1 and HG (<i>muci70</i> has 60% reduction of pectin in mucilage)	unknown	(Voiniciuc et al., 2018)
MUM4 (AT1G53500)	UDP-L-Rha synthase; required for the synthesis of the RG1 backbone	Activated indirectly by TTG1/AP2	(Western et al., 2004) (Oka et al., 2007)
Pectin modification			
PMEI6 (AT2G47670)	Regulates PME activity and the degree of HG methyl-esterification	Directly activated by STK directly and positively regulated by MUM1/LUH	(Saez-Aguayo et al., 2013) (Ezquer et al., 2016)
PMEI14 (AT4G25250)	Esterase inhibitor; targets PMEs to regulate HG methyl esterification.	Directly activated by MYB52	(Shi et al., 2018)
FLY1 (AT4G28370)	RING E3 ubiquitin ligase; activates PMEs and the degree of HG methyl-esterification.	Unknown	(Voiniciuc et al., 2013)
BXL1 (AT5G49360)	α -L-arabinofuranosidase; cleaves <i>t</i> -Ara from RG1 polymers.	Positively regulated by MUM1/LUH	(Arsovski et al., 2009a) (Huang et al., 2011)
MUM2 (AT5G63800)	β galactosidase; cleaves <i>t</i> -Gal from RG1 polymers.	Positively regulated by MUM1/LUH	(Dean et al., 2007) (Walker et al., 2011)
SBT1.7 (AT5G67360)	Protease; targets PMEs to regulate HG methyl esterification.	Directly activated by MYB52 Positively regulated by MUM1/LUH	(Ranocha et al., 2014) (Shi et al., 2018)

Cellulose biosynthesis			
COBL2 (AT3G29810)	Proposed: involved in cellulose synthesis (<i>cobl2</i> has similar defects to <i>cesa5</i>).	Unknown	(Ben-Tov et al., 2015)
CESA3 (AT5G05170)	Cellulose synthase; required for cellulose synthesis in mucilage (<i>cesa3</i> has altered cellulose distribution)	Unknown	(Griffiths et al., 2015)
CESA5 (AT5G09870)	Cellulose synthase; required for cellulose synthesis in mucilage	Unknown	(Sullivan et al., 2011)
Hemicellulose biosynthesis			
IRX10 (AT1G27440)	Involved in xylan synthesis; elongation of xylan backbone	Positively regulated by KNAT7	(He et al., 2018)
MUCI10 (AT2G22900)	Galactosyltransferase-like enzyme; galactosylates glucomannan in mucilage	Unknown	(Voiniciuc et al., 2015b)
IRX9 (AT2G37090)	Involved in xylan synthesis; elongation of xylan backbone	Regulated by KNAT7	(He et al., 2018)
MUCI21/MUM5 (AT3G10320)	Involved in xylan synthesis; xylan branching	Unknown	(Voiniciuc et al., 2015a) (Ralet et al., 2016)
IRX14 (AT4G36890)	Involved in xylan synthesis; elongation of xylan backbone	Regulated by KNAT7	(He et al., 2018) (Voiniciuc et al., 2015a) (Hu et al., 2016)
CSLA2 (AT5G22740)	Glucomannan synthase; synthesises glucomannan.	Unknown	(Yu et al., 2014)
Polysaccharide secretion			
ECH (AT1G09330)	Forms a complex with YIP4A and YIB4B to export mucilage to the apoplast	Unknown	(McFarlane et al., 2013)
SEC8 (AT3G10380)	Forms a complex with EXO70A1 and ROH1 for the secretion of mucilage into the apoplast	Unknown	(Kulich et al., 2010)
EXO70A1 (AT5G03540)	Forms a complex with SEC8 and ROH1 for the secretion of mucilage into the apoplast	Unknown	(Kulich et al., 2010)
ROH1 (AT1G63930)	Forms a complex with SEC8 and EXO70A1 for the secretion of mucilage into the apoplast	Unknown	(Kulich et al., 2010)
YIP4A (AT2G18840)	Forms a complex with ECH and YIB4B to export	Unknown	(McFarlane et al., 2013)

YIP4B (AT4G30260)	mucilage to the apoplast Forms a complex with ECH and YIB4A to export mucilage to the apoplast	Unknown	(McFarlane et al., 2013)
Miscellaneous function			
SALT OVERLY SENSITIVE5 (SOS5)/FLA4 (AT3G46550)	Fasciclin-like arabinogalactan-protein (AGP); proposed to mediate mucilage adherence through arrangement of cellulosic rays	Unknown	(Griffiths et al., 2016) (Harpaz-Saad et al., 2011)
PER36 (AT3G50990)	Class III peroxidase; proposed to weaken MSC cell wall for mucilage extrusion	Activated by NAR1 and NARS2	(Kunieda et al., 2013)
PRX56 (AT5G15180)	Class III peroxidase; unknown function (<i>prx56</i> do not extrude mucilage)	Unknown	(Ranocha et al., 2014)

Table 1.2. Summary of TF families elevated in *luh* compared to Col-0 in 7 DPA seeds.

TFs responding to a loss of *MUM2* were excluded to produce a list of TFs families that are responding only to a loss of *LUH*. TF families proposed as potential *MUM2* repressors are shaded in blue (Jayawardana, 2016). Adapted from Jayawardana (2016).

TF family	Number of DEGs
MADS-box	10
MYBs	8
AP2-like	7
NAC	7
bHLH	5
C2H2	5
Homeodomain	5
WRKY	2
bZIP	2
DNA-binding with one zinc finger proteins (DOF)	1
Miscellaneous	13

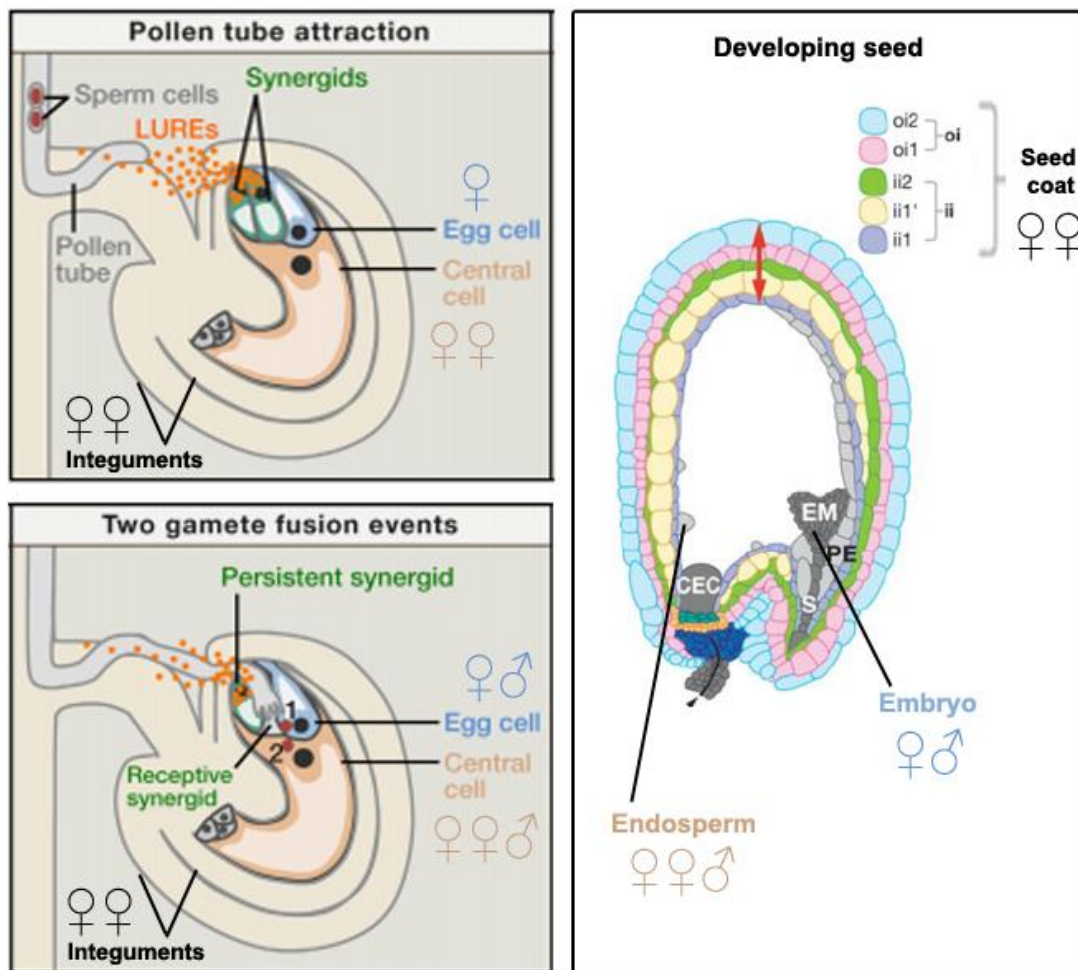


Figure 1.1. Double fertilisation in *Arabidopsis*.

The pollen tube penetrates the ovule via the micropyle and releases two sperm cells which fertilise the egg cell and central cell to eventually produce the embryo and endosperm, respectively, in the developing seed. The outer and inner integuments proceed to differentiate into the seed coat, which is of maternal origin. oi: Outer integument, ii: inner integument. Adapted and modified from: Sprunck & Dresselhaus et al., 2015, Lepiniec et al., 2006.

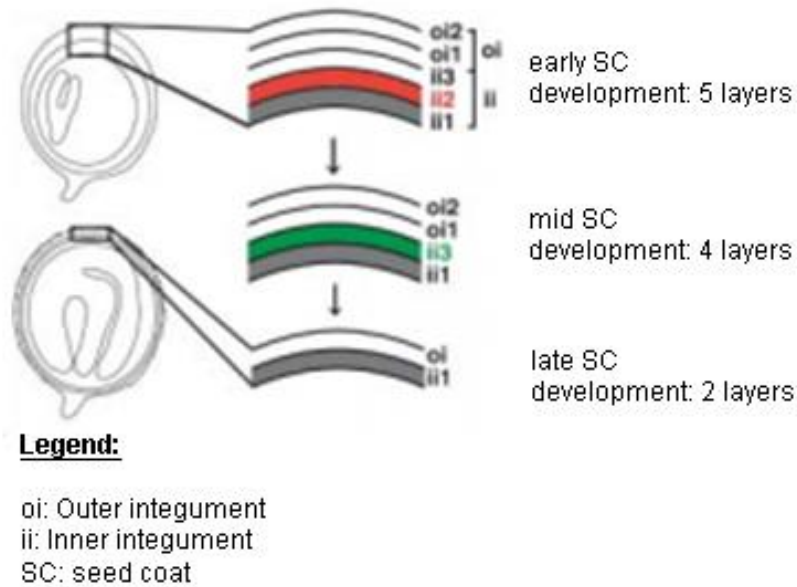


Figure 1.2. Development of the seed coat layers in *Arabidopsis*.

The inner integument (ii1-3) and outer integument (oi1-2) give rise to a total of five layers in the early stages of SC development. The inner integument is compressed into a single layer called the tegmen (ii1) where tannin is synthesised, while the outer integument is compressed into a single layer to form the testa (oi) that produces and secretes pectinaeous mucilage polysaccharides by late SC development. Adapted and modified from: Nakaune et al., 2005.

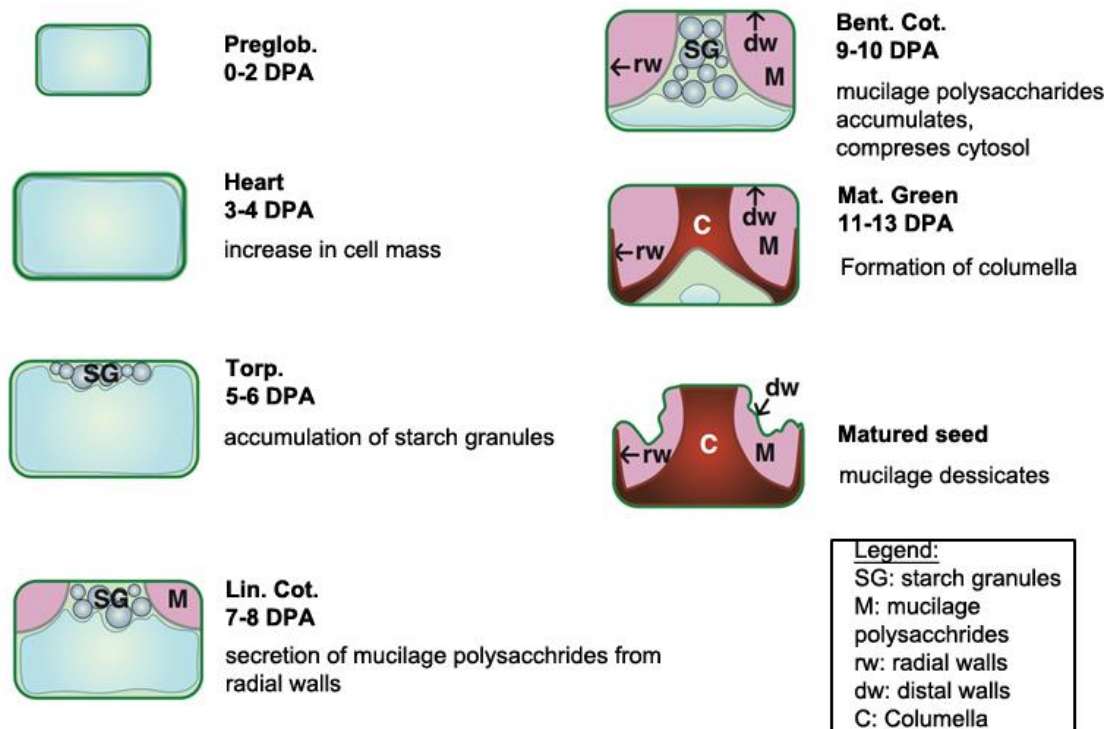


Figure 1.3. Stages of MSCs development in *Arabidopsis*.

Diagrammatic representation showing the events that occur during MSC development. In early stages (0-4 DPA), MSCs increase in cell size, and then begin accumulating starch granules during the mid-stages (5-6 DPA). Mucilage polysaccharides is secreted into the apoplast adjacent to the radial walls from 7 DPA onwards, where it accumulates and eventually displaces the cytoplasm into the centre of the cell. Secondary wall material is deposited into this cytoplasmic column, eventually replacing it with a volcano-like structure called a columella (11 DPA onwards). As the seed matures, the mucilage polysaccharides desiccate in the radial pockets, ready for extrusion on imbibition. Adapted and modified from: North et.al., 2014

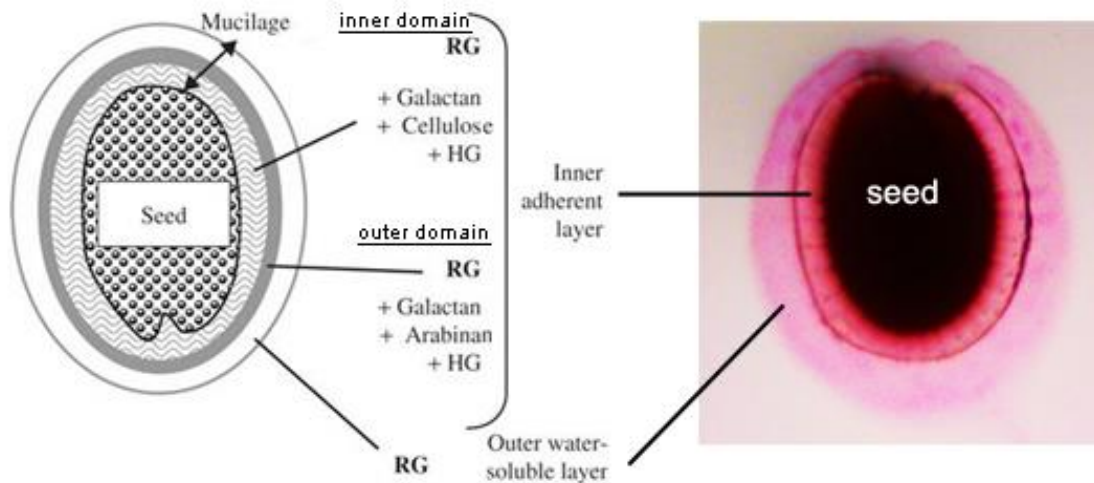


Figure 1.4. Composition of *Arabidopsis thaliana* seed coat mucilage.

Diagram showing the distribution of polysaccharides in the inner and outer mucilage layers. The outer soluble layer is mostly composed of unbranched RG1. The inner adherent layer has two main continuous domains - the inner and outer domain. The inner domain consists of RG1, cellulose, and moderately methyl-esterified HG, rendering it the least soluble portion of extruded mucilage. In the outer domain, RG1 is branched with Gal and Ara side chains while HG is strongly methyl-esterified. Adapted and modified from: Macquet et al., 2007a.

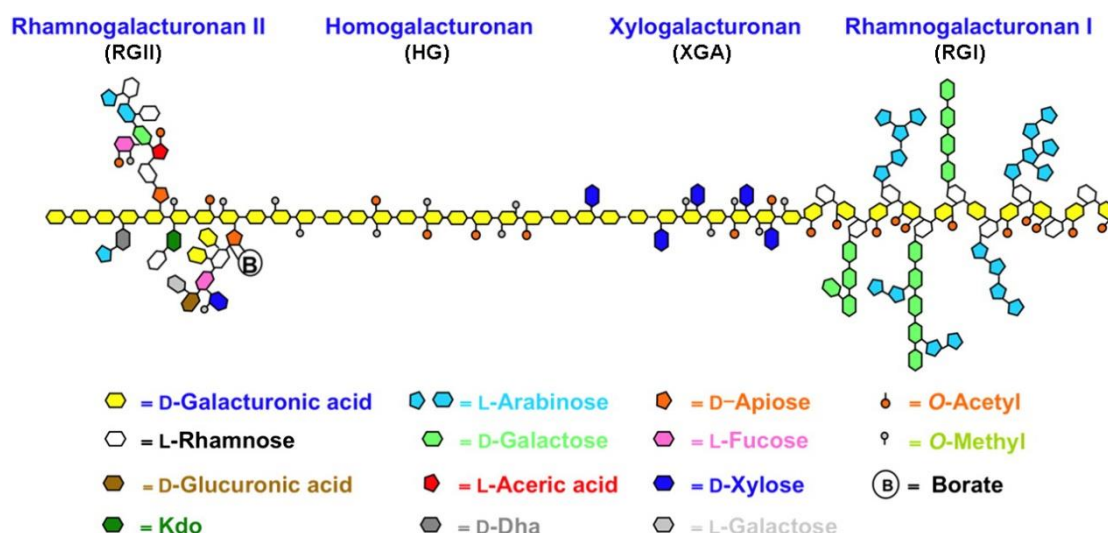


Figure 1.5. Different types of pectin found in *Arabidopsis*.

There are four different types of pectin based on structure – RGII, HG, XGA and RG1. XGA and RGII are not present in *Arabidopsis* mucilage and will not be discussed further. HG is the simplest form of pectin, with a backbone of 4-GalA that can be methyl-esterified at the C-6 position and acetylated at the O-2 and O-3 position to increase its solubility. RG1 is the only pectin that has a backbone of alternating chain of Rha and GalA and can be substituted at about half of its Rha residues by Ara and Gal side chains. Adapted and modified from Harholt et al., 2010.

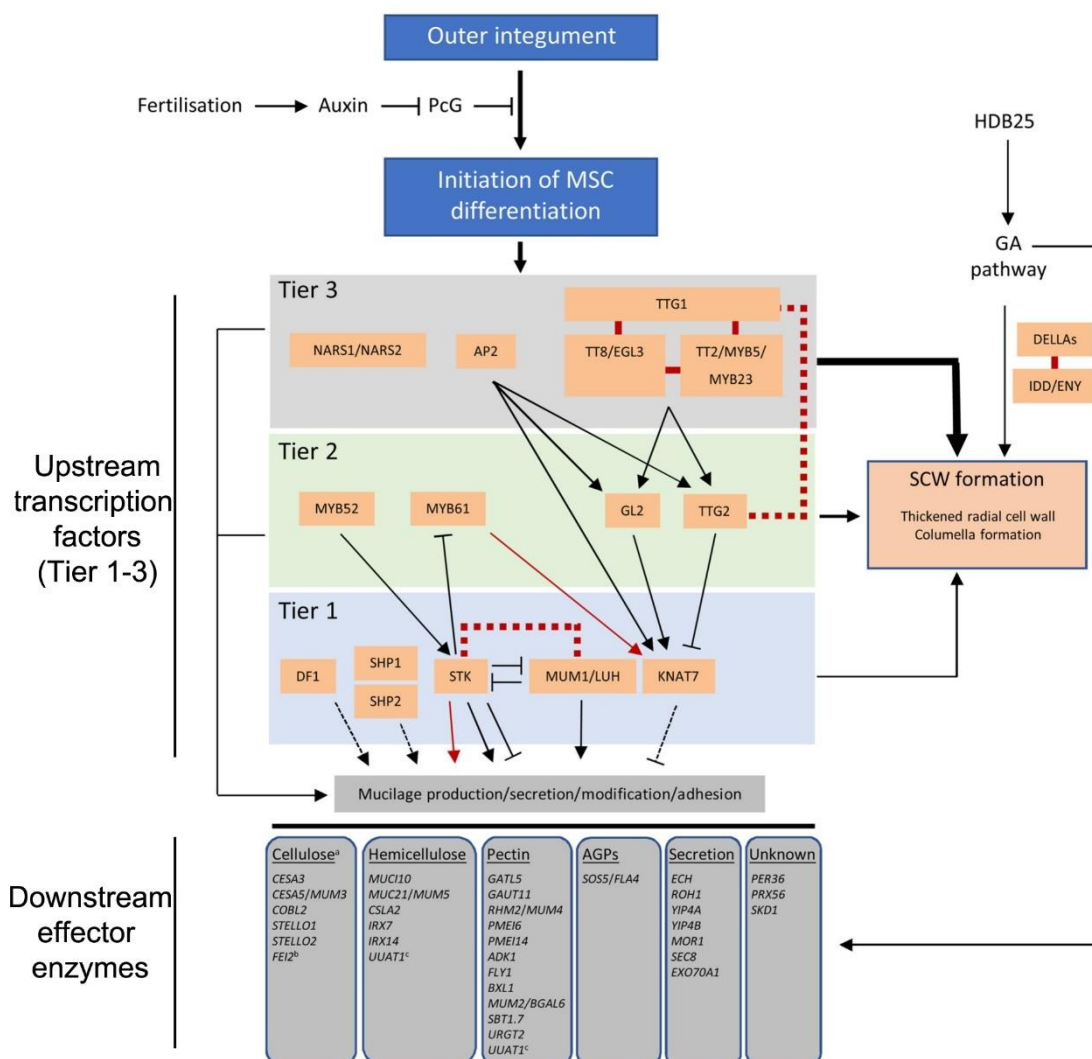


Figure 1.6. Proposed transcriptional network regulating MSC differentiation and mucilage polysaccharides production in *Arabidopsis*.

Upstream TFs are categorised into three tiers according to their impact on MSC development and mucilage polysaccharides production/secretion/modification (see text for details) while downstream effector enzymes are indicated in grey boxes. Arrows within tier 1-3 indicate direct or indirect activation, with confirmed direct interactions indicated by red arrows, probable interactions by black arrows and inferred by dashed lines. Thick red lines between boxes indicate protein-protein interactions. Blunt lines indicate direct or indirect repression. Adapted and modified from Golz et al., 2018.

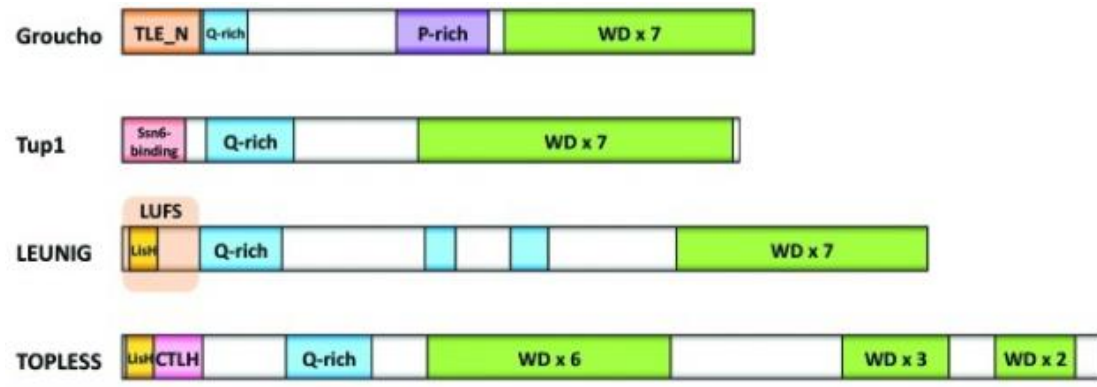


Figure 1.7. Structure of Gro/Tup1 family members in *Arabidopsis*.

All Gro/Tup1 members possess conserved WD-40 repeats (indicated in green) at the C-terminus and have glutamine (Q)-rich regions (indicated in blue) but differ substantially at the N-terminus. The two sub-classes LUG and TPL possess a LisH motif (indicated in gold) at the N-terminus unlike Gro and Tup1. In LUG, the LisH motif forms part of a larger LUFS domain while TPL has a CTHL domain closely associated with its LisH motif. Taken from Lee & Golz., 2012.

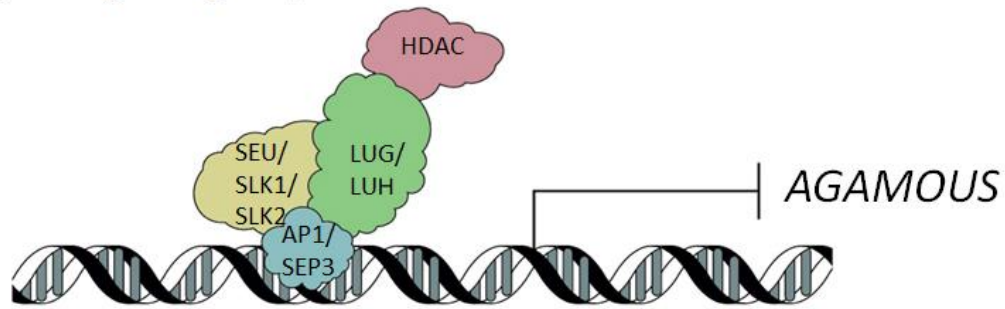
LUG/LUH regulatory complex

Figure 1.8. The LUG/LUH regulatory complex in *Arabidopsis*.

LUG/LUH (in green) are Gro/Tup1 co-repressors that recruit HDACs to repress target genes. As LUG/LUH cannot bind to DNA, they require co-adaptor proteins SEU/SLKs (in yellow) to facilitate binding to TFs (in blue) or may bind to TFs directly (not shown). In this example, TFs AP1 and SEP3 facilitate the recruitment of the LUG regulatory complex to regulatory sequences of *AG* for its inhibition in the outer whorls of the developing flower (Sridhar et al., 2006).

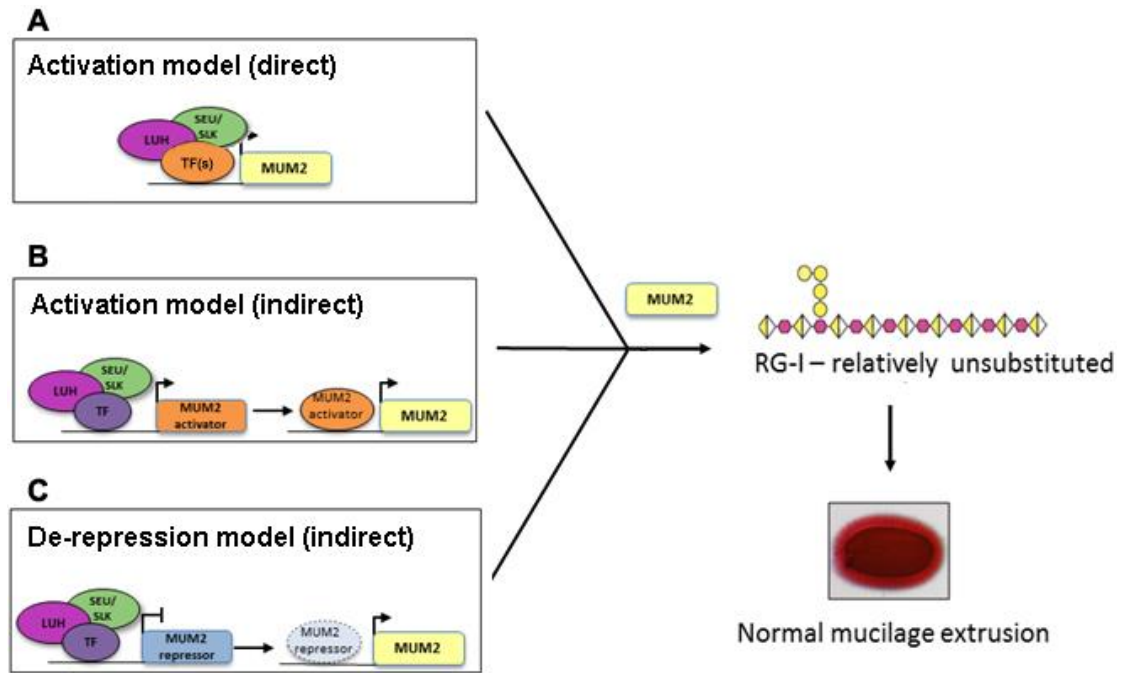


Figure 1.9. Proposed models of *MUM2* regulation by LUH.

A: Direct activation model, in which the LUH regulatory complex binds to *MUM2* regulatory sequences triggering its activation.

B: Indirect activation model, in which the LUH regulatory complex binds to and activates an intermediate *MUM2* activator, that proceeds to activate *MUM2*.

C: De-repression model, in which the LUH regulatory complex binds to and inhibits a *MUM2* repressor directly, therefore releasing repression on *MUM2*, leading to indirect activation of *MUM2*.

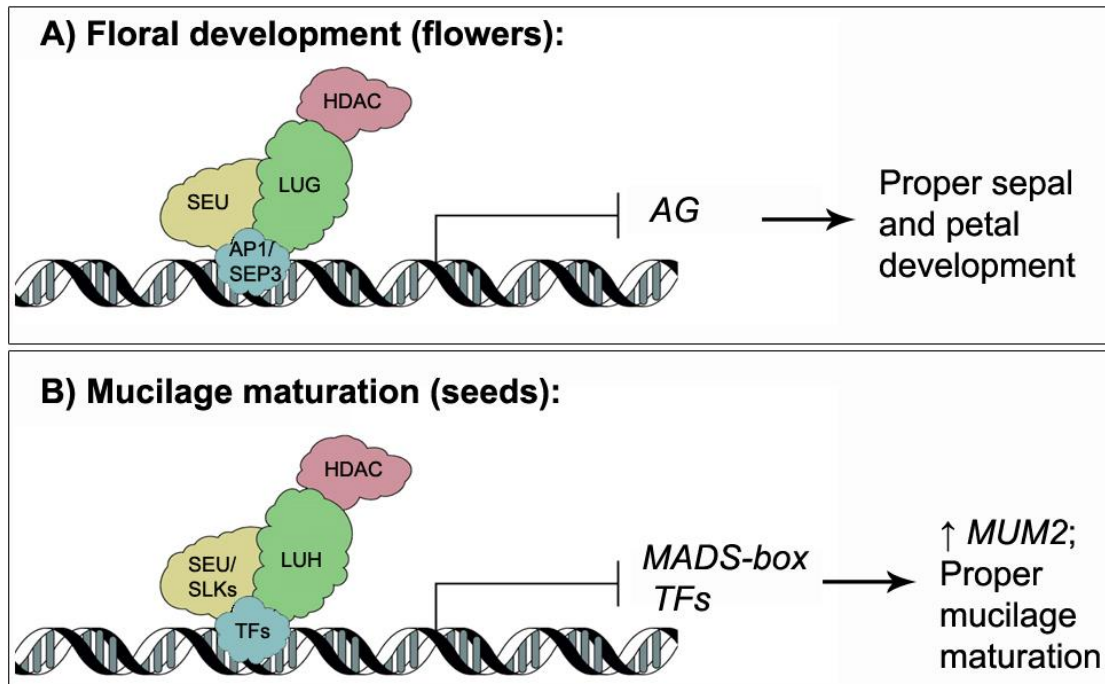


Figure 1.10. Proposed model of *MUM2* regulation in the seed in seeds.

A: LUG regulatory complex binds to and inhibits a MADS-box TF, *AG*, to allow for proper sepal and petal development in flowers (Sridhar et al., 2004).

B: Based on the function of LUG in flower development, it is proposed that LUH activates *MUM2* through de-repression in seeds. Here, MADS-box TFs are suggested to be *MUM2* repressors, that are themselves repressed by the LUH regulatory complex (Jayawardana, 2016).

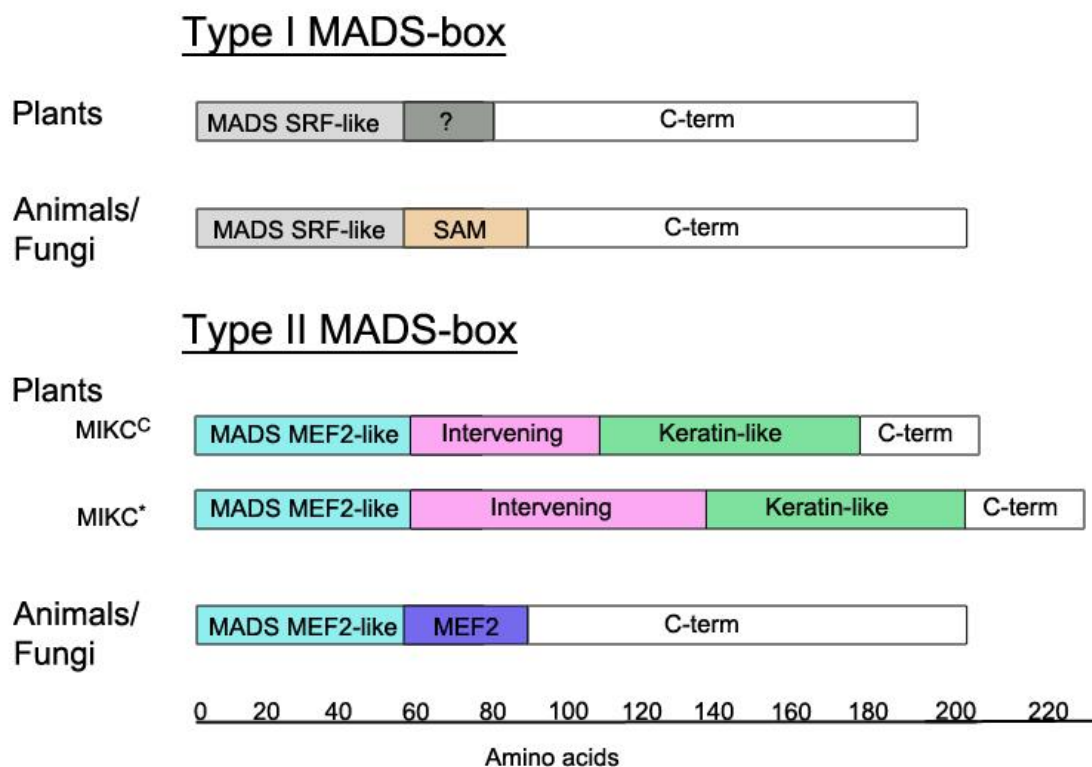


Figure 1.11. Structures of MADS Type I & II TFs.

Type I MADS-box TGs are characterised by their SRF-like MADS domain. Adjacent to their MADS domain in plants is an unknown conserved region while in animals/fungi a SAM domain is present. Type II MADS-box TFs are MEF2-like and known as MIKC-type proteins due to their conserved structure containing a MEF2-like MADS domain (M), intervening domain (I), keratin-like domain (K) and C-terminal (C). In classical MIKC (MIKC^C), intervening domain is shorter than the more recently discovered MIKC^{*} counterparts. In animals, a MEF2 domain is present next to the MADS domain, followed by the C-terminal (Shore & Sharrocks, 1995; West et al., 1997). Length in amino acids is indicative only.

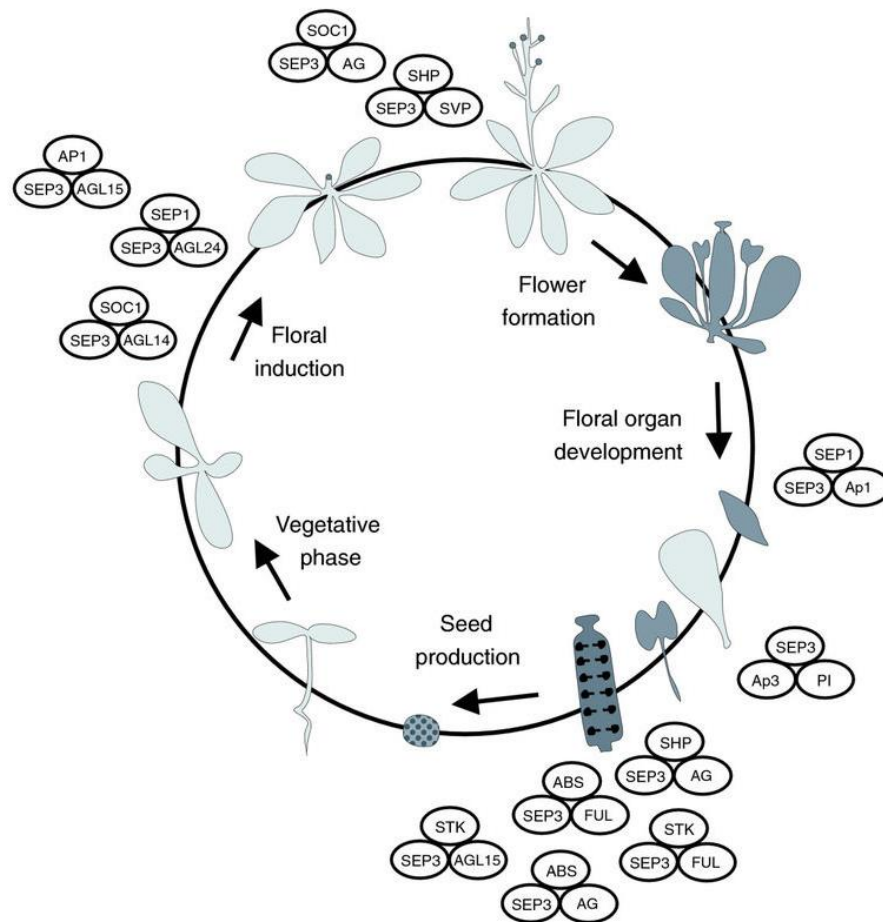


Figure 1.12. Roles of MADS-box TF complexes during plant development.

MADS-box TFs have roles in *Arabidopsis* throughout its development and includes induction of flowering, floral organ development and seed production. Taken from Immink et al. 2009.

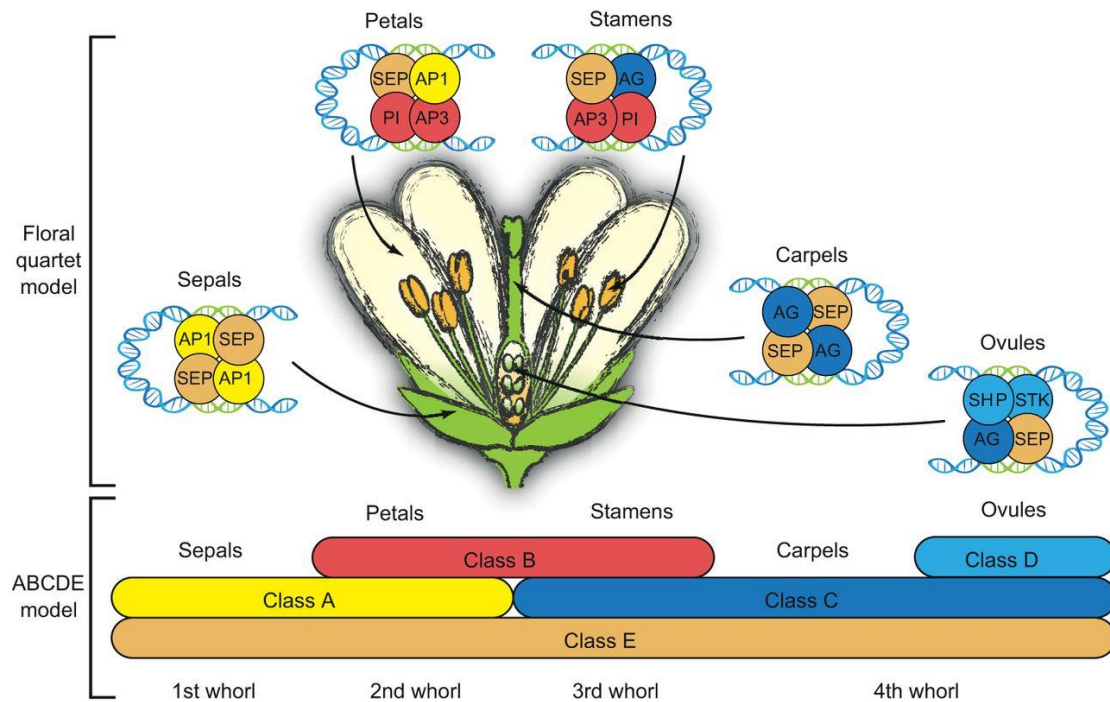


Figure 1.13. ABCDE model of floral development.

Class A TFs (in yellow) are expressed in whorls 1 and 2 where they regulate sepal and petal development. Class B TFs (in red) are expressed in whorls 2 and 3 where they specify the development of petals and stamens, while class C TFs (in blue) are expressed in whorls 3 and 4 where they regulate the identity of stamens and carpels. Two new classes of TFs include class D (in light blue), that are expressed in developing ovules that arise in whorl 4, and class E (in orange) that are expressed throughout the four whorls and form tetrameric complexes with class A, B and C MADS-box TFs.

Overlapping regions indicate that classes of these TFs participate in floral organ identity, for example, sepals are specified by class A and E TFs, petals by class A, B and E TFs, stamens by class B, C and E TFs, carpels by class C and E TFs and lastly, ovules by class D, C and E TFs (see text for details). Adapted from Theissen et al., 2016.

Chapter 2

Materials and methods

Materials and methods

2.1 Plant materials & methods

Wildtype *Arabidopsis thaliana* plants used in this study were either in a Columbia-0 (Col-0) or Landsberg *erecta* (Ler) background. T-DNA insertion lines in a Col-0 background include *luh-4* (SALK_097509), *mum2-1* (SALK_060221), *sep3-4* (CS910405), *sep1-233* (SALK_121233C), *shp1-1* (CS3842), *shp1-058* (SALK_098058), *shp1-457* (SALK_200457), *shp2-1* (CS3846), *shp2-772* (SALK_047772), *shp1-1 shp2-1* double mutants (CS3844), *ap3-387* (SALK_036387), *stk* (SALK_206790C), *df1-1* (SALK_106258), *myb61-556* (SALK_106556) and *kna7-098* (SALK_002098). Point mutation *luh-5* (*mum1-1*, CS3903) was used in a Col-2 background.

In a *Ler* background, T-DNA insertion lines include *ttg2-1* (CS277), *shp1-1* (CS3841), *shp2-1* (CS3845) and *shp1-1 shp2-1* double mutants (CS3843), while point mutations include *ttg1-1* (CS89) and *gl2-1* (CS65). All mutant lines were obtained from the Arabidopsis Biological Resource Centre (ABRC).

A standard lab strain of *Nicotiana benthamiana* was used for transient expression assays.

2.1.1 Plant growth conditions

Seeds were either grown in soil mixed with perlite (2:1 ratio) in punnets, or sterilised and sown on 0.5x Murashige and Skoog (MS) plates (0.5g/mL Murashige and Skoog basal medium, 0.7% (w/v) phytoigel (Sigma Aldrich) and adjusted to pH 5.7 with 1 M KOH).

Seeds sown directly onto soil in punnets were watered with nutrient solution (5 mM KNO₃, 2.5 mM KH₂PO₄, 2.5 mM K₂HPO₄, 2 mM MgSO₄·7H₂O, 2 mM Ca(NO₃)₂·4H₂O, 50 µM FeNaEDTA, 70 µM H₃BO₃, 14 µM MnCl₂·4H₂O, 500 nM CuSO₄·5H₂O, 1 µM ZnSO₄·7H₂O, 10 µM NaCl and 10 nM CoCl₂·6H₂O) and covered with cling-wrap to maintain humidity. Next, seeds were stratified at 4°C in the dark for two days and then transferred to a growth room under continuous light at 18-20°C. Humidity was maintained by cling-wrap and removed with the emergence of cotyledons. Germinating seedlings were watered three times a week; once with nutrient solution and twice with water.

To avoid fungal contamination on plates, seeds sown onto 0.5x MS plates containing 0.7% (w/v) phytogel were first sterilised using chlorine gas (generated by adding 3 mL HCl to 100 mL sodium hypochlorite in a fume hood) for at least four hours. Plates with sown seeds were then stratified at 4°C in the dark for two days and subsequently transferred to a growth cabinet under continuous light at 21°C for approximately one week. Germinating seedlings were then transferred to soil, watered with nutrient solution and covered with cling-wrap to maintain humidity for two days. After two days, cling-wrap was removed. Resultant plants were watered three times a week with nutrient solution once a week and water twice a week.

2.1.2 Floral dip transformation

Transformation of *Arabidopsis* plants was performed using the *Agrobacterium*-mediated floral dip procedure (Clough & Bent, 1998; Zhang et al., 2006). Binary vector constructs were first generated and transformed into *Agrobacterium tumefaciens* strains AGL1 or C58 via electroporation as described in **Section 2.2.12**. Next, a single colony was inoculated into a starter culture of 10 mL Luria Bertani (LB) medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl) containing appropriate selection antibiotics in a 50 mL conical tube, and incubated overnight at 28°C, 200 rpm. The following day, the starter culture was used to

inoculate 300 mL LB with appropriate selection antibiotics in a 1 L Erlenmeyer flask and incubated for 2 days at 28°C with shaking at 200 rpm until an optical density (OD₆₀₀) of at least 1 was reached. The culture was then centrifuged at 4,000 x g for 20 mins to collect the cells which were subsequently resuspended in 800 mL of freshly prepared *Agrobacterium* infiltration medium (5% (w/v) sucrose, 0.05% (v/v) Silwet L-77 (OSi Specialties), 10 mM MgCl₂). The resuspended culture was then poured into a foil tray before submerging the inflorescence tips of *Arabidopsis* plants in this solution for 2 mins. Dipped plants were laid horizontally in a clean tray and covered with cling-wrap overnight to provide sufficient humidity for *Agrobacterium* infection. The next day, cling-wrap was removed, and plants were placed in an upright position and watered with nutrient solution to recover. Plants were maintained under growth conditions described in **Section 2.1.1** until mature and resultant T₁ seeds collected. Lastly, transgenic seedlings were identified by sowing the T₁ seeds at a high density on soil before spraying with a solution of Basta® (100 µg/mL; Bayer).

2.1.3 Generation of crosses

Arabidopsis lines were crossed by pollinating the carpel of a designated maternal parent with pollen from a designated paternal parent. Forceps were used to remove anthers from an unopened, unfertilised flower that was to be used as the pollen recipient. Next, pollen from the paternal flower was then manually transferred onto the stigma of the emasculated flower. Successful fertilisation of the carpel was indicated by the production of anthocyanin at the base of the carpel the following day. Seeds were subsequently collected on maturation of the silique approximately three weeks-post-fertilisation. Seeds from crosses were sowed on 0.5x MS plates and the resulting seedlings transferred to soil after approximately two weeks of growth.

2.1.4 Extraction of genomic DNA from *Arabidopsis*

Genomic DNA of *Arabidopsis* plants was extracted according to the Edward's protocol (Edwards et al., 1991). This involved collecting fresh young leaf tissue in a 1.5 mL tube and macerating the tissue with a micropestle before the addition of 400 μ L extraction buffer (200 μ M Tris-HCl, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% (w/v) SDS). Samples were vortexed for 5 sec before centrifuging at 14,000 x g for 3 mins. The resultant supernatant was transferred to a fresh 1.5 mL tube and mixed with an equal volume of isopropanol before incubating at room temperature for 5 mins. Precipitated nucleic acids were subsequently collected by centrifugation at 14,000 x g for 5 mins. Pelleted nucleic acids were air-dried for 1 hour before resuspension in 25 μ L TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5). Resuspended genomic DNA was stored at -20°C until needed.

2.2 General molecular materials & methods

2.2.1 Growth media

LB medium in both liquid and solid (1.5% agar) form was used for the culture of *Escherichia coli* and *Agrobacterium tumefaciens*. To identify recombinant *E. coli* strains, appropriate antibiotics were added to the medium such as ampicillin (100 µg/mL), spectinomycin (100 µg/mL) or kanamycin (50 µg/mL). For recombinant *A. tumefaciens* strains, antibiotics such as rifampicin (50 µg/mL), spectinomycin (100 µg/mL), gentamycin (25 µg/mL), kanamycin (50 µg/mL) or tetracycline (10 µg/mL) were routinely added in various combinations according to the construct being introduced into the cells. In addition, Super Optimal broth (SOB) with catabolite repression (SOC) media (0.5% (w/v) yeast extract, 2% (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, with the addition of 20 mM glucose post-autoclaving) was used to recover both *E. coli* and *A. tumefaciens* cells immediately following transformation via either heat shock or electroporation, respectively.

Yeast strains were grown in YPDA medium (2% (w/v) tryptone, 1% (w/v) yeast extract, 2% (w/v) glucose, 0.003% (w/v) adenine hemisulfate, 1.5% (w/v) agar for solid medium). Selective synthetically defined (SD) plates (Clontech) were reconstituted according to manufacturer's instructions, with SD double dropout (DDO) plates lacking both Trp and Leu used to select for successfully co-transformed yeast strains. Finally, both quad dropout (QDO) plates lacking adenine (Ade), histidine (His), Leu and Trp amino acids and QDO plates containing 2 mM 3-aAmino-1,2,4-triazol (3-AT) were used to assay for successful protein-protein interactions.

2.2.2 Chemicals & reagents

Laboratory chemicals were obtained from Sigma-Aldrich, Promega, Thermo Fisher Scientific, Ajax, Bioline, Merck, Oxoid and Scientifix. Common enzymes such as KOD Hot Start DNA Polymerase, Q5® High-Fidelity DNA Polymerase, restriction enzymes (RE), DNA ligase and Calf Intestinal Phosphatase (CIP) were purchased from Novagen, New England Biolabs (NEB) and Promega, respectively. Antibiotics were reconstituted according to manufacturer's instructions at 1000 x working concentration.

Kits such as the Gibson Assembly® System were purchased from NEB, SuperScript III One-Step RT-PCR System from Invitrogen, Tetro cDNA Synthesis Kit and SensiMix SYBR® No-ROX Kit from Bioline, Spectrum™ Plant Total RNA Kit from Sigma-Aldrich, Wizard® Plus Minipreps DNA Purification System, Wizard® SV Gel and PCR Clean-Up System and Dual-Luciferase® Reporter Assay System from Promega.

Oligonucleotides were designed using the A plasmid Editor (ApE) tool (<http://biologylabs.utah.edu/jorgensen/wayned/apex/>), evaluated using OligoAnalyser (Integrated DNA Technologies (IDT, USA) and ordered from either Sigma-Aldrich or Macrogen.

2.2.3 Plasmid vectors

Plasmids used in creating constructs for plant transformation purposes such as generating transgenic lines in *Arabidopsis*, use in transactivation and bimolecular fluorescence complementation (BiFC) assays in *N. benthamiana* include shuttle vector pMIGRO (JG171) with ampicillin resistance and binary vectors pMLBART (JG13) with spectinomycin resistance and pGREENII (JG1189) with kanamycin resistance (Gleave, 1992; Hellens et al., 2000; Moore et al., 1998). Both binary vectors contain a BaR gene under the control of a NOS promoter which confers

Basta® herbicide resistance (Gleave, 1992; Hellens et al., 2000). Helper plasmid pSOUP (JG946), containing tetracyclin resistance, was co-transformed with pGREENII to enable replication in *Agrobacterium* (Hellens et al., 2000).

For yeast-based assays, pGBKT7, pGADT7 Matchmaker vectors (Clontech) and pTFT1 were used as bait, prey and ternary factor plasmids, respectively.

2.2.4 Bacterial strains and growth conditions

A laboratory stock of *E. coli* TOP10 cells with streptomycin resistance were maintained and used for general cloning purposes. *A. tumefaciens* strains C58 with rifampicin and gentamicin resistance and AGL1 with ampicillin resistance were used for transformation of *Arabidopsis* and *N. benthamiana*, with C58 predominantly used for the stable transformation of *Arabidopsis* whereas AGL1 was used exclusively for transactivation and BiFC assays in *N. benthamiana*. Both strains were grown in LB containing appropriate antibiotics at 28-30°C, with liquid cultures shaken at 200 rpm until appropriate OD₆₀₀ was reached, typically for 2 days.

Saccharomyces cerevisiae strain Y2HGold (Clontech) was grown in liquid YPDA at 30°C, with agitation at 250 rpm until optimal OD₆₀₀ of 0.4-0.5 was reached. Transformed Y2HGold cells were plated on SD plates supplemented with appropriate combinations of amino acids and incubated at 30°C for 3 days.

2.2.5 Polymerase chain reaction (PCR)

For cloning purposes, sequences of interest were amplified with high fidelity DNA polymerases such as KOD HiFi Hot Start DNA Polymerase (Novagen) or Q5® High-Fidelity DNA Polymerase (NEB) from genomic DNA or cDNA templates. All PCR reactions were performed using DNA Engine Peltier Thermal Cyclers (BioRad).

Reaction volumes of 50 μ L were set up according to the manufacturer's instructions and then cycled through the following conditions: initial denaturation of 94-98°C for 2 mins, followed by 35-40 cycles of denaturation at 94-98°C for 15 sec, annealing at 55-60°C for 20-30 sec and 20 sec/kb of extension at 72°C. Final extension was performed for 2-5 mins at 72°C. Resultant PCR products were separated and visualised on 1% (w/v) agarose gels containing 0.5 μ g/mL ethidium bromide, run in either Tris/Borate/EDTA (TBE) or Tris/Acetate/EDTA (TAE) buffer and subsequently cloned and verified by DNA sequencing before further use.

Laboratory stocks of non-proof reading DNA Taq polymerase were used for both genotyping and colony PCR purposes. Reaction volumes of 10 μ L were set up with 1 μ L of template material (gDNA/cDNA) and amplified with Taq polymerase under the following conditions: 2 mins of initial denaturation at 94°C, followed by 40 cycles of 20 sec denaturation at 94°C, 30 sec of annealing at 55°C and 30 sec/kb of extension at 72°C. Final extension was performed for 5 mins at 72°C. Resultant PCR products were analysed as above.

2.2.6 Plasmid DNA extraction

Plasmid DNA was extracted from 3-5 mL of recombinant *E. coli* cultures using the Wizard® Plus Minipreps DNA Purification System (Promega) according to the manufacturer's instructions.

2.2.7 Restriction digestion

For DNA cloning purposes, single and double restriction enzyme (RE) digestions of 1-2 μ g of DNA were incubated at 37°C for 2-4 hours and overnight for diagnostic purposes. A typical digestion reaction contained 1 unit (U) of each RE in a final concentration of 1x optimal RE buffer supplied by the manufacturer. To promote

complete digestion, 0.1 mg/mL bovine serum albumin (BSA) was added. When cloning, restriction digests were inactivated by heat treatment at 65°C for 20 mins and/or purified with the Wizard® SV Gel and PCR Clean-Up kit (Promega). Where necessary, digested products for cloning were separated on a 1% (w/v) TAE agarose gel. In contrast, RE products for diagnostic purposes were run directly on a 1% (w/v) TBE agarose gel following RE digestion.

2.2.8 Vector dephosphorylation

Dephosphorylation of vectors was performed immediately following RE digestion to prevent self-ligation. Briefly, 1 U of CIP (Promega) was added to ≤ 1 pmol of vector and a final concentration of 1x CutSmart® buffer in a 20 μ L total volume. Reactions were then incubated at 37°C for 30 mins. All dephosphorylation reactions were subsequently heat-inactivated at 74°C for 15 mins before storage at -20°C.

2.2.9 DNA ligation

DNA ligation reactions were performed in a 10 μ L reaction volume that included a vector to insert molar ratio of 1:3, 1 U of T4 DNA ligase (NEB) and a final concentration of 1x T4 ligase buffer. Ligations were incubated at 16°C overnight in 0.2 mL tubes using a DNA Engine Peltier Thermal Cycler and stored at -20°C until needed.

2.2.10 DNA sequencing

DNA sequencing was performed in a 20 μ L reaction volume consisting of 1x BigDye Terminator sequencing buffer V3.1 (Applied Biosystems), 500 nM sequencing primer and either 400 ng of plasmid DNA or 50 ng of PCR product. This reaction was cycled

in a DNA Engine Peltier Thermal Cycler under the following conditions: 5 mins of initial denaturation at 94°C, followed by 30 cycles of 10 sec denaturation at 96°C, 5 sec of annealing at 50°C and 4 min of extension at 72°C. The sequencing reaction then was precipitated with 80 µL of 75% (v/v) isopropanol in a 1.5 mL tube, vortexed, and incubated at room temperature for 20 mins. This mixture was then centrifuged at 16,000 x g for 20 min at room temperature before washing the DNA pellet with 200 µL 75% (v/v) isopropanol and removing residual isopropanol by heating at 90°C for 1 min. The sequencing reaction was then sent to the Melbourne Translational Genomics Platform (MTGP, Department of Pathology, University of Melbourne) or Macrogen (Korea) for electrophoresis. Sequencing results were analysed using the ApE software.

2.2.11 Construct generation

Constructs were generated by incorporating target DNA sequences into a suitable plasmid using seamless or Gateway cloning. These cloning procedures require an overlap of 15-20 bp between the vector and target DNA or the presence of compatible *att* sites, respectively. These sites were incorporated into the target DNA using PCR and genomic DNA or cDNA as a template as described in **Section 2.2.5**. PCR products were then purified using a Wizard® SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. Purified inserts were introduced into shuttle vectors such as pMIGRO (JG171), pGADT7, pGBKT7 and pTFT1 using a 5 µL Gibson® assembly reaction (NEB) or through a Gateway reaction (Thermo Fisher Scientific) according to the manufacturer's instructions. Resultant constructs were subsequently transformed into *E. coli* for propagation and confirmed by DNA sequencing before transformation into the appropriate cell line for downstream uses.

Constructs for plant transformation require the use of a binary vector such as pMLBART (JG13) to allow for *Agrobacterium*-mediated transformation. This involved transferring recombinant DNA from pMIGRO into pMLBART (JG13).

Briefly, pMIGRO constructs were digested with the appropriate RE and gel-purified to obtain the transgene cassette, which was subsequently ligated into pMLBART (JG13) cut with *Not* I and dephosphorylated. Alternatively, transgene cloning into binary vector pGREENII (JG1189) was achieved using the Gibson® assembly method.

2.2.12 Transformation of bacteria

For *E. coli* transformations, ≥ 200 ng of plasmid DNA was incubated with 50 μ L of chemically competent *E. coli* cells in a 1.5 mL tube on ice for 5 mins before applying heat shock at 42°C for 30 sec. Cells were placed back on ice for 2 min to recover before 1 mL SOC was added, and the mixture was incubated at 37°C, with shaking at 200 rpm for 1 hour. Following incubation, transformed cells were pelleted by brief centrifugation before being plated onto LB media supplemented with the appropriate antibiotics and incubated overnight at 37°C.

A 50 μ L aliquot of electrocompetent *Agrobacterium* was first incubated with ≥ 50 ng of construct in a 1.5 mL tube on ice for 15 mins. The mixture was then transferred to a pre-chilled electroporation cuvette for electroporation under standard conditions: 2.5 kV voltage, 25 μ F capacity and 200 Ω resistance. Cells were immediately resuspended in 1 mL SOC following electroporation and transferred to a 1.5 mL tube for incubation at 30°C, 200 rpm for 2 hours. Recovered cells were then plated onto LB plates supplemented with the appropriate antibiotics and incubated for 2-3 days at 30°C.

2.3 Quantitative reverse transcription PCR (qRT-PCR)

2.3.1 Tissue isolation for RNA extraction

Both whole seed and manually isolated seed coat tissue were collected for RNA extraction. This involved marking flowers at anthesis and then collecting seeds at either 3, 7 or 10 days post-anthesis (DPA). Microscopic examination of seeds showed that 3 DPA seeds contained heart stage embryos, whereas linear cotyledon stage embryos were present in 7 DPA seeds and 10 DPA seeds contained mature green stage embryos. Seeds at linear cotyledon (7 DPA) stage were harvested to obtain seed coat material for qRT-PCR experiments, whereas whole seeds at heart (3 DPA), linear cotyledon (7 DPA) and mature green (10 DPA) stages were collected for RNA-seq and qRT-PCR. For both whole seeds and isolated seed coats, three biological replicates were collected in a manner described below.

Developing siliques were placed on double-sided tape attached to thick cards in such a way that the replum was facing upwards. For whole seed collections, approximately 10 siliques were cut open using a hypodermic needle and seeds transferred to labelled 1.5 mL tubes containing 500 μ L RNAlater storage solution (Ambion). Tubes were kept on ice until tissue collection was completed and then stored at -80°C until needed. For seed coat material, about 17 siliques were cut open and their immature seeds were transferred to 50 μ L of wash solution (5% (w/v) sucrose in 0.1% (v/v) RNAlater) placed on a microscope slide. Developing embryos were separated from the seeds by applying gentle pressure via a glass coverslip. The embryo-less seeds were then transferred to 1.5 mL tubes containing 500 μ L RNAlater storage solution. All samples were kept on ice until tissue collection was completed and subsequently stored at -80°C until needed.

2.3.2 RNA extraction

RNA extractions of samples for qRT-PCR and RNA-seq analyses were performed on the same day to minimise handling variation. Prior to RNA extraction, all samples were thawed on ice. To enable separation of seed material from storage solution, 1 mL of nuclease-free water was added to each sample to reduce the viscosity of RNAlater. Seed material was then pelleted by centrifugation at 16,000 x g for 5 mins before removal of the supernatant. For effective tissue grinding, acid-washed sand was added to each seed pellet and subsequently homogenised using micropestles that were pre-chilled in liquid nitrogen. RNA extractions of homogenised seed tissue were performed using the Spectrum plant total RNA kit (Sigma-Aldrich). An initial 100 µL aliquot of lysis solution was added during the grinding process to aid with the homogenisation before the remaining 400 µL of lysis buffer was added. The manufacturer's instructions were followed for all subsequent steps, and RNA was eluted in 50 µL elution buffer. All RNA samples were subjected to DNase treatment using a TURBO DNA-free kit (Ambion) according to the manufacturer's instructions. Finally, RNA integrity was assessed by two methods described in the following section before use in downstream applications.

2.3.3 RNA quality control

The integrity of extracted total RNA was first assessed using a Thermo Scientific NanoDrop 1000 spectrophotometer with a ratio ≥ 2.0 for 260/280 nm readings and ≥ 2.0 -2.2 for 260/230 nm considered high quality. Samples that achieved this threshold were passed to the second quality control step in which 500 ng of RNA per sample was separated on a non-denaturing 1% (w/v) TBE agarose gel to look for the presence of intact ribosomal RNA subunits (25S and 18S). Only samples with both a brighter ethidium bromide stained 25S band (approximately twice as bright as the 18S band) and absence of smearing were considered non-degraded and subsequently used for downstream applications. For those samples subjected to RNA sequencing, RNA

integrity was assessed further using an Agilent 2100 Bioanalyser with a threshold of 7 RIN (RNA integrity number) deemed acceptable for subsequent analysis.

2.3.4 qRT-PCR primer design

Primers used for qRT-PCR were designed with the PrimerQuest Tool (Integrated DNA Technologies (IDT, USA), evaluated with OligoAnalyser 3.1 (IDT) and purchased from Sigma-Aldrich (Australia). Coding DNA sequences (CDS) of targets were obtained from The Arabidopsis Information Resource (TAIR; <https://www.arabidopsis.org/>).

Where possible, primers (size=18-25, T_m = 55-60°C) for qRT-PCR were designed so that the amplicon spanned an intron to allow genomic DNA contamination to be assessed. Amplicon lengths were typically ~100 bp, but ranged from 75 bp to 150 bp. A 3' GC clamp (maximum 2 nt) was also included in the primer design where possible. Potential primer sequences were next evaluated using OligoAnalyser and only those that failed to self/hetero-dimerise, form hairpins, or had little/no non-specific binding to other regions of the *Arabidopsis* genome were selected for use.

2.3.5 qRT-PCR experimental design & parameters

Contaminating genomic DNA from extracted RNA samples was removed with DNase using a TURBO DNA-free kit (Ambion) according to the manufacturer's instructions. All DNase-treated samples were quantified using a Thermo Scientific NanoDrop 1000 spectrophotometer and assessed as described in **Section 2.3.3** and stored at -80°C until needed.

First-strand cDNA was generated from 1 µg of DNase-treated RNA from isolated seed coat samples using the SuperScript III One-Step RT-PCR System (Invitrogen)

and from whole seed samples using a Tetro cDNA Synthesis Kit (Bioline) according to the manufacturer's instructions. All cDNA synthesis reactions were performed using DNA Engine Peltier Thermal Cyclers in 0.2 mL PCR tubes.

qPCR was performed using a SensiMix SYBR® No-ROX Kit (Bioline) according to the manufacturer's instructions, with 1 µl of 10 x diluted cDNA template, a final concentration of 1x SensiMix, nuclease-free water and 250 nM of primers in a 10 µl reaction volume. Housekeeping genes used for normalisation include *TUBULIN7* (AT2G29550) for seed coat samples and *ACTIN2* (AT3G18780) for whole seed samples. Three biological replicates were amplified in triplicates for each gene in a 96-well plate using the CFX96 Touch™ Real-Time PCR Detection System (Biorad) under the following cycling parameters: initial denaturation at 95°C for 10 mins, followed by 45 cycles of denaturation at 95°C for 15 sec, annealing at 60°C for 15 sec, extension at 72°C for 15 sec and a plate read step. A melt curve gradient was undertaken as a final step (65°C to 95°C, with 0.5°C per 5 sec) with single peaks in samples, excluding the no template control, indicating specific gene amplification with minimal levels of primer dimer and/or non-specific amplification products. qPCR results were exported to Excel and the relative expression of each gene normalised against housekeeping genes using the comparative CT method ($\Delta\Delta C_T$) (Livak & Schmittgen, 2001). Statistical significance was calculated using the unpaired student's *t*-test ($p>0.05$)

2.4 Microscopy

2.4.1 Light microscopy

All light microscopy was performed using the Nikon (SMZ800) dissecting microscope and images captured using the Nikon ACT-2U (ver 1.70) software. Mature seeds stained with 0.02% (w/v) Ruthenium Red were resuspended in 0.1% (w/v) agar to reduce light reflection and were placed on a microtitre plate cover for imaging under 30 x and 63 x magnification. Floral and silique samples were imaged under 20-40 x magnification.

2.4.2 Compound fluorescence microscopy

Seed coat and embryos from transgenic *Arabidopsis* lines and leaf sections from infiltrated *N. benthamiana* leaves were examined using an Olympus BX60 fluorescence microscope under 100 x and 200 x magnification, with blue light excitation at 515 nm. *Arabidopsis* seed samples were first mounted on microscope slides in 10% (v/v) glycerol, with gentle pressure applied to coverslips to separate seed coats from embryos while thin leaf sections from *N. benthamiana* were also mounted onto microscope slides in 10% (v/v) glycerol with the abaxial side facing upwards. To enable direct comparisons across all samples, images were captured under the same intensity setting. Samples that required additional exposure in order to produce detectable fluorescence were post-processed using Adobe Photoshop (ver. 12.0), with overall exposure across all sample sets meant for direct comparisons increased at the same rate.

2.4.3 Scanning electron microscopy (SEM)

To examine the morphology of mucilage secreting cells (MSCs), five dried mature *Arabidopsis* seeds per background were scattered onto carbon tape and manually

oriented to lie along the basal side using forceps and then gold-coated for 3 mins using the Xenosput Sputter Coater (Dynavac Engineering. SEM was performed at 5.00 kV using the FEI/Philips XL-30 Field Emission scanning electron microscope under 1200 x magnification.

2.5 Ruthenium Red staining of mucilage

A stock solution of 0.02% (w/v) Ruthenium Red (Sigma-Aldrich) dissolved either in water or 50 mM Tris-HCl, pH 7.5 (McFarlane et al., 2014) was prepared fresh and filtered through a 0.45 µm Polyethersulfone (PES) syringe filter (Millex-HP) prior to use. Visualising the outer soluble layer of mucilage involved placing a seed in a drop of 0.02% (w/v) Ruthenium Red solution that was incubated at room temperature for 5 mins with no agitation, before imaging using the Nikon (SMZ800) dissecting microscope. To visualise the inner adherent layer, up to 50 seeds were placed in a tube containing 1 mL 50 mM Tris-HCl pH 7.5 and agitated on an orbital shaking incubator at 300 rpm, 27°C for 20 mins. Seeds were then washed with deionised (DI) water before being placed in 0.02% (w/v) Ruthenium Red solution on an orbital shaking incubator at 300 rpm, 27°C for 40 mins. Stained seeds were washed again in water before suspension in 0.1% (w/v) agar. All microscopy was performed under 30 x and 63 x magnification and captured by Nikon software ACT-2U (ver 1.70).

In addition, selected mutants with a mucilage release phenotype, along with Col-0 (wildtype) seeds were treated with either ethylenediaminetetraacetic acid (EDTA) or acid/alkali prior to Ruthenium Red staining to weaken the cell wall of the seed coat and hence improve release of mucilage from these lines. For EDTA pre-treatments, ~20 seeds were exposed to 50 mM EDTA, pH 8.0 with agitation at 300 rpm, 27°C for 20 mins. For the acid/alkali pre-treatments, ~20 seeds were exposed to 0.2 M HCl at 27°C, 300 rpm agitation for 20 mins, followed by a wash in DI water, and then incubated at 27°C in a solution of 0.5 mg/mL NaBH₄, 0.05 M NaOH with agitation at 300 rpm for 5 mins.

2.6 Yeast-based assays

2.6.1 Preparation of competent yeast

A frozen stock of *Saccharomyces cerevisiae* Y2HGold was streaked onto YPDA plates and incubated at 30°C for 3 days until growth was observed. A single, fresh colony of Y2HGold from this plate was inoculated into a starter culture of 3 mL YPDA and incubated at 30°C overnight at 250 rpm. The starter culture was subsequently inoculated into 100 mL of fresh YPDA the following day and incubated at 30°C, 250 rpm until OD₆₀₀ of 0.4–0.5 was achieved. Cells were pelleted at 700 g for 5 mins at room temperature to remove the medium and was followed by a wash step using a total of 60 mL sterile DI water. Pelleted cells were subsequently resuspended in 3 mL of freshly prepared 1.1x TE/LiAc solution (10 mM Tris-HCl, 1 mM EDTA and 100 mM LiAc) and transferred to two 1.5 mL tubes for centrifugation at 5,000 x g, 15 sec. Supernatant from each tube was discarded, and pellets were resuspended in a total of 1.2 mL of freshly prepared 1.1x TE/LiAc solution.

2.6.2 Co-transformation of yeast Y2HGold (Yeast 2-Hybrid)

Each set of candidate prey (pGADT7 constructs) and bait (pGBKT7 constructs) was individually co-transformed into competent Y2HGold cells to test for interactions. Briefly, Yeastmaker carrier DNA (Clontech) was denatured at 95°C for 5 mins and was combined at a final concentration of 10 µg/µL with 100 ng of bait and prey plasmids in a 1.5 mL tube. A 50 µL aliquot of freshly prepared competent Y2HGold cells was added and mixed by pipetting before the addition of 500 µL PEG/LiAc solution (10 mM Tris-HCl, 1mM EDTA, 100 Mm LiAc and 50% (w/v) PEG 3350). The contents were mixed gently and incubated at 30°C for 30 mins, with gentle agitation every 10 mins. Next, 20 µL of DMSO was added and mixed with the cells before a 15 min heat shock at 42 °C, with mixing performed every 5 mins. Yeast cells were pelleted at 500 g for 15 sec and supernatant was discarded and first resuspended in 1 mL 0.9% (w/v) NaCl, followed by a second centrifugation and a final

resuspension in 100 μ L 0.9% (w/v) NaCl. A 100 μ L aliquot of transformed cells was then plated onto SD/-Leu/-Trp (DDO) plates for selection and incubated at 30°C for 3-5 days until colonies were observed.

2.6.3 Sequential co-transformation of yeast Y2HGold (Yeast 3 Hybrid)

Sequential transformation of ternary factor (TFT) constructs (pTFT1 constructs) were performed following co-transformation of candidate prey and bait constructs as described in **Section 2.6.2**. These sequential transformants were then plated onto SD/-Leu/-Trp/-Ade (TDO) plates for selection and incubated at 30°C for 3-5 days until colonies were observed.

2.6.4 Protein interaction testing

Successfully co-transformed Y2HGold colonies were re-streaked onto fresh SD/-Leu/-Trp (DDO) plates and incubated at 30°C for 2 days to obtain fresh colonies. A small amount of cells from each re-streaked colony was selected using a sterile P200 tip and diluted in 10 μ L of sterile UHQ water (Millipore Milli-Q grade, 18 M Ω). Cell suspensions were further diluted 10 x and 5 μ L was plated onto a fresh SD/-Leu/-Trp/-Ade/-His (QDO) plate and a fresh QDO with 2 mM 3-AT plate. Both QDO plates were incubated at 30°C for 5 days and monitored daily for positive growth of colonies. Successful protein-protein interactions were identified through colony growth on both QDO and QDO + 3-AT plates, with known interactors SEU (bait) and AP1/SEP3 (prey) serving as positive interaction controls (Sridhar et al., 2006).

2.7 Transactivation assay

2.7.1 Infiltration of *Nicotiana benthamiana*

Constructs containing both Firefly *LUCIFERASE* (*FLUC*) expressed under the positive control *MYB PROTEIN75* promoter (*MYB75_{pro}*) or the *MUM2* promoter-intron and internal control Renilla *LUCIFERASE* (*RLUC*) under the control of the Cauliflower mosaic virus 35S promoter (*CaMV 35S_{pro}*) were introduced into *Agrobacterium* strain AGL1. These cell lines, along with AGL1 expressing relevant constructs carrying candidate transcription factors (TFs), were grown from 5 mL starter cultures in 20-150 mL of LB with appropriate antibiotics at 28-30°C with shaking at 200 rpm overnight. On all cell cultures reaching an OD₆₀₀ of ≥ 1 , cells were pelleted at 4,000 x g for 5 mins at room temperature and resuspended in an appropriate amount of infiltration buffer (10 mM MES pH 5.6, 10 mM MgCl₂ and 1 mM acetosyringone) such that final OD₆₀₀ was 4/mL. Relevant combinations of resuspended AGL1 cells carrying transactivation constructs and TFs were mixed in equal ratios with an AGL1 line carrying the p19 suppressor from tomato bushy stunt virus (TBSV) and supplemented with infiltration buffer where necessary to a final OD₆₀₀ of 1/mL. These cell mixtures were incubated at room temperature for 2 hours and subsequently infiltrated into 3-4 week old *N. benthamiana* plants on the abaxial side of fully expanded leaves. Each set of transactivation combinations was infiltrated into a single leaf on three independent plants (three biological replicates), using a fresh leaf per combination set. Infiltrated leaves were labelled appropriately and processed at six days post infiltration (DPI).

2.7.2 Dual luciferase assay

Transactivation was assessed using a Dual-Luciferase® Reporter Assay System (Promega). A disc (24 mg) of infiltrated *N. benthamiana* leaf tissue at 6 DPI was produced using a clean hole punch and subsequently snap frozen in liquid nitrogen before grinding to a fine powder with sterile, pre-chilled pestles in 1.5 mL tubes and

resuspended in 100 μ L 1 x passive lysis buffer containing 2 mM DDT. Samples were spun down at 1,600 x g for 5 mins, and the supernatant was transferred to fresh 1.5 mL tubes and stored at -20°C until needed.

A two-step luciferase assay was performed in duplicate white 96-well plates with 20 μ L of 100 x diluted samples added per well in triplicates to each plate. Luciferase Assay Reagent II (LAR II) (Promega) was reconstituted according to the manufacturer's instructions and 40 μ L added to each sample in the first plate. FLUC activity was measured across samples in the first plate for 5 mins using a FLUOstar® Omega microplate reader (BMG LABTECH) at 3,600 gain and analysed using MARS Data Analysis Software (Ver 5.11, BMG LABTECH). In the second plate, 40 μ L of 1 x Stop & Glo reagent was added to each sample and RLUC activity was measured across samples for 5 mins in the same manner. Expression of FLUC activity was normalised against RLUC activity to account for differences in transformation efficiency between samples. DNA-protein interactions were evaluated to be positive if addition of TF(s) resulted in a statistically significant increase of FLUC/RLUC ratio compared to the equivalent experimental set lacking TF(s). Statistical significance was calculated using unpaired student's *t*-test ($p > 0.05$).

2.8 Bimolecular fluorescence complementation (BiFC)

BiFC constructs containing the CDS of putative interacting proteins fused to either N-VENUS or C-VENUS halves using seamless cloning were first introduced into *Agrobacterium* AGL1. Transformed cells were then infiltrated to *N. benthamiana* leaves in the same manner as described in **Section 2.7.1**. Each set of BiFC combinations was infiltrated into a single leaf of three plants, with a fresh leaf used per combination set. A 1 cm² section was removed from the infiltrated leaf and examined for presence of VENUS fluorescence using the Olympus BX60 fluorescence microscope as outlined in **Section 2.4.2** at 3, 6 and 9 DPI with comparison to known positive and negative controls at each stage. Interactions were only evaluated as positive if fluorescence in nuclei were observed when compared to positive and negative controls.

2.9 Carbohydrate linkage analysis

2.9.1 Extraction of mucilage from mature seeds

Both the outer soluble and inner adherent layer of *Arabidopsis* mucilage were extracted in separate steps using a modified acid/alkali protocol (Walker et al., 2011). A total of 300 mg of Col-0 (wildtype), *mum2*, *shp1 shp2* and *sep3* seeds were weighed out, with 30 mg of seeds placed into ten 1.5 mL tubes per line. A weak acid treatment of 1.5 mL 0.2 M HCl was added per tube and incubated at 27°C for 20 mins, with shaking at 300 rpm to weaken the cell wall for maximum mucilage extrusion and extraction of the outer soluble layer. Samples were spun down at 5,000 x g for 5 sec to separate seeds from the soluble fractions, which were collated into individual 50 mL conical tubes (labelled as soluble fraction) per line. Acid-treated seeds were washed three times with DI water (1 mL per tube) with each wash combined with the previously collected soluble fractions.

Tubes containing acid-treated seeds were further treated with 1.5 mL freshly prepared 0.05 M NaOH containing 0.5 mg/mL NaBH₄ and incubated at 27°C for 5 mins at 300 rpm to encourage inner adherent mucilage extrusion from mutant lines. Seeds were spun down at 5,000 x g for 5 sec and washed three times with 1 mL DI water. Next, 1 mL of DI water was added per tube and sonicated for 5 sec using a 3 mm probe attached to a Branson Sonifier 450 at amplitude setting 3 for wildtype seeds and 2 for mutant seeds. Seeds were pelleted by centrifuging at 5,000 x g for 5 sec and the supernatant transferred to fresh 50 mL conical tubes (labelled as adherent fraction) per line. Three subsequent washes with 1 mL DI water per tube were combined with the adherent fraction. To confirm that seeds were not ruptured, and mucilage had been extracted, seeds were stained with Ruthenium Red.

The soluble fractions were neutralised with NaOH and the adherent fractions with glacial acetic acid, before storing at -20°C until needed.

2.9.2 Preparation and gas chromatography–mass spectrometry (GC-MS) of extracted mucilage

Neutralised mucilage fractions were dialysed overnight using Cellu-Sep T2 (8030-32, Membrane Filtration Products, Inc) washed extensively with UHQ water and freeze-dried. Next, samples were carboxyl-reduced and subsequently methylated according to established protocols (Kim & Carpita, 1992; Pettolino et al., 2012). GC-MS samples were resuspended in 10 μ L dichloromethane and 1 μ L aliquots were loaded onto a high polarity BPX70 column for analysis on the Dionex BioLC ICS 3000 system as previously described (Lau & Bacic, 1993). Polysaccharide content was estimated from the methylation data according to published methods (Shea et al., 1989).

2.10 RNA-seq

2.10.1 Tissue collection and RNA extraction

RNA was extracted from whole seeds at heart (3 DPA), linear cotyledon (7 DPA) and mature green (10 DPA) stages from Col-0 (wildtype) and *luh-5* mutants whereas RNA was only extracted from the linear cotyledon stage (7 DPA) of *shp1 shp2* double mutants. Developing whole seeds at these stages were isolated from siliques manually as described in **Section 2.3.1** and staged according to embryo morphology. Three biological replicates were collected for all genotypes, with ten siliques-worth of whole seeds per biological replicate. Isolated whole seeds were collected in 500 µL of RNeasy lysis solution (Qiagen) and kept on ice to avoid RNA degradation. All biological replicates within each genotype were collected on the same day to avoid background variation and were snap-frozen in liquid nitrogen before storage at -80°C. RNA extractions were performed according to **Section 2.3.2** prior to RNA-seq.

2.10.2 Mapping of RNA-seq reads

At least three micrograms of total RNA per sample was submitted to the Australian Genome Research Facility (AGRF), where library generation and sequencing was performed. Each library was barcoded and sequenced using 50 bp single-end reads in two dedicated lanes of a single flow cell of an Illumina Hi-seq 2000 sequencer, which served as technical replicates. Data processing steps were performed by Dr Momeneh Foroutan (WEHI).

Quality of the FastQC files was assessed using FastQC and MultiQC (Ewels et al., 2016). Trimming of adaptor sequences were not performed as it typically results in short read lengths (<30 bp) that lead to poorer alignments at multiple locations, impacting downstream analysis such as differential expression (Williams et al., 2016). Sequence alignment was performed using RSubread in R/Bioconductor to align reads to the TAIR10 reference genome (May 2017) (Liao et al., 2013). The parameters for

alignment were set such that the maximum allowed mismatch was 3, maximum indels was 5, and only unique read mapping was allowed. Genome annotation file (downloaded from RefSeq NCBI on 2nd June 2017) along with the Feature Count function available in the RSubread package were used to calculate count level data and hence the number of reads per gene. The average sequencing depth was 19 X (19 million reads), with a minimum value of 17 X (17 million reads) and a maximum value of 20 X (20 million reads), which is within the recommended sequencing depth of 10-25 million reads for differential expression profiling (Liu et al., 2014; Robinson & Oshlack, 2010). Benjamini-Hochberg (BH) was used to correct for multiple hypothesis testing and control false discovery rate (FDR) at the level of 0.05.

2.10.3 Identification of differentially expressed (DE) genes

Quality control and differential expression analysis were performed using the EdgeR package in R/Bioconductor (McCarthy et al., 2012; Robinson et al., 2010). Several visualisation functions in the EdgeR package such as Multi-Dimensional Scaling (MDS) and Principal Component Analysis (PCA) plots were used to assess the quality and variance of the data. Based on MDS and PCA plots, one biological replicate from *luh-5* at 10 DPA was considered as an outlier and was omitted from downstream analysis. Next, pair-wise comparisons of Col-0, *luh-5* and *shp1 shp2* across the same stages (3, 7 and 10 DPA for Col-0 and *luh-5*, 7 DPA for *shp1 shp2*) were performed in EdgeR, which was used to filter and normalise count data values to obtain a list of DEGs.

Filtering and normalisation steps were done in a separate manner, with the assumption that each biological set (whole seeds at 3, 7, 10 DPA) is biologically different. The separate analysis only identified genes that were above the expression threshold of 1 count-per-million (CPM), with a lower log-fold change (LFC) of 1.1 across all three biological samples. The LFC threshold of 1.1 was applied as it was the highest threshold that allowed for the detection of *SHP2* being DE in *luh-5* 7 DPA seeds,

which were known to be biologically relevant and accurate through other lines of evidence in the Golz lab.

As the library sizes and sequencing depth between samples were similar, the limma-trend pipeline was applied (Law et al., 2014). The limma-trend pipeline controls for type I error well by modelling variance at the gene level (Law et al., 2014). The *t*-tests Relative to a Threshold (TREAT) method was also used to obtain biologically relevant DE genes (McCarthy & Smyth, 2009). TREAT, which is an extension of the empirical Bayes moderated *t*-statistic that incorporates a biologically relevant fold-change threshold, results in the detection of a higher number of biologically relevant genes as DE (McCarthy & Smyth, 2009).

2.10.4 Statistical analysis

Statistical comparison of *luh-4* and *luh-5* was performed by Prof. Martina Doblin (University of Technology, Sydney). Briefly, gene expression of mutants was compared using nonmetric multidimensional scaling (nMDS) ordination based on a Bray-Curtis distance similarity measure once Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values for each replicate had been square root-transformed. Variation in gene expression was analysed with one-way analysis of similarity (ANOSIM). The ANOSIM *R* statistic (Clarke, 1993) is scaled to take a value between -1 and +1 and is interpreted as the strength of the difference between groups. *R* values are directly comparable among different studies. Pairwise comparisons were used to test for differences between pairs of samples when ANOSIM was significant.

Chapter 3

Identifying and validating differentially expressed genes in *luh* seeds

Identifying and validating differentially expressed genes in *luh* seeds

3.1 Introduction

As described in **Chapter 1**, there are three possible models describing how LEUNIG HOMOLOG (LUH) promotes *MUCILAGE MODIFIED2* (*MUM2*) activity in the developing seed coat. The first is a direct model in which the LUH directly binds to the *MUM2* regulatory regions and promotes transcription of this gene (Fig. 1.9A). According to this model, LUH functions as an activator rather than a repressor (Fig. 1.9). The two other models discussed in **Chapter 1, Section 1.3.1** propose that LUH regulates *MUM2* indirectly by modulating activity of downstream TFs (Fig. 1.9B & C). In one scenario, the LUH-regulated TF acts as a *MUM2* activator (Fig. 1.9B); whereas in the other, the downstream TF functions as a *MUM2* repressor whose activity is limited by LUH. In the latter model, LUH promotes *MUM2* expression through de-repression, which is achieved by repressing the activity of the *MUM2* repressor (Fig. 1.9C). Based on similarity to LEUNIG (LUG), it has been assumed that LUH likely functions as a transcriptional repressor (Walker et al., 2011), making de-repression the most appropriate model to describe how LUH regulates *MUM2* in the seed coat. A key prediction arising from the de-repression model is the existence of a *MUM2* repressor that is negatively regulated by LUH and is therefore elevated in *luh* mutants, resulting in a corresponding decrease of *MUM2* expression. Hence, one way that a candidate repressor(s) may be identified is to look for elevated TFs in a *luh* mutant background. This approach was taken by Dr Nadeeka Jayawardana, a former PhD student in the Golz group, and involved the use of RNA-sequencing (RNA-seq) to identify differentially expressed genes (DEGs) in developing *luh* mutant seeds following comparison to wildtype Col-0 seeds. This analysis was performed at the linear cotyledon stage (7 days post anthesis; DPA), a time when mucilage polysaccharides production and secretion is just beginning and coincides with the

activation (putative de-repression) of *MUM2* (Jayawardana, 2016; Western et al., 2000).

The initial RNA-seq analysis also included *mum2* mutant seeds, which allowed for the identification of genes that respond to the loss of *MUM2*, but not LUH activity in the *luh* mutant background. These genes were excluded from further analysis as they were not directly responsive to LUH. Based on the results of the RNA-seq experiments, 1,800 genes were found to be differentially expressed in *luh* seeds with 897 being elevated and 903 reduced (Jayawardana, 2016). As the putative LUH-regulated *MUM2* repressor is predicted to be elevated in a *luh* background, the list of DEGs was assessed for TFs displaying increased expression. This identified 63 of 128 differentially expressed TFs that were increased in *luh* seeds (Jayawardana, 2016), with many of the elevated TFs belonging to large TF families such as MADS-box (10 members), APETALA2-like (AP2-like; 4 members), MYB PROTEINS (MYB; 7 members), NAC DOMAIN CONTAINING PROTEINS (NAC; 7 members), basic helix-loop-helix (bHLH, 5 members), Cys2His2 (C₂H₂; 5 members), homeodomain (5 members) and BASIC LEUCINE ZIPPER DOMAIN (bZIP; 2 members). As significant experimentation is required to establish whether any of these TFs function as a *MUM2* repressor, the aim of this chapter is to use multiple approaches to reduce the number of TFs that need to be tested.

Previous analysis of the *MUM2* promoter/intron1-reporter revealed little or no activity in the mucilage secreting cells (MSCs) of the seed coat prior to 7 DPA, but high activity between 7 DPA and 10 DPA (Jayawardana, 2016) (Fig. 3.1A). As this observation implies that de-repression of *MUM2* occurs from 7 DPA onwards, it was reasoned that a candidate *MUM2* repressor will be active prior to 7 DPA, and that after 7 DPA, activity of the repressor (at least in the MSCs) is reduced (Fig. 3.1B). It therefore follows that in *luh* mutants, where *MUM2* is not de-repressed, the repressor remains active after 7 DPA (Fig. 3.1B). Given this prediction, RNA-seq at different developmental timepoints (3, 7 and 10 DPA) may be able to identify which of the 63 previously identified TFs are elevated in *luh* seeds at 7 DPA and beyond. An

additional dimension to these experiments was the selection of a different *luh* mutant allele (*luh-5*) to be used for the RNA-seq experiment compared to the previous experiment which used *luh-4*, an RNA null mutant with a T-DNA insertion in the 7th exon (Huang et al., 2011; Walker et al., 2011). In contrast, *luh-5* is a strong allele that arises from a nonsense mutation at Glu-97 in the 6th exon and was selected on the basis that it is phenotypically indistinguishable from *luh-4* (Walker et al., 2011). By repeating the RNA-seq experiment at 7 DPA using *luh-5*, it becomes possible to compare elevated TFs across the two different mutant backgrounds and identify TFs that are consistently mis-expressed in both *luh-4* and *luh-5*. It was reasoned that such a comparison increases the stringency of identifying TFs that are regulated by LUH and may therefore reduce the number of TFs that are considered to be *MUM2* repressors. Combining this approach with the timepoint experiments has the potential to significantly reduce the list of candidate *MUM2* repressors to a point where further analysis of all TFs becomes possible.

In addition to the above-mentioned approaches, it is also necessary to establish which of these TFs are specifically mis-expressed in MSCs, and to confirm that they are regulated by LUH in this cell layer. This can be achieved using quantitative reverse transcription-polymerase chain reaction (qRT-PCR) to look at the expression of selected genes in isolated Col-0 (wildtype) and *luh* mutant seed coat tissue, as well as promoter-reporter analysis. For the latter experiments, three tandemly fused VENUS proteins with a C-terminal N7 nuclear localisation signal (hereafter called VENUSn) was selected for use as previous work has shown that this reporter produces a strong and localized signal in the nucleus (Heisler et al., 2005). Nuclear accumulation is particularly attractive for studies of MSCs as the size of the nucleus, unlike the cytoplasm, remains relative constant throughout development and hence minimises errors associated with a fluctuating cytoplasm volume.

Based on the above analyses, a small number of MADS-box TFs were considered good candidates for being *MUM2* regulators. Subsequent analysis of mucilage release from seeds of MADS-box TF mutants suggest that they play a role in mucilage

polysaccharides production/release. In addition, a novel transgenic approach was developed so that the consequences of reducing *AGAMOUS* (*AG*) activity in just the seed coat could be examined. This approach was needed because *AG* functions early in flower development to promote stamen and carpel formation and in its absence, causes sterility (Yanofsky et al., 1990). The consequences of *AG* knockdown are described, which, surprisingly, show that *AG* has a role in regulating seed size but not mucilage polysaccharides production in *Arabidopsis*. Finally, based on the results reported in this chapter, several MADS-box TFs were identified as putative *MUM2* regulators and their function investigated further in later chapters.

3.2 Results

3.2.1 RNA-seq on *luh-5* compared to *luh-4*

Based on previous work by Dr. Nadeeka Jayawardana, *MUM2* shows high expression in the MSCs from the linear cotyledon stage (7 DPA) onwards in a Col-0 (wildtype) background (Fig. 3.1A) (Jayawardana, 2016). This activity is apparently absent in the MSCs, but still present in the inner endothelial layers from the heart stage (3 DPA) onwards in a *luh* background (Fig. 3.1A). Having established that the *MUM2* activity is low in the early stages of seed coat development (before 7 DPA) in wildtype seeds (Fig. 3.1A), it is likely that the putative *MUM2* repressor is highly expressed at these early seed coat developmental stages to prevent early activation of *MUM2* (Fig. 3.1B). Subsequently, the putative *MUM2* repressor is repressed by LUH in the later seed coat developmental stages (7 DPA and beyond) thereby allowing *MUM2* to be expressed (Fig. 3.1B). Hence, in a *luh* background, it is predicted that the putative *MUM2* repressor will remain expressed during the later stages of seed coat development (7 DPA and beyond).

To examine the temporal expression patterns of candidate *MUM2* repressors, RNA-seq was performed on *luh-5* and Col-0 whole seeds at heart (3 DPA), linear cotyledon (7 DPA) and mature green (10 DPA) stages. This analysis served two purposes: firstly, to identify differentially expressed TFs that are mis-expressed at different developmental stages and secondly, to identify a common set of differentially expressed TFs within two different *luh* mutants (*luh-4* and *luh-5*) at the linear cotyledon stage of development. As *luh-4* and *luh-5* mutants have indistinguishable mucilage extrusion phenotypes, it is expected that they will display similar changes in gene expression. One caveat is that *luh-4* and *luh-5* are in slightly different genetic backgrounds with *luh-4* in a Col-0 background and *luh-5* in a Col-2 background (Huang et al., 2011; Walker et al., 2011). While background differences may influence the identification of DEGs, especially since Col-0 was used as a wildtype control for both experiments, it is expected that a comparison between the two datasets will identify genes that are consistently mis-expressed and hence lead to a

reduced list of high-confidence DEGs from which elevated TFs can be identified for further analysis.

To enable direct comparisons between the *luh-4* and *luh-5* RNA-seq datasets, raw data from both mutants and Col-0 were processed using the same pipeline as described in **Chapter 2, Section 2.10.3**. A standard expression threshold of 1 count-per-million (CPM) and a log-fold change (LFC) threshold of 1.1 was applied across all samples. This analysis generated a list of DEGs detected in *luh-4* and *luh-5* relative to the Col-0 control at 7 DPA. DEGs detected in *luh-4* and *luh-5* were compared to identify genes that are mis-expressed in both backgrounds (Table 3.1). It is expected that some differences in the DEGs detected in *luh-4* and *luh-5* might arise from experimental variation. For instance, RNA extracted from the *luh-4* and *luh-5* mutants were of slightly different quality, with RNA integrity number (RIN) values above 8 for *luh-4* and above 7 for *luh-5* (Appendix 3, Supplementary Table 1) (Jayawardana, 2016). There might also be subtle differences in the developmental stage of *luh-4* and *luh-5* seeds used for each experiment as plants were grown several years apart in growth rooms that could potentially have a different light intensity and temperature to when the first experiment was performed. Because of this, it is possible that 7 DPA may actually correspond to slightly different stages of seed development in the second experiment. To determine the extent of variation between the *luh-4* (performed in 2014) and *luh-5* experiments (performed in the current study), FPKM values for all genes expressed in the Col-0 controls (2014 and current study) were compared to each other using a one-way analysis of similarity test, ANOSIM (**Chapter 2, Section 2.10.4**). ANOSIM determined that the *luh-4* and *luh-5* mutants were different to Col-0 (Global R = 1.0, p = 0.0002) but not significantly different from one another (pairwise test: *luh-5* vs *luh-4*, p = 0.1).

As expected, the RNA-null allele, *luh-4*, produced a higher fold decrease of *LUH* expression (124.6-fold) compared to the non-sense allele *luh-5* (2.2-fold) (Appendix 3, Supplementary Table 2 & 3). However, a similar fold decrease in *MUM* expression is observed in each mutant, with *luh-4* producing an 8.8-fold decrease of *MUM2* while

luh-5 produced a 10.2-fold decrease (Appendix 3, Supplementary Table 2 & 3). Again, this is expected as both mutant alleles are non-functional (Sitaraman, 2004). Interestingly, more DEGs were detected in *luh-5* (4371) than *luh-4* (2988), with a set of 1519 DEGs common to both *luh-4* and *luh-5* but not *mum2-1* (Fig. 3.2A, Appendix 3, Supplementary Table 2, 3 & 4). These differences in DEGs are most likely due to the experimental variation arising between the two RNA-seq experiments, although some may be a consequence of the Col-2 rather than Col-0 background of the *luh-5* allele as described in the section above.

Another potential issue with the data arising from the RNA-seq analysis is that sampling whole seeds may result in genes from some seed compartments (embryo/endosperm) being over-represented in the list of DEGs relative to those of the seed coat. While this would not normally be a problem if LUH functions solely in the seed coat, published literature has shown that LUH functions redundantly with LUG to promote embryo development and hence clearly has roles outside of the seed coat (Sitaraman et al., 2008). Given that this study is mostly interested in DEGs that are mis-expressed in the seed coat, DEGs were sorted according to their compartment-specific activity. This was achieved using a publicly available dataset that lists the regional expression pattern of genes within the developing *Arabidopsis* seed (Belmonte et al., 2013). This extensive microarray dataset includes seven sub-regions of the wildtype Wassilewskija (Ws-0) seed (embryo proper, suspensor, micropylar endosperm, peripheral endosperm, chalazal endosperm, chalazal seed coat and general seed coat) across five different seed development stages (pre-globular, globular, heart, linear cotyledon and mature green) and is hereafter referred to as the Belmonte dataset. One limitation of the Belmonte dataset is that the spatial and temporal expression of genes are only available for the Ws-0 background, hence the dataset can only be used to indicate the likely spatial expression of genes within Col-0 seed as differences between the two backgrounds cannot be ruled out. Nevertheless, the Belmonte dataset can still be used to indicate which DEGs are likely to be active in the *Arabidopsis* seed coat. Overall, use of the Belmonte dataset identified 1149/1519

DEGs (75.6%) common to both *luh-4* and *luh-5* but not *mum2-1* that are expressed in, but not limited to the seed coat (Fig. 3.2B).

Next, DEGs that are common to both *luh-4* and *luh-5* were compared to a group of 63 TFs that were originally identified as being elevated in *luh-4* seeds (Jayawardana, 2016). Of these 63 genes, seven were not found to be differentially expressed when the previous *luh-4* dataset was analysed using the new, more stringent RNA-seq pipeline, and hence were removed from further analysis. The remaining 56 genes were sorted according to whether they are elevated in both *luh-4* and *luh-5* (35/56; 62.5%) and therefore were higher-confidence DEGs or only elevated in *luh-4* but not *luh-5* (21/56; 37.5%) (Table 3.1).

To identify the most likely candidate *MUM2* repressors, TFs denoted as being predominantly expressed in, but not limited to the seed coat, that were elevated in *luh* mutants (Table 3.1, purple shading) were selected for further characterisation (Belmonte et al., 2013; Winter et al., 2007). Specifically, these TFs consisted of 15 TFs elevated in both *luh-4* and *luh-5* mutants and five that were *luh-4*-specific (Table 3.1, red text). The latter were selected because they include genes of interest such as MADS-box TFs *SEP3* (*AT1G24260*), *SHP1* (*AT3G58780*) and *AG* that have roles in floral development and therefore may have unidentified roles in seed development, or other genes that have known roles in MSC development such as *NAC-REGULATED SEED MORPHOLOGY2* (*AT1G52880*, *NARS2*) and *NAC080* (Jayawardana, 2016; Kunieda et al., 2008).

A detailed analysis of expression of these TFs in seeds using publicly available microarray data revealed 17/20 TFs with broad expression in the general seed coat and three that are primarily expressed in the chalazal region of the seed coat (Table 3.2) (Belmonte et al., 2013; Winter et al., 2007). The three TFs primarily expressed in the chalazal region, *VND-INTERACTING 2* (*AT5G13180*, *VNI2*), *BASIC LEUCINE ZIPPER DOMAIN44* (*AT1G75390*, *bZIP44*) and *VERDANDI* (*AT5G18000*, *VDD*) were excluded from further analysis as they were considered unlikely to play an

important role in *MUM2* regulation given that *MUM2* is expressed uniformly throughout the inner layers of the seed coat from 3-7 DPA and then extends into the MSC from 7 DPA onwards (Jayawardana, 2016) (Winter et al., 2007). In addition, due to limited quantities of available cDNA for qRT-PCR analysis, two other shortlisted genes, *AT5G60140* and *ILI1 BINDING BHLH 1 (AT2G43060, IBH1)* were also excluded from further analysis on the grounds that they are not associated with known mucilage defects or with MSC tool box genes in published literature (Table 3.2, red text) (Francoz et al., 2015).

Expression of the remaining 15 elevated TFs were examined in isolated seed coat tissue using qRT-PCR, with a subset then being carried through to promoter-reporter analysis (**Sections 3.2.2 and 3.2.3**). In parallel with qRT-PCR, differential expression of these 15 elevated TFs were also examined in an RNA-seq time-course (described in **Section 3.2.1.1**) performed on Col-0 and *luh-5* whole seeds across early to late developmental stages to provide additional insight into their regulation. This RNA-seq time-course allows for the identification of TFs that are elevated in *luh* mutants beyond 7 DPA as predicted of a candidate LUH-regulated *MUM2* repressor (Fig. 3.1B). Collectively, these analyses were performed to identify a small subset of LUH-regulated TFs that are elevated in both *luh-4* and *luh-5* and hence are strong candidates for being *MUM2* repressors according to the model described in **Chapter 1, Section 1.3.1** (Fig. 1.9C).

3.2.1.1 Characterising gene expression changes in *luh* mutants at different stages of seed development

To further understand the transcriptional changes that occur in *luh* mutant seeds beyond linear cotyledon stage (7 DPA), RNA-seq analysis was performed on seeds at both heart (3 DPA) and mature green (10 DPA) stages of development. A total of 4676 DEGs were detected in early stages (heart), and 2793 DEGs in late stages (mature green) of *luh-5* mutant seeds (Fig. 3.3A & Appendix 3, Supplementary Table

2, 6 & 7). Interestingly, each developmental stage appeared to be somewhat distinct from the others, with many DEGs detected at only a single time-point (3 DPA: 2403/4676, 51.4%; 7 DPA: 1203/4371, 27.5%; 10 DPA: 940/2793, 33.6%). However, a set of 757 DEGs common to all developmental stages was identified of which 626 are seed coat-expressed (Fig. 3.3B).

Given that the wildtype backgrounds for the electronic Fluorescent Pictograph (eFP) browser (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) and RNA-seq time-course differed (Ws-0 vs Col-0/Col-2), the RNA-seq time-course was used to evaluate the general accuracy and utility of the eFP dataset (Winter et al., 2007) across early to late developmental stages for the LUH-regulated TFs identified in **Section 3.2.1**. As the RNA-seq time-course was performed on whole seeds and the eFP browser information is derived from individual seed tissues, comparison of these datasets was used to validate temporal expression patterns of these genes in their respective wildtype backgrounds. The comparative analysis between the eFP dataset and RNA-seq time-course on Col-0 and *luh-5* developing seeds at the heart, linear cotyledon and mature green stage revealed that the expression pattern of 7/15 candidate *MUM2* repressors shortlisted from Table 3.2 had similar temporal expression, although were not quantitatively similar to each other (Table 3.3). This indicates that the eFP browser was generally useful in predicting temporal, but not quantitative expression patterns. Of the remaining 8/15 candidate *MUM2* repressors, four TFs displayed patterns that mostly matched in the different wildtype backgrounds (SHP1, MYB3, TTG2 and NARS2), while the other four TFs, *STK*, *SEP1*, *AP2* and *TT8*, had RNA-seq time-course expression patterns that significantly differed from the eFP dataset (Table 3.3). This difference was attributed to a significantly large increase of expression in the mature green stage in the eFP dataset but not the RNA-seq time-course, which indicates potential differences in the staging of mature green seeds or differences associated with the wildtype backgrounds used in the eFP (Ws-0) and RNA-seq (Col-0) datasets. Overall, results from this comparative analysis indicated that the eFP dataset is generally a useful guide for predicting temporal expression patterns across developing seeds.

Next, the data from the RNA-seq time-course experiment was queried to allow assessment of the temporal expression pattern of the elevated TFs at early (heart) and late (mature green) developmental stages in *luh* mutants. Given that *MUM2* is expressed highly in the seed coat from the linear to the bending cotyledon stage onwards (Jayawardana, 2016) (Fig. 3.1A), it is likely that a *MUM2* repressor would be most active during the early stages of seed development, with little or no activity from the linear cotyledon stage onwards (Fig. 3.1B). As LUH is predicted to modulate the activity of the *MUM2* repressor (Fig. 1.9C), the repressor might be expected to be elevated from the linear cotyledon stage onwards in *luh* mutants (Fig. 3.1B). An initial inspection of the RNA-seq data revealed only four of the 15 LUH-regulated TFs, *SHP2*, *AP2*, *MYB3* and *TT8*, to be elevated across all three developmental stages in *luh-5* mutants (Table 3.4). These expression patterns correspond to the predicted activity of a *MUM2* repressor in *luh* seeds (Fig. 3.1B). Three other TFs have a similar pattern, with *STK*, *SEP1* and *TTG2* elevated across both heart and linear cotyledon stage, but not the mature green stage (Table 3.4). These TFs are still good candidate *MUM2* repressors as *MUM2* promoter activity is decreased slightly at the mature green stage (Fig. 3.1B), meaning that *MUM2* repressors may be active at this stage following release from LUH repression and therefore may not be elevated in *luh* at 10 DPA. The remaining TFs (Table 3.4) are not consistent with the predicted activity of a *MUM2* repressor in the *luh* background (Fig. 3.1B). However, as this RNA-seq time-course was performed on whole seeds, it is possible that these TFs are not detected as mis-expressed due to the dilution effect of other tissue transcripts (i.e. endosperm, embryo). For example, the expression of a TF in different tissues in the seed besides the seed coat in opposing directions can lead to the masking of the actual mis-expression in the seed coat. For this reason, all 15 TFs were also examined by qRT-PCR on isolated seed coat tissue to obtain tissue-specific resolution of their mis-expression (**Section 3.2.2**).

In summary, the RNA-seq time course analysis identified seven LUH-regulated TFs that are elevated in the general seed coat and are consistently elevated at two or more

stages of seed development in *luh-5* mutants, which aligns with the expected properties of a *MUM2* repressor (Table 3.4, bold text).

3.2.2. Validating selected TFs using qRT-PCR analysis

Based on the RNA-seq analysis and database mining, 15 TFs were identified as being elevated in *luh* mutants and expressed in, but not necessarily restricted to the seed coat (Table 3.2). Given that the RNA-seq experiments were performed on RNA isolated from whole seeds, it is important to confirm that these genes are indeed expressed in the seed coat, specifically MSCs, as well as being elevated in *luh* mutants. Of the number of approaches that could be taken to validate the RNA-seq analysis, qRT-PCR on isolated seed coat material is perhaps the easiest where expression of a number of genes needs to be assessed. Additionally, due to its high sensitivity relative to RNA-seq, qRT-PCR is particularly well-suited for analyses in which the amount of starting material is low (Onate-Sanchez & Vicente-Carbajosa, 2008). Complicating this approach, however, are difficulties associated with the isolation of seed coat material: the lack of specialised equipment such as a laser-directed micro dissector and issues associated with manual separation of seed tissues such as RNA degradation and the induction of stress responses that may alter gene expression (Khan et al., 2014). This can be partially overcome by using liquid nitrogen to pre-chill all associated equipment involved in the isolation of the seed coat, as well as the use of RNA stabilisation storage solution such as RNeasy as described in **Chapter 2, Section 2.10.1**. Manual separation also cannot completely remove endosperm tissue from isolated seed coat tissue (Dean et al., 2011).

For this study, RNA was extracted from both seed coat and embryo tissue isolated from *luh-4* and Col-0 seeds at the linear cotyledon stage (7 DPA) of development. In particular, RNA from embryos was used to investigate instances where candidate *MUM2* repressors were elevated in *luh* whole seeds but not in *luh-4* seed coats by examining their expression in embryos. To determine the activity of a gene in

different genetic backgrounds and/or tissues, it is necessary to measure amplification of the target gene relative to at least one housekeeping gene that has equivalent expression (denoted by ΔC_t values within ± 1) across the various sampled backgrounds (Freeman et al., 1999; Radonic et al., 2004). While it is best practice to use more than one reference gene when assessing gene expression by qRT-PCR, two housekeeping genes *TUBULIN7* (*TUB7*, AT2G29550) and *ACTIN2* (*ACT2*, AT3G18780) were tested, and only *TUB7* was found to be stably expressed (denoted by ΔC_t values within ± 1) in both *luh-4* and Col-0 seed coats and embryos. Purity of seed coat and embryonic RNA was subsequently tested by examining expression of the seed coat-specific gene *DIRIGENT PROTEIN1* (*DP1*, AT4G11180), and the embryo-specific gene *AT1G04660* (Esfandiari et al., 2013; Winter et al., 2007). This revealed an 1873-fold enrichment for *DP1* in the seed coat samples compared to embryonic samples and a 16-fold enrichment for *AT1G04660* in the embryonic samples compared to seed coat samples. These values exceed those calculated from publicly available data in which 805-fold and 6.5-fold enrichments were expected in these respective tissues (Winter et al., 2007), suggesting the seed tissue isolation performed in this study had been successful. A list of primers used for qRT-PCR is given in Appendix 3, Supplementary Table 11.

Having verified that RNA isolated from seed coat and embryo showed the expected enrichment for genes expressed in those regions, qRT-PCR was next used to examine expression of the 15 elevated TFs (Table 3.2) in seed coat tissue isolated from Col-0 and *luh* mutants (Table 3.5). Twelve genes had elevated expression in the *luh-4* samples, which is consistent with data obtained by RNA-seq (Table 3.5, in bold text). Interestingly, expression of *SEP1*, *TT2* and *NARS2* was significantly higher in the qPCR analysis of *luh* mutants relative to Col-0 (>50 fold) when compared to the fold elevation detected by RNA-seq analysis. This difference might reflect the more sensitive nature of qRT-PCR analysis compared to RNA-seq, and/or the difference in source material used for each experiment (isolated seed coat versus whole seeds, respectively) (Liu et al., 2011). Overall, four MADS-box TFs, *SEP3*, *SH2*, *AP3* and *STK*, along with tier 3 TFs *MYB5*, *TT2* and *NARS2* (**Chapter 1, Section 1.2.1**) were

selected for further examination via promoter-reporter analysis. Of the MADS-box TFs, *AG* was excluded due to the complexity of generating a promoter-reporter construct; the 2nd intron is known to be required for proper expression (Sieburth & Meyerowitz, 1997). Meanwhile, *SEP1* was excluded on the basis of being similar to *SEP3*, which was of particular interest due to its ability to complex with the other MADS-box TFs of interest such as *AP3* (Jetha et al., 2014), *SHP2* and *STK* in *Arabidopsis* (Favaro et al., 2003).

Interestingly, *SHP1* was identified as being elevated by RNA-seq in *luh-4* at the linear cotyledon stage, but not at any developmental stage in *luh-5* mutants (Table 3.1 & 3.4), suggesting that *SHP1* is either not truly elevated or only marginally elevated in *luh-4* seeds (Table 3.1). An examination of the raw ΔCt readings of *SHP1* expression in the three biological replicates of both Col-0 and *luh-4* revealed that the third biological replicate produced an outlier (ΔCt value difference of -2 from closest triplicate value), which could contribute to the masking of statistically significant differential expression detected in *luh-4* by qRT-PCR (Appendix 3, Supplementary Table 9). Removal of this outlier did not result in a statistically significant differential expression, possibly due to the lower sample number. Thus, to ascertain if there is differential expression of *SHP1* in the *luh* seed coat, *SHP1* was included in downstream analysis using promoter-VENUSn as outlined below (**Section 3.2.3**).

Lastly, both the expression of *TRANSPARENT TESTA GLABRA2* (*AT2G37260*, *TTG2*) and *APETALA2* (*AT4G36920*, *AP2*), were found to be reduced in *luh-4* seed coat, even though RNA-seq analysis indicated that they were elevated in *luh-4* and *luh-5* whole seeds (Table 3.1, red text). A possible explanation for this observation is that while *TTG2* and *AP2* may be reduced within the seed coat, they are significantly elevated in other compartments of the seed. As the RNA-seq experiment assessed expression in the whole seed, compartment-specific expression differences are not apparent. In support of this hypothesis, there was an observed 1.6-fold increase of *TTG2* expression in *luh-4* embryos compared to wildtype (Appendix 3, Supplementary Table 10). Unfortunately, due to a lack of embryonic cDNA,

expression of *AP2* in embryos could not be assessed within the timeframe of this project. Given that both genes had reduced expression in *luh-4* seed coat, they were excluded from any further analysis. In summary, Fig. 3.4 provides an overview of the screening process of the initial 56 genes to the eventual selection of the eight genes for further examination with promoter-reporter constructs.

Overall, qRT-PCR analysis confirmed that 12 TFs are mis-expressed in the seed coat of *luh* mutants. Due to the labour-intensive nature of obtaining and studying promoter-reporters in *Arabidopsis*, seven of these verified genes, along with *SHP1* (Table 3.5, shaded in purple), were selected for further promoter-reporters analysis (**Section 3.2.3**) to provide greater spatial resolution of their differential expression in *luh* and Col-0 seed coat.

3.2.3 Investigating expression of selected TFs using promoter-reporter analysis

To provide further evidence that the TFs identified by RNA-seq and confirmed by qRT-PCR are elevated in the seed coat of *luh* mutants, and to determine their precise spatial and temporal pattern of mis-expression, promoter-*VENUSn* reporters for a subset of eight TFs were generated and introduced into Col-0 and *luh* backgrounds (Fig. 3.4). A major drawback of the promoter-reporter approach, however, is the length of time required for the generation of stable transgenic lines (Voss et al., 2013). In addition, the positional effect from the site of transgene integration is associated with variation in reporter activity within the same batch of transformants (Gelvin, 2003; Szabados et al., 2002). As a consequence, promoter-reporter comparisons between different genetic backgrounds/tissues requires generating multiple independently transformed lines. An alternative approach is to introgress the same transgene into two different genetic backgrounds through crossing with the resulting reporter activity in the two backgrounds being directly comparable. Accordingly, *luh* mutant were transformed with a *VENUSn* promoter-reporter, and T₁ hemizygous plants subsequently crossed to both *luh* and wildtype plants. Resulting progeny were

either homozygous for *luh* (*luh/luh*) or heterozygous (*luh/+*; phenotypically wildtype), and both hemizygous for the reporter following selection. This approach results in both classes of progeny (*luh/luh* and *luh/+*) possessing the promoter-reporter at exactly the same chromosomal position and hence differences in reporter activity can be attributed to genetic background rather than positional effects. This approach has been used extensively throughout this thesis to investigate promoter activity of candidate LUH-regulated TFs in the *luh* mutant and a *luh/+* background.

Promoter-*VENUSn* reporters served a dual purpose of confirming promoter activity in the seed coat and revealing differential activity in *luh* and *luh/+* backgrounds (Fig. 3.5 & 3.6). Due to the time-consuming and labour-intensive process of generating promoter-*VENUSn* reporters, only eight genes across several TF families were analysed. This included five MADS-box TFs (*AP3*, *SHP1*, *SHP2*, *STK* and *SEP3*), two MYB TFs (*MYB5* and *TT2*) and one NAC-domain TF (*NARS2*). Promoter regions of these genes (generally 1-3 kb upstream of the coding region depending on intergenic sequence length) were amplified from genomic DNA by PCR and cloned seamlessly into shuttle vector *pMIGRO* containing the *VENUSn* reporter (JG1000) as described in **Chapter 2, Section 2.2.11**. The promoter-reporter cassette was then introduced into the *NotI* site of binary vector *pMLBART* (JG13) using standard restriction-ligation methodology. All primers and plasmids used to generate these constructs are listed in Appendix 3, Supplementary Table 12 & 13, except for *SEP3*, as Col-0 seeds containing the promoter and genomic region of *SEP3* fused to GFP (*gSEP3-GFP*) were kindly provided by Gerco Angenent (Wageningen University and Research). These promoter-*VENUSn* reporter constructs were then introduced into *luh-5* mutants by floral dip transformation and transgenic T₁ seedlings identified following Basta selection.

Developing seeds from T₁ plants were then examined for fluorescence. Sixteen transgenic *luh-5* plants were examined per transgenic background, and one representative fluorescent line (based on fluorescence strength and presence of a single T-DNA) was selected for further analysis, which included crossing to both Col-

0 and *luh-5*. Similarly, *gSEP3*-GFP seeds were crossed to *luh-5* to generate a *luh-5/+* line hemizygous for the *gSEP3*-GFP transgene. The line was then backcrossed to *luh-5*, and the resulting progeny sown onto soil following selection for the GFP reporter on hygromycin plates. Transgenic *luh-5/+* and *luh-5* individuals were identified by PCR-based genotyping and fluorescence in developing seeds examined.

One expectation of a *MUM2* repressor is its expression in the epidermal layer of the seed coat (MSCs) during the early stages, but not late stages of wildtype seed development (Fig. 3.1B). In contrast, it is expected that the *MUM2* repressor will be ectopically expressed during the late stages of *luh* seed development (Fig. 3.1B). Complicating these expectations is the observation that the *MUM2* promoter is also active in other seed compartments such as the inner seed coat layers and embryo tissue (Jayawardana, 2016; Macquet et al., 2007b). Thus, putative *MUM2* repressors may also be present in these compartments, although the extent to which they would be repressed by LUH in this tissue is presently not known. Taken together, candidate *MUM2* repressors are expected to be present in the MSCs of both Col-0 and *luh-5* developing seeds, and significantly elevated in the MSCs of *luh-5* at later stages of seed development.

Examination of promoter activity in early to late stages of seed development revealed that *AP3*, *SEP3*, *STK*, *SHP1*, *SHP2* and *NARS2* were active in the MSCs from early (globular; 2 DPA) to late stages (mature green, 10 DPA) in both *luh/+* and *luh* backgrounds (Fig. 3.5 & 3.6). Promoter activity of *AP3*, *SHP1* and *SHP2* were noticeably elevated in the MSCs of *luh-5* compared to wildtype from the linear cotyledon stage (7 DPA) onwards (Fig. 3.5) although all the aforementioned genes apart from *STK* displayed this elevated pattern upon closer inspection (Fig. 3.7). These observations support the RNA-seq analysis that identified *AP3*, *SEP3*, *SHP1*, *SHP2* and *NARS2* as being elevated in *luh* seeds at the linear cotyledon stage of development and provides greater spatial resolution of their expression in the seed coat that is consistent with the predicted activity of putative *MUM2* repressors in *luh*. No noticeable changes were observed in *STK* promoter activity, with comparable

levels of fluorescence signal detected in both *luh/+* and *luh* backgrounds across all stages (Fig. 3.5). Subsequent inspection of *STK_{pro}::VENUSn* fluorescence signal in corresponding embryos revealed increased VENUSn activity in *luh* compared to *luh/+*, which may account for the observed increased in *STK* expression detected by both RNA-seq and qRT-PCR (Appendix 3, Supplementary Fig. 1). This indicates that LUH represses *STK* in the embryo rather than the seed coat, and therefore *STK* is unlikely to be involved in *MUM2* regulation in the MSCs.

The remaining two promoters, *MYB5_{pro}* and *TT2_{pro}*, were active in the seed coat only at certain stages (Fig. 3.6), and not in the embryo (Appendix 3, Supplementary Fig. 2). For instance, *MYB5_{pro}* activity was initially observed in the inner seed coat layers, only appearing in the MSCs in both *luh/+* and *luh* mutants at later stages (linear cotyledon), albeit with increased fluorescence signal in *luh* (Fig. 3.6 & 3.7). As *MYB5_{pro}* is only active in the inner seed coat layers, and not in the MSCs during early developmental stages (Fig. 3.6), it is unlikely that MYB5 functions as a *MUM2* repressor (Fig. 3.1B), where it should keep *MUM2* repressed at early stages according to the model proposed in Fig. 3.1. Similarly, the promoter activity of *TT2* was observed from mid (torpedo, 5 DPA) to late stages (mature green) in the MSCs of *luh/+* seeds but was absent in the MSCs of corresponding *luh* mutants and only elevated in the inner seed coat layers/endosperm (Fig. 3.6 & 3.7). While *TT2_{pro}::VENUSn* had elevated activity in *luh* mutants at the linear cotyledon stage, which is consistent with RNA-seq data, this increase was limited to the inner seed coat layer/endosperm with no expression of *VENUSn* in the MSCs (Fig. 3.7). This indicates that LUH may be regulating *TT2* in the endothelium rather than in the MSCs. *TT2* promoter activity in *luh* is therefore not consistent with the expectation of a *MUM2* repressor.

Overall, elevated promoter activity of the majority of MADS-box TF promoters such as *AP3*, *SEP3*, *SHP1* and *SHP2* is consistent with the results of the RNA-seq/qRT-PCR analyses. Finding that many of these TFs are repressed by LUH in the MSCs is a key prediction arising from the depression model of *MUM2* regulation, suggesting

that one or more of these TFs may function as a *MUM2* repressor. To investigate this possibility further, mucilage extrusion phenotypes of lines defective in MADS-box TF activity were examined in the following section.

3.2.4 Examining mucilage defects in mutants of candidate *MUM2* repressors

According to the de-repression model, seeds lacking the putative *MUM2* repressor are still expected to release mucilage, as *MUM2* is no longer repressed. Previous work has shown that lines carrying *35S_{pro}::MUM2* produce seeds with normal mucilage release, suggesting that increased *MUM2* levels have no negative consequences for mucilage production and subsequent release (Dean et al., 2007; Macquet et al., 2007b; Walker et al., 2011). Thus, crossing mutants of candidate *MUM2* repressors with *luh* should restore mucilage extrusion due to loss of the *MUM2* repressor and the consequential increase in *MUM2* expression. At the time of this study, mutations in MADS-box TF family were not known to produce mucilage defects whereas some NAC-domain and MYB TFs such as NARS2 and MYB5 had well characterised phenotypes (Gonzalez et al., 2009; Kunieda et al., 2008). In addition, MADS-box TFs were particularly attractive as *MUM2* repressors as their function in the seed coat was unknown, while the other identified TFs have been clearly associated with mucilage polysaccharides regulation.

Apart from the sterile mutant *ag* (Yanofsky et al., 1990), MADS-box TF mutants were first assayed for their mucilage phenotypes using Ruthenium Red staining (see **Chapter 2, Section 2.5** for method). A range of mucilage extrusion defects were observed, which is inconsistent with these TFs as *MUM2* repressors. Briefly, *sep3* (CS910405) produced seeds with a slightly deformed shape, subtle mucilage defects in the inner adherent layer and severe outer mucilage defects, with a marked reduction in mucilage release compared to Col-0, while *shp2-1*, *shp1-1 shp2-1* double mutants and *stk* (SALK_206790C) all had pronounced extrusion defects associated with the inner adherent layer, resulting in visibly less mucilage extrusion (Fig. 3.8 & Appendix

3, Supplementary Fig. 3D). In contrast, *ap3* (SALK_036387C), a male-sterile mutant, produced seeds that extruded normal amounts of mucilage when pollinated with wildtype pollen (Appendix 3, Supplementary Fig. 3 B). *sep1* (SALK_121233C) and *shp1-1* also had a wildtype pattern of mucilage extrusion (Appendix 3, Supplementary Fig. 3 G & H). As the mucilage defects of *stk* has previously been characterised and shown to be a consequence of reduced *PMEI6* activity (Ezquer et al., 2016), no further characterisation of this line was performed as the *stk* defects do not apparently reflect changes in *MUM2* activity. Interestingly, although *shp1-1* alone does not produce visible mucilage defects, the double *shp1-1 shp2-1* mutant produces a phenotype that is more severe than that of the single *shp2-1* mutant, suggesting that SHP1 has some role in regulating mucilage extrusion redundantly with SHP2. To confirm the phenotypes of *shp1-1* and *shp2-1*, additional alleles of these mutants (*shp1-058C*, *shp1-457C*, *shp2-772C*) were identified and tested for mucilage release. In each case, the mucilage release phenotype was similar to the other mutant allele (Appendix 3, Supplementary Fig. 3G-I), confirming that SHP1, SHP2 and SEP3 have a role in regulating mucilage release, but not apparently as *MUM2* repressors.

Mucilage defects in these mutants may arise from 1) compositional changes in RG-I, 2) an alteration in cell wall composition/organisation resulting in different mucilage hydration properties, 3) changes in the amount of mucilage polysaccharides synthesised in the MSCs, and/or 4) increased mechanical strength of MSC radial walls. To test the last of these possibilities *shp1-1*, *shp2-1*, *shp1-1 shp2-1* and *sep3* mutant seeds were pre-treated with either ethylenediaminetetraacetic acid (EDTA; Fig. 3.8L-O) or acid/alkali (Fig. 3.8Q-T) to weaken the cell wall (Saez-Aguayo et al., 2014) facilitating the release of mucilage prior to Ruthenium Red staining. Both pre-treatments resulted in additional release of mucilage from Col-0, *shp1-1*, *shp2-1*, *shp1-1 shp2-1* and *sep3* mutant seeds compared to non-treated controls (Fig. 3.8). While the mucilage release phenotype of *shp1-1* was similar to wildtype, *shp2-1*, *shp1-1 shp2-1* and *sep3* mutants released additional mucilage compared to the non-treated control (Fig 3.8). These observations suggest that changes to the hydration properties of mucilage polysaccharides may account for the mucilage release defects

seen in these mutant lines, which is reminiscent of *luh* and *mum2* mutants (Walker et al., 2011).

3.2.4.1 Genetic interactions

As MADS-box TFs form quartets in *Arabidopsis* to facilitate transcriptional regulation (Fig. 1.13) and act in a partially redundant manner (Theissen & Saedler, 2001), higher order mutants were generated to determine if they produce stronger mucilage extrusion defects. To explore the possibility of genetic interactions between MADS-box TFs in regulating mucilage polysaccharides production, examination of mucilage extrusion defects was expanded to sterile *ap1* (Irish & Sussex, 1990) and *pistillata* (*pi*) (Krizek & Meyerowitz, 1996) mutants. Both AP1 and PI are MADS-box TFs that are involved in specifying floral organ identity and were not detected as differentially expressed by RNA-seq in *luh* mutants. Although sterile, *ap1* and *pi* have functional carpels and were pollinated with wildtype pollen to generate seeds. The resulting seeds have seed coats that are maternally derived and hence are homozygous for *ap1* and *pi*. In both cases, wildtype-like mucilage was released from these seeds (Appendix 3, Supplementary Fig. 3E & F). Both AP1 and PI are therefore unlikely to have major roles in the mucilage regulatory process.

Next, MADS-box TF mutants that were identified as having mucilage extrusion defects were crossed to generate the following higher order mutant combinations: *sep3 shp1*, *sep3 shp2* and *sep3 shp1 shp2*. In addition, to determine if SEP1, the only other SEP member that is expressed at high levels in the seed coat based on eFP data (Winter et al., 2007), may function redundantly with SEP3 for mucilage polysaccharides regulation, crosses of *sep1* to *sep3* were included. Microscopic analysis of *sep1 sep3* seeds following Ruthenium Red analysis revealed not only a lack of mucilage extrusion, but severely deformed seeds, indicating that SEP1 and SEP3 likely have roles in upstream seed patterning that subsequently impact downstream mucilage polysaccharides production (Fig. 3.9H). Given that single mutant *sep1* produces wildtype-like seeds (Appendix 3, Supplementary Fig. 3C), it is likely that SEP1 and SEP3 act redundantly, with SEP3 being the predominantly active

member controlling early seed development. Interestingly, crosses between *sep* and *shp* mutants such as *shp1 sep3* and *shp2 sep3* double mutants enhanced the mucilage polysaccharides defects of both *shp1* and *shp2*, suggesting that SEP3 may work in the same pathway of SHP1 and SHP2 to regulate mucilage release (Fig. 3.9B, F & C, G). Further evidence supporting this hypothesis is the enhanced mucilage extrusion defect, observed in seeds arising from the *sep3 shp1 shp2* triple mutant, relative to the respective single mutants and double mutant *shp1 shp2* (Fig. 3.9K, B-D & E). An in-depth examination of the possibility that MADS-box TFs act cooperatively will be covered in **Chapter 5**.

3.2.5 Identification of seed coat-specific promoters

As *ag* mutants lack inner whorl floral organs that produce the gametes, it is not possible to assess AG activity in the seed coat. To circumvent the need for AG during flower development, an approach was taken to knock-down AG activity specifically in the MSCs using an artificial microRNA (amiRNA) under the control of a seed coat-specific promoter. Before this approach could be undertaken, a suitable seed coat-specific promoter needed to be identified. In order to test the effectiveness of the selected promoters, their capacity to knock down *LUH* and *MUM2* activity using amiRNA targeting these genes was assessed.

DP1 (*AT4G11180*) was chosen based on published data indicating that this promoter has seed coat specificity (Esfandiari et al., 2013), whereas *GAMMA-INTERFERON-RESPONSIVE LYSOSOMAL THIOL REDUCTASE* (*AT4G12960*, *GILT*) was the top seed coat-specific candidate identified by interrogating the Belmonte dataset across all seed developmental stages (Belmonte et al., 2013; Esfandiari et al., 2013). The identification of *GILT* as a suitable seed coat-specific promoter involved several steps: Firstly, all genes with any embryo expression (denoted as present; PP), and all genes with no seed coat expression (denoted as absent; AA) were removed from the Belmonte dataset, with the remaining genes arranged from highest to lowest

expression. This generated a list of seed coat-specific genes that were next compared against the RNA-seq analysis of *luh-4* whole seeds at 7 DPA to identify seed coat-specific genes responding to the loss of LUH. These genes were removed from further consideration as the promoters of these genes would likely be influenced by an amiRNA targeting LUH. As the Belmonte dataset only covers expression of genes limited to seeds, data from the eFP browser was examined and genes with expression in other plant structures excluded (Winter et al., 2007). This step was necessary to prevent the selection of promoters that are active in other parts of the plant which may cause broad plant developmental defects when used to drive amiRNAs. These combined analyses yielded a small list of candidate seed coat-specific genes, with *GILT* as the best candidate due to its strong expression pattern across all stages of the seed (Fig. 3.10).

3.2.5.1 Characterisation of seed coat-specific promoter activity

As mentioned above, a criterion for using a seed coat-specific promoter in a *luh* background is that the promoter must not be mis-regulated in *luh* mutants and therefore maintain similar expression compared to Col-0 across all seed developmental stages. Thus, the next step was to validate the expression of the two seed coat-specific promoters, *DP1_{pro}* and *GILT_{pro}*. This involved performing qRT-PCR on *luh-4* seed coat material at the linear cotyledon stage (7 DPA) to examine the gene activity as a proxy for promoter activity. The qRT-PCR analysis was performed as described in **Chapter 2, Section 2.3**, and a list of qRT-PCR primers is shown in Appendix 3, Supplementary Table 11. Seed coat-specific promoter *AT1G_{pro}* (*AT1G62000*; *TESTA ABUNDANT 1*) (Tsai et al., 2017) was also included in this study as it had only been previously partially characterised and shown to be active in the later stages (from bending cotyledon onwards) of seed development (Jayawardana, 2016).

RNA-seq data did not detect *AT1G* or *GILT* as being differentially expressed in *luh-4* seeds at 7 DPA but revealed a 5.1-fold decrease in *DP1* (Appendix 3, Supplementary Table 2). Consistent with this, a 27.4-fold decrease in *DP1* expression in *luh-4* seed coat material was detected by qRT-PCR (Fig. 3.11 & Appendix 3, Supplementary Table 9). Next, *AT1G_{pro}*, *DP1_{pro}* and *GILT_{pro}-VENUSn* reporters were generated and introgressed into both Col-0 and *luh-5* backgrounds by floral dip transformation and then crossing as described in **Section 3.2.3** to enable direct comparisons of reporter transgenes between Col-0 and mutant backgrounds. Primers used for making these constructs are listed in Appendix 3, Supplementary Table 12, and the description of constructs used in the making of these reporters is given in Appendix 3, Supplementary Table 13, with the exception of *AT1G_{pro}::VENUSn* which was kindly provided by Dr. Nadeeka Jayawardana. These reporter lines served dual functions – to validate activity of the *AT1G*, *DP1* and *GILT* promoters in both Col-0 and *luh* backgrounds at early to late developmental stages in the seed coat as well as to evaluate their differential expression in *luh* mutants.

In terms of promoter activity, *GILT_{pro}* is the only promoter that is highly active in all stages of seed coat development, while *DP1_{pro}* and *AT1G_{pro}* are both active from the linear cotyledon stages onwards (Fig. 3.12). The pattern of *AT1G_{pro}*, *DP1_{pro}* and *GILT_{pro}* activity closely followed the expression pattern in wildtype predicted from the Belmonte dataset (Fig. 3.10) and observed in the RNA-seq on Col-0 (data not shown), while a decrease of fluorescence was only observed in *DP1_{pro}::VENUSn* reporters in *luh-5* from the linear cotyledon stage onwards (Fig. 3.12 & 3.13). Furthermore, expression of all three promoters driving VENUSn were absent in embryos (Appendix 3, Supplementary Fig. 4), again confirming the seed coat-specific expression of these promoters that was inferred from the Belmonte dataset. While the pattern of fluorescence suggested that *AT1G_{pro}*, *DP1_{pro}* and *GILT_{pro}* are only expressed in the MSCs in developing Col-0 and *luh-5* seeds, compound microscopy does not provide sufficient resolution to exclude expression in deeper layers of the seed coat (Fig. 3.12 & 3.13). Interestingly, activity of *GILT_{pro}* was detected at low levels in other plant structures such as the carpel, stamen, trichome tips and ovules

(Fig. 3.14) (Winter et al., 2007). Consequently, *GILT_{pro}*, along with *DP1_{pro}* and *AT1G_{pro}* were further assessed for utility in driving amiRNAs as discussed in the next section.

3.2.5.2 Utility of seed coat-specific promoters in driving artificial micro-RNAs

Seed coat-specific promoters *DP1_{pro}*, *GILT_{pro}* and *AT1G_{pro}* were next used to drive amiRNAs targeting either *LUH* or *MUM2*, to evaluate their capacity to knock down gene activity during seed development. Based on characterisation in **Section 3.2.5.1**, only *GILT_{pro}* has the potential to be effective in all developmental stages, while *DP1_{pro}* is predicted to be ineffective at early developmental stages but may be effective at later stages. Results of previously characterised *AT1G_{pro}* amiRNA lines showed that this promoter was only effective at knocking down *MUM2* but not *LUH* (Jayawardana, 2016) (Table 3.6).

amiRNAs targeting *LUH* (*amiR-LUH*) and *MUM2* (*amiR-MUM2*) were previously generated and shown to be active, that is, able to cause a mucilage extrusion defect when placed under the control of the Cauliflower mosaic virus promoter (CaMV 35S_{pro}) with 80.8% (n=26) and 88.8% (n=27) of lines displaying moderate or severe mucilage phenotypes (Jayawardana, 2016; Stahle et al., 2009) (Table 3.6). Both *amiR-LUH* and *amiR-MUM2* were obtained from laboratory stocks and placed downstream of *DP1_{pro}* in shuttle vector pMIGRO (JG1330), and downstream of *GILT_{pro}* in binary vector pGREENII (JG1595) via seamless cloning. Next, *DP1_{pro}* driving *amiR-LUH* and *amiRMUM2* in pMIGRO were moved into the *NotI* site of the binary vector pMLBART using standard restriction-ligation approach. All constructs were subsequently introduced into Col-0 plants via *Agrobacterium* C58. A list of all primers, constructs and plasmids used can be found in Appendix 3, Supplementary Table 14 & 15. Primers, constructs and plasmids associated with *AT1G_{pro}* were provided by Dr Nadeeka Jayawardana. Resultant transgenic plants were selected by Basta selection and assayed for mucilage defects via Ruthenium Red staining.

As expected, only *GILT_{pro}* proved sufficiently active to induce mucilage extrusion defects in Col-0 when targeting knock down of either *LUH* or *MUM2* with 67.5% (n=40) and 75% (n=40) of T₁ lines displaying mucilage extrusion defects respectively, compared to 80.8% (n=26) and 88.8% (n=27) respectively when the *35S_{pro}* was used (Table 3.6). While *DPI_{pro}* and *AT1G_{pro}* both failed to induce mucilage defects when driving *amiR-LUH* (Table 3.6 & **Section 3.2.5.1**), *AT1G_{pro}* produced a high rate (83.9%; n=62) of mucilage defects when driving *amiR-MUM2* (Table 3.6).

In summary, *GILT_{pro}* is the only seed coat-specific promoter capable of reducing gene activity in all seed coat stages when used to drive an amiRNA, whereas *AT1G_{pro}* is only effective at later stages of seed coat development. For this reason, *GILT_{pro}* was selected to drive an amiRNA targeting *AG*, which according to the eFP browser is active in the early stages of seed coat development (Winter et al., 2007).

3.2.6 Analysis of *AG* knockdown in seed coat

amiR-AG in the *miR-164* backbone (Alvarez et al., 2006), was cloned into binary vector pGREENII downstream of the *35S_{pro}* (JG1188) and subsequently introduced into Col-0 plants to evaluate the effectiveness of the amiRNA to induce floral defects. Selection of transgenic plants were performed using Basta, and T₁ plants were examined for floral defects associated with reduced *AG* activity. This analysis revealed varying degrees of floral defects with 19/40 lines having strong *ag* phenotypes (47.5%) with a deformed flower replacing the carpel and stamens, 10 lines with mild *ag* phenotypes (25%) with short, deformed carpels and homeotic transformation of stamens to petals, and 11 lines (27.5%) with no *ag* phenotype (Table 3.7). Collectively these phenotypes are consistent with reduced *AG* activity. Therefore, with a penetrance of 72.5% (n=40), *amiR-AG* was considered to be effective in knocking down *AG*.

Having established the effectiveness of *amiR-AG*, the amiRNA was placed downstream of *GILT_{pro}* in binary vector pGREENII (JG1595) by seamless cloning to generate construct *GILT_{pro}::amiR-AG* (JG1599). This was subsequently moved into *Agrobacterium* C58 by electroporation (JG1605) and used to transform Col-0 plants via floral dip. Primers and plasmids associated with construction of *GILT_{pro}::amiR-AG* can be found in Appendix 3, Supplementary Table 14 & 15. Resulting Basta-resistant T₁ seeds were collected on maturation, and about 20 seeds per line were assayed for mucilage defects using Ruthenium Red staining. Interestingly, unlike other MADS-box TF mutants that produced mucilage defects (Fig. 3.8, 3.9 & Supplementary Fig. 3), lines expressing *amiR-AG* under the control of *GILT_{pro}* had normal mucilage release (Table 3.8). Twelve of 40 lines (30%) had noticeably larger seed size relative to wildtype control Col-0 seeds and were weighed alongside Col-0 to determine if increase in seed size was correlated with increase in seed weight. Weights of 300 seeds from each line (weighed as three independent replicates of 100 in three tubes to serve as triplicates) were determined (Fig. 3.15 & Appendix 3, Supplementary Table 16). Increased seed weight was calculated as a fold change compared to Col-0, with lowest to highest fold change being 1.32- to 1.89-fold (Appendix 3, Supplementary Table 16).

Next, all 12 lines were sowed out for segregation analysis, with three T₂ lines analysed per line. Only single T-DNA insert lines carried forward to the next generation to determine if the increased seed size/weight phenotype was heritable, as single inserts allows for easier comparison and are also more stably inherited in subsequent generations. A total of six lines were found to have single inserts, and the resulting T₂ lines were analysed in the same manner as described for T₁ seeds (Appendix 3, Supplementary Table 16). All six T₂ lines were found to have significantly increased seed weight relative to Col-0 ($p < 0.01$) (Fig. 3.15). Given that T₂ lines maintained increased seed weight, the seed size phenotype appears to be stably inherited across generations.

3.3 Discussion

3.3.1 Experimental discrepancies between *luh-4* and *luh-5*

While ANOSIM determined that the RNA-seq data obtained from *luh-4* and *luh-5* mutants at 7 DPA were not significantly different from each other ($p=0.1$), several discrepancies were noted between the two alleles. For example, a higher amount of DEGs were detected in *luh-5* compared to *luh-4* (4371 vs 2988) (Fig. 3.2). While ANOSIM ruled out a statistical difference between the two alleles, the difference in the number of biological replicates between *luh-4* and *luh-5* mutants may influence the amount of DEGs detected in each background. For example, in the set for *luh-4*, four biological replicates were sampled, while for the set with *luh-5*, three biological replicates were sampled. Thus, the higher number of biological replicates in *luh-4* may account for less DEGs detected due to higher stringency (i.e. Four replicates in *luh-4* likely introduces more variation than three replicates in *luh-5*, leading to a higher elimination of DEGs on this basis). Interestingly, although *luh-4* (124.6-fold) produced a greater decrease in *LUH* than *luh-5* (2.2-fold), a similar decrease in *MUM2* was seen in *luh-5* (10.2-fold) and *luh-4* (8.8-fold) confirming that both alleles are strong (Appendix 3, Supplementary Table 2 & 3). There are two factors that may account for the drastic difference in *LUH* transcripts between *luh-4* and *luh-5*. The mechanism of non-sense mediated decay, which facilitates the removal of non-sense transcripts bound by ribosomes (Nickless et al., 2017), may account for the reduction of *luh-5* transcripts resulting in a 2-fold difference to wildtype. However, this explanation seems unlikely as non-sense mediated decay usually results in larger fold change (Nickless et al., 2017). A second possible explanation is that *LUH* is autoregulatory and required to promote its own expression and thus in the absence of functional *LUH*, *luh-5* expression is reduced. Given that both *luh-4* and *luh-5* alleles condition a similar mucilage extrusion phenotype and decrease in *MUM2* expression, it seems reasonable to compare DEGs arising in both mutant backgrounds. Despite differences in the total number of DEGs identified, ANOSIM analysis suggests these are not statistically significant. DEGs shared between these mutants therefore represent a core set of genes regulated either directly or indirectly by *LUH*.

3.3.2 The role of LUH in the developing seed coat

LUH is a Groucho/Tup1-like transcriptional corepressor that is expressed throughout the developing seed (Walker et al., 2011). As the main notable defect of *luh* mutants is the lack of mucilage extrusion that is associated with the loss of *MUM2*, LUH appears to be a key *MUM2* activator (Dean et al., 2007; Macquet et al., 2007b; Walker et al., 2011). Given that LUG inhibits AG, a MADS-box TF in floral development (Sridhar et al., 2004), and that several MADS-box TFs were elevated in *luh* seeds, it appeared that LUG and LUH have conserved functions in repression, with LUH acting as a repressor temporally but not spatially in the seed coat. Therefore, a de-repression model (Fig. 1.9C) was proposed for *MUM2* regulation. Briefly, this involved LUH lifting repression mediated by a putative *MUM2* repressor from 7 DPA onwards to allow for proper *MUM2* expression. Hence, a prediction arising from this model is the regulation of genes from 7 DPA onwards. While RNA-seq on both *luh-4* and *luh-5* at 7 DPA (mid seed development) reveal that absence of *LUH* results in a large proportion of genes (2988 and 4371 genes respectively) being misexpressed (Fig. 3.2A), with majority of these genes (1140/1519) expressed throughout the seed coat (Fig. 3.2B), DEGs were also detected at 3 DPA, indicating this prediction was incorrect. This however does not necessarily invalidate the de-repression model, as LUH may have other roles in the seed coat beyond *MUM2* regulation. Therefore, the RNA-seq datasets, particularly the time-course on *luh-5*, can be used to discuss the possible roles of LUH within the developing seed coat in the absence of generic redundancy arising from LUG. Broadly, this involves comparisons of the number of DEGs obtained across the different developmental stages as an approximate measure of where LUH might have the most regulatory impact. The RNA-seq time course on *luh-5* reveals that *LUH* appears particularly active at 3 DPA (4676 DEGs), followed closely by 7 DPA (4371 DEGs) and is decreased at 10 DPA (2793 DEGs) (Fig. 3.3). This mis-expression pattern in *luh-5* is coherent with the general expression of *LUH* according to the eFP browser, which displays peak *LUH* expression in the general seed coat at the globular stage (2 DPA) that gradually decreases across the later developmental stages (Winter et al., 2007). Together, these data point to the necessity of *LUH* for the correct expression of genes in wildtype seeds and therefore may act as

a regulator in the developing seed coat, particularly in the early (3 DPA) to mid (7 DPA) developmental stages. DEGs appeared to be elevated (645) and decreased (828) at similar levels in *luh* mutants at 7 DPA (Fig. 3.2A), suggesting that LUH may possibly act as both a transcriptional activator and repressor in the seed coat. However, as this study focused on elevated TFs (Table 3.1 & 3.4) to identify the proposed putative *MUM2* repressor described in Fig. 1.9C, this discussion will primarily focus on elevated TFs.

Evidence that may support *LUH* as a regulator in the seed coat beyond the role of *MUM2* regulation is the increase of expression in a wide range of early acting tier 3 TF families in *luh* mutants at 7 DPA (Table 3.1), beyond a stage that they are typically expressed in Col-0 (Table 3.2): MYB, WRKY, AP2-like/Ethylene Responsive Factor (ERF), bHLH, homeodomain, bZIP and NAC-domain. These tier 3 TFs have diverse roles in the developing seed coat such as epidermal patterning (*MYB5*, *TT2*), specifying seed morphology (*NARS2*, *AP2*), proanthocyanin development (*TT2*, *TT8*, *TTG2*), germination (*bZIP44*) and mucilage polysaccharides regulation (*GL2*, *MYB5*) (Baudry et al., 2004; Gonzalez et al., 2016; Iglesias-Fernandez et al., 2013; Kunieda et al., 2008; Li et al., 2009; Ohto et al., 2009) (**Chapter 1, Section 1.2.1**). Given that disruptions to these tier 3 TFs occur generally throughout early (3 DPA) to late (10 DPA) (Appendix 3, Supplementary Table 6 & 7) in *luh* mutants, LUH appears to be limiting their levels and is involved in temporal regulation in the seed. While disruptions to these tier 3 TFs at both early to late stages would be expected to produce morphological defects, *luh* mutants have relatively normal MSC morphology (Huang et al., 2011; Walker et al., 2011) and are similar to tier 1 mutants (**Chapter 1, Section 1.2.3**) (Golz et al., 2018). It is possible that these tier 3 TFs while repressed by LUH, are still active enough to ensure normal seed development void of morphological defects. Therefore, it is unlikely that LUH is repressing these tier 3 TFs in the seed coat for the purpose of regulating downstream mucilage polysaccharides regulation, such as the proper expression of *MUM2*. Instead, as LUH is known to regulate abiotic stress responses in *Arabidopsis* plants (Shrestha et al., 2014), the role of LUH in the seed coat may be simply to respond to

environmental cues in the seed coat tissue, perhaps in coordination with seed coat development. For instance, deviations in temperature has been observed to affect overall seed (both embryo and seed coat) development, with an increase in temperature resulting in faster seed development; i.e. linear cotyledon stage is reached at 5 DPA instead of 7 DPA (unpublished data). Conversely, a decrease in temperature results in slower overall seed development; i.e. linear cotyledon stage reached at 9 instead of 7 DPA (unpublished data). As embryo and seed coat development are closely associated (Haughn & Chaudhury, 2005), it is possible that LUG and LUH may have roles in coordinating abiotic responses in the embryo, while LUH may coordinate abiotic responses in the seed coat. These scenarios require further testing but are not in the scope of this thesis. Lastly, in addition to repressing tier 3 TFs, LUH also represses MADS-box TFs (Table 3.1), which are classified as tier 1 TFs (Golz et al., 2018) (**Chapter 1, Section 1.2.3**). Hence, an obvious model for the activation of *MUM2* by LUH lies in a regulatory pathway mediated by MADS-box TFs and is discussed in the next section.

Future work may include performing RNA-seq on just the seed coat, as RNA-seq was performed on whole seeds. As dissecting the seed coat is labour intensive and likely induces stress responses in the tissue whilst being susceptible to endosperm contamination, approaches such as translating ribosome affinity purification sequencing (TRAP-seq) may facilitate the transcriptomic analysis of cells within the seed coat (Reynosio et al., 2015). Briefly, this might involve using a seed coat-specific promoter such as *GILT_{pro}* (**Chapter 3, Section 3.2.5**) to express a tagged ribosomal protein and precipitating the translating mRNA, collectively known as polysomes, and sequencing the isolated mRNA through RNA-seq. This will allow for clean, seed coat-specific transcriptomic data which may augment the understanding of LUH role in the seed coat.

3.3.3 The roles of MADS-box TFs in the seed coat

Interrogation of RNA-seq datasets (Table 3.1 & 3.4) to identify elevated TFs in *luh* mutants as candidate *MUM2* repressors led to the identification of several MADS-box TFs (**Section 3.2.4**), suggesting a then novel role of MADS-box TFs in mucilage polysaccharides regulation. While this role was subsequently confirmed in published literature that demonstrated both STK (Ezquer et al., 2016) and SHP1 and SHP2 (Ehlers et al., 2016) have roles in mucilage polysaccharides production, this data was not available at the time of the study. A key characteristic of a putative *MUM2* repressor(s) is the predicted lack of mucilage defects associated with mutations in these genes, but a rescue of mucilage extrusion in double mutants lacking both *LUH* and candidate *MUM2* repressor(s). The latter arises because in double mutants, loss of *MUM2* repressor activity is epistatic to mutations causing an increase in *MUM2* repressor activity. Therefore, the mucilage extrusion defects observed in MADS-box TF mutants *sep3*, *shp2*, *stk*, *shp1 shp2* and *sep1 sep3* (Fig. 3.8 & 3.9) in this study are inconsistent with MADS-box TFs being *MUM2* repressors. Instead, these observations are consistent with MADS-box TFs functioning as *MUM2* activators, or as regulators of other downstream mucilage polysaccharides effector(s). Both scenarios have been investigated using transcriptomic approaches on selected MADS-box TF mutants and this work is described in **Chapter 4**. In both instances, the observation of an enhanced mucilage defect seen in *sep3 shp1 shp2* triple mutants and *shp1 shp2* double mutants compared to single *shp2* and *sep3* mutants (Fig. 3.8) means that MADS-box TFs SEP3, SHP1 and SHP2 must be acting in a partially redundant manner to regulate mucilage polysaccharides and is consistent with MADS-box TFs forming quartets for gene regulation in *Arabidopsis* (Immink et al., 2009; Theissen & Saedler, 2001).

Based on the premise that MADS-box TFs, particularly SEP3, SHP1 and SHP2 are actually *MUM2* activators, they may act in a manner that is either independent or dependent on *LUH*. There are two possible scenarios where SEP3, SHP1 and SHP2 are dependent on *LUH* for *MUM2* activation – 1) directly in a complex with *LUH*

(Fig. 1.9A) or 2) indirectly as an intermediate *MUM2* activator activated by LUH (Fig. 1.9B). Given that these MADS-box TFs are repressed, and not activated by LUH at the linear cotyledon stage (7 DPA) onwards (Table 3.1, 3.4 & 3.5), a stage coinciding with the onset of *MUM2* expression in the MSCs (Fig. 3.1A) (Jayawardana, 2016), the indirect activation model described in Fig. 1.9B does not apply to the MADS-box TFs. In any case, the repression of MADS-box TFs (a potential *MUM2* activator) by LUH is odd, given that LUH activates *MUM2*. A possible reason is that LUH regulates these MADS-box TFs for other roles in the seed coat, such as the repression of *AG* to maintain appropriate seed size as discussed below. Eliminating the indirect activation model as a mechanism for *MUM2* regulation (Fig. 1.9B) leaves two possible models for the role of SEP3, SHP1 and SHP2 in *MUM2* regulation. The first is where the MADS-box TFs activate *MUM2* independent of LUH, while the second model has them working in a complex with LUH to activate *MUM2* (Fig. 1.9A). To investigate if SEP3, SHP1 and SHP2 may be acting in a complex with LUH, yeast-based assays such as yeast 2-hybrid (Y2H) and Y3H can be used to test interactions between LUH, SEU, SEP3, SHP1 and SHP2 and is described in **Chapter 5**.

Besides a role in mucilage polysaccharides regulation, this study identified roles of MADS-box TFs in determining seed shape (SEP1, SEP3) and a novel role in determining seed size (*AG*) in *Arabidopsis*. The deformed seed shape in *sep1 sep3* double mutants (Fig. 3.9H) resemble *stk* (Favaro et al., 2003) and *aberrant testa shape* (*ats*) mutants (Leon-Kloosterziel et al., 1994). STK is required for the normal development of the funiculus, with disruptions to this tissue driving the seed shape defect seen in *stk* mutants (Favaro et al., 2003), whereas *ats* mutants have disruptions to inner and outer integuments during ovule development (Leon-Kloosterziel et al., 1994). Interestingly, STK is also implicated in the development of ovules with ARABIDOPSIS B-SISTER (ABS), as *stk abs* mutants lack the entire inner integument layer (Ehlers et al., 2016). The similarity in mutant seed shape in *sep1 sep3* double mutants to both *stk* and *ats* implies that MADS-box TFs SEP1 and SEP3 may be involved in both funiculus development and early integument development.

Furthermore, STK interacts with SHP1 and SHP2 to produce mutants with enhanced ovule defects (Favaro et al., 2003), while ATS interacts genetically with epidermal-acting TFs such as GL2 and AP2, with double mutants *ats gl2* and *ats ap2* producing enhanced seed shape deformation (Leon-Kloosterziel et al., 1994). Therefore, future work should consider possible genetic interaction between SEP1/SEP3 with ABS, SHP1, SHP2, ATS, GL2 and AP2 through generation of several mutant combinations, and evaluating if an enhanced deformed seed shape is observed. Observation of enhanced defects will indicate that these genes act in separate pathways, while a lack of enhancement will suggest that these genes lie in the same genetic pathway. Next, further characterisation of *sep1 sep3* mutants is also necessary, perhaps by histological sections across seed developmental stages to pinpoint where these disruptions occur and identify the affected seed coat tissues, as well as assess if embryo development is affected. Additionally, while *sep1 sep3* double mutants require examination under scanning electron microscopy (SEM) to confirm if formation of their MSCs are disrupted, this scenario is likely, given the poor Ruthenium Red staining observed (Fig. 3.9H). Given that *sep1* mutants are aphenotypic (Appendix 3, Supplementary Fig. 3C), and *sep3* mutants have mild seed deformation (best seen in Fig. 3.8O), the synergy seen between the mutants is consistent with SEP1 and SEP3 acting in a partially redundant manner in regulating seed shape, with SEP3 as the pre-dominantly active form. Meanwhile, MADS-box TF AG is implicated in the control of seed size, as knock-downs of AG directed by artificial microRNAs resulted in a significant increase in seed size and weight, but with wildtype MSCs formation inferred from Ruthenium Red staining, which showed mucilage extrusion (Table 3.8 & Fig. 3.15). This will be further discussed in **Section 3.3.5**.

3.3.4 The utility of *GILT_{pro}* as a seed coat-specific tool

Having initially selected MADS-box TFs for further study as putative *MUM2* repressors (**Section 3.2.4**), a seed coat-specific promoter was needed as a tool to

specifically knock-down *AG* in the seed coat to circumvent sterility associated with *ag* mutants (Liu & Meyerowitz, 1995). Ideally, such a seed coat promoter would be capable of knocking down genes expressed throughout the seed coat developmental stages to allow for the characterisation of genes that function at particular time points. While *GILT_{pro}* was finally identified and proved useful in revealing the role of *AG* in the seed coat, there are limitations associated with the use of *GILT_{pro}* as a seed coat-specific tool. For instance, while *GILT_{pro}* is able to knock-down both early-acting (*LUH*) and late-acting (*MUM2*) genes with reasonable success in a Col-0 background (Table 3.6), it was only able to drive complementation of *mum2*, but not *luh* when used to express *MUM2* and *LUH* respectively (data not shown). This indicates that while *GILT_{pro}* is highly active throughout all seed coat developmental stages (Fig. 3.12), it is more active in the later stages. This is an important caveat that potential users of *GILT_{pro}* as a seed coat-specific tool for complementation assays must consider. While *GILT_{pro}* is active in other plant structures, notably the stamens and ovules (Fig. 3.14), no disruptions were observed to these structures in this study when used to drive amiRNAs or heterologous expression. However, if *GILT_{pro}* is to be used to affect the activity of genes in the seed coat that also have roles in stamen/ovule development, then it is possible that this approach will lead to stamen/ovule defects.

While *GILT_{pro}* has two major limitations – the inability to complement the activity of early-acting genes (e.g. *LUH*) and its activity in other tissue beyond the seed coat, there was no other alternative seed coat promoter available that is present across all seed developmental stages (Fig. 3.10). There are therefore two main proposals to address this issue: 1) the addition of enhancer sequences to increase the transcriptional strength of *GILT_{pro}* and 2) the generation of seed coat-specific synthetic promoters. The first proposal involves the use of enhancers, which are short regions of *cis*-acting regulatory elements that can be found in the vicinity of target genes (upstream/within/downstream) and confer spatiotemporal patterns of gene activity by recruiting cell-specific TFs to enhance its expression (Blackwood & Kadonaga, 1998; Pennacchio et al., 2013). Therefore, enhancers, such as the 68 bp viral enhancer sequence from tobacco mosaic virus (TMV) that is capable of enhancing expression

of target genes by 2- to 10-fold (Turner & Foster, 1995), can be added in the vicinity of *GILT_{pro}* to ensure its increased activity. While the first proposal may address the ability of *GILT_{pro}* in driving early-acting genes in the seed coat, it risks the increase activity in other tissues of the plant. A worst-case scenario is the increased activity in these tissues if *GILT_{pro}* is used to drive CDS of genes also expressed in these regions, leading to developmental defects in structures such as the stamen and ovules, culminating in sterile mutants. Thus, the first proposal should only be attempted if *GILT_{pro}* is too weak for its intended use. The second proposal involves the use of synthetic promoters in lieu of *GILT_{pro}* as a seed coat-specific tool. This proposal requires the re-organisation of the ‘cis-architecture’ of a DNA promoter by a combination of adjusting the position of the TATA box, possibly including desirable elements from another promoter and site-directed mutagenesis to increase the efficiency of native promoters (Dey et al., 2015). While this has the potential of resolving both limitations reported in *GILT_{pro}*, creating and testing the utility of synthetic seed coat promoters is time consuming and may not be warranted given the reported utility of *GILT_{pro}* in this study.

Overall, the identification and validation of *GILT_{pro}* as a seed coat-specific tool is useful for basic and applied research in seed coat biology. However, it is important to verify that *GILT_{pro}* (and any tissue-specific promoter) is stably expressed in mutant backgrounds intended for use, as differential expression of such tools in a selected mutant background would negate its utility.

3.3.5 AG has roles in controlling seed size in *Arabidopsis*

AG is a MADS-box TF that is well characterised in specifying carpel identity according to the ABC model (Coen & Meyerowitz, 1991). As the loss of AG results in sterility and the resulting lack of seeds, the role of AG in controlling seed size in *Arabidopsis* has been masked until this study. The increase in seed mass (weight and size, Table 3.8 & Fig. 3.15) on knock-downs of AG directed by seed coat-specific

GILT_{pro} in Col-0 therefore shows that AG is necessary for specifying seed size. As the knock-down of AG took place in the maternally-derived seed coat, AG is likely acting maternally and not zygotically in regulating seed size. A survey of published literature identifying maternal acting genes controlling seed size include genes involved in ubiquitin pathways (*DA1*, *DA2*) (Li et al., 2008; Xia et al., 2013), hormone signalling (*AUXIN RESPONSE FACTOR2*; *ARF2*) (Okushima et al., 2005) and other functions (*AINTEGUMENTA*; *ANT*, *AP2* and *TTG2*) (Garcia et al., 2005; Jofuku et al., 2005; Mizukami & Fischer, 2000). Both *da1* and *da2* mutants result in *Arabidopsis* plants with enlarged leaves and seeds, which are enhanced in *da1 da2* mutants (Li et al., 2008), indicating that DA1 and DA2 act in the same genetic pathway or in the same complex (Xia et al., 2013). *DA1* encodes a ubiquitin receptor while *DA2* encodes a E3 ubiquitin ligase, both of which are observed to restrict cell proliferation in the integuments of ovules, possibly by targeting positive regulators of cell proliferation to the proteasome (Li et al., 2008; Xia et al., 2013). Similar to DA1 and DA2, ARF2 is a negative regulator of cell proliferation in the ovule integuments, with *arf2* mutants producing enlarged plant size, seed size and abnormal floral morphology resulting in reduced fertility (Okushima et al., 2005). ARF2 likely inhibits *ANT*, a gene promoting cell division as expression of *ANT* is prolonged in *arf2* mutants (Schruff et al., 2006). Furthermore, overexpression of *ANT* is associated with an increase in seed size, while *ant* mutants produce seeds with reduced sizes (Mizukami & Fischer, 2000). Meanwhile, epidermal acting TFs AP2 and TTG2 have opposing roles in regulating cell elongation in integuments for eventual seed size: AP2 has roles in limiting cell elongation in integuments, with *ap2* mutants displaying larger seeds than wildtype (Jofuku et al., 2005). In contrast, *ttg2* mutant seeds are smaller and are associated with a loss of cell wall elongation in the integuments (Garcia et al., 2005). A simple way to determine if AG is involved in regulating any of these genes is to perform qRT-PCR on AG known-down mutants and to look for differential expression of these genes. Exploring genetic interactions can be achieved by crossing mutants with enlarged seeds as *da1*, *da2*, *arf2* and *ap2* and evaluating for enhanced seed size. Therefore, future directions will include the characterisation of AG knock-down mutants and the study of genetic interactions between AG knock-down and the aforementioned genes.

Such characterisations include determining if there is a synergistic increase in seed size and determining whether this is due to increases in integument cell size and/or embryo size and if enlarged seeds produce higher oil content for potential use in agronomy.

Lastly, enlarged seeds represent one way of increasing crop yield and may have applications in agronomy such as the economically viable canola. Several challenges are associated in applying the AG knock-down to canola, which include the successful knock-down of AG in canola, possibly driven by *GILT_{pro}* if *ag* mutants of canola are sterile as predicted. Assuming the success in removing/decreasing AG expression in canola, verification of an enhanced seed size is also necessary, and if this results in higher oil yield in canola seeds.

3.6 Conclusion

Based on the de-repression model, MADS-box TFs were initially identified as *MUM2* repressors in RNA-seq datasets looking for TFs elevated in *luh* mutants. However, the mucilage extrusion defects observed in seeds of several MADS-box TF mutants, most notably *sep3*, *shp2* and *shp1 shp2* casts doubt as to whether these TFs are indeed *MUM2* repressors. Therefore, further characterisation of these mutants is warranted to determine if the observed mucilage extrusion defects result from a disruption in *MUM2* expression and/or by misregulation of other causative factors. Besides the role of MADS-box TFs in mucilage polysaccharides regulation, this study identified novel roles of MADS-box TFs in the seed such as specifying seed shape (*SEP1*, *SEP3*) and seed size (*AG*) in *Arabidopsis*. Thus, it is possible that *LUH* represses MADS-box TFs such as *AG* to maintain proper seed size in roles beyond mucilage polysaccharides regulation. While *ag* mutants are sterile, the identification and evaluation of a robust seed coat-specific promoter, *GILT_{pro}*, allowed for the knock-down of *AG* in Col-0 specifically in the seed coat and is thus demonstrated as a useful tool for both basic and applied seed biology research, particularly since *AtGILT_{pro}* has

been demonstrated as possessing seed coat outer integument-specific promoter in canola (Wu et al., 2011). These AG knock-downs produced an enlarged seed phenotype and holds potential agronomy value, especially if this can be applied to economically viable crops such as canola.

Future directions beyond the scope of this thesis includes the characterisation of *sep1 sep3* double mutants and AG knock-downs to determine how SEP1 and SEP3 control seed shape and how AG controls seed size respectively. Broadly, this may involve the use of histological sections on both backgrounds to determine where disruptions to normal seed development occurs, the examination of differentially expressed genes in both backgrounds to determine what genes these TFs may regulate and the examination of genetic interactions through crossing. Specifically, *sep1 sep3* can be crossed into mutants with abnormal seed shape to check for genetic interactions while AG knock-downs may be crossed to other enlarged seed phenotypes to identify possible interactors.

Table 3.1. List of TFs elevated in *luh* seeds at the linear cotyledon stage of development.

Of the 56 TFs elevated at the linear cotyledon stage of seed development in *luh-4*, 35 are also elevated in *luh-5* mutants. FPKM values $\pm$ SEM are given in parenthesis alongside the fold increase. Expression profiles of these TFs in the general seed coat was determined by mining publicly-available microarray datasets (Belmonte et al., 2013). Of the 20 TFs that are denoted as predominantly expressed in, but not restricted to the seed coat (purple highlighting) by the Belmonte dataset, 15 are also elevated in both *luh-4* and *luh-5* mutants. The five TFs that are solely elevated in *luh-4* belong to TF families in which other members are either involved or potentially involved in MSC development and/or mucilage polysaccharides production (red text). ND - gene not detected as being differentially expressed in that background. See Appendix 3, Supplementary Table 2 & 3 for the full list of DEGs and TFs.

Locus	Gene	Fold elevated (FPKM values)	
		<i>luh-4</i>	<i>luh-5</i>
MADS-box (10)			
AT1G17310	<i>AGL100</i>	4.5 (1.9)	ND (0.7)
AT1G24260	<i>SEPALLATA3 (SEP3)</i>	1.9 (22.3)	ND (28.0)
AT2G14210	<i>AGAMOUS-LIKE44 (AGL44)</i>	1.9 (3.1)	2.4 (3.1)
AT2G42830	<i>SHATTERPROOF2 (SHP2)</i>	1.5 (94.2)	1.5 (147.1)
AT3G54340	<i>APETALA3 (AP3)</i>	3.3 (84.3)	2.4 (98.5)
AT3G58780	<i>SHATTERPROOF1 (SHP1)</i>	1.5 (20.7)	ND (25.6)
AT4G09960	<i>SEEDSTICK (STK)</i>	1.6 (26.8)	2.0 (38.5)
AT4G18960	<i>AGAMOUS (AG)</i>	2.1 (8.0)	ND (10.2)
AT5G13240	<i>MADS AFFECTING FLOWERING1 (MAF1)</i>	1.5 (16.1)	1.3 (17.8)
AT5G15800	<i>SEPALLATA1 (SEP1)</i>	1.7 (72.6)	1.5 (96.2)
AP2-like/ERF (4)			
AT1G12890	-	2.1 (2.6)	2.6 (6.0)
AT4G36920	<i>APETALA2 (AP2)</i>	1.5 (29.8)	1.4 (32.3)
AT5G18090	-	1.4 (4.9)	1.5 (2.5)
AT5G60140	-	1.6 (63.9)	1.4 (114.8)
MYB (7)			
AT1G22640	<i>MYB-PROTEIN3 (MYB3)</i>	1.9 (33.4)	1.9 (37.6)
AT2G25230	<i>MYB-PROTEIN100 (MYB100)</i>	5.2 (1.0)	N.D. (0.9)
AT2G46410	<i>CAPRICE (CPC)</i>	1.9 (12.0)	1.9 (16.6)
AT3G13540	<i>MYB-PROTEIN5 (MYB5)</i>	1.8 (61.0)	1.4 (81.7)
AT4G01680	<i>MYB-PROTEIN55 (MYB55)</i>	2.1 (3.4)	1.9 (4.1)
AT5G35550	<i>TRANSPARENT TESTA2 (TT2)</i>	1.8 (26.7)	1.5 (36.9)
AT5G40360	<i>MYB-PROTEIN115 (MYB115)</i>	1.8 (14.8)	ND (7.0)

Locus	Gene	Fold elevated (FPKM values)	
		luh-4	luh-5
WRKY (2)			
AT2G37260	TRANSPARENT TESTA GLABRA2 (TTG2)	1.9 (22.2)	1.7 (35.3)
AT2G46400	WRKY-PROTEIN46 (WRKY46)	5.7 (4.2)	6.6 (5.1)
NAC (7)			
AT1G52880	NAC FAMILY PROTEIN2 (NARS2)	1.6 (112.1)	ND (116.4)
AT3G04070	NAC DOMAIN CONTAINING PROTEIN47 (NAC047)	2.1 (4.0)	2.2 (4.0)
AT3G10500	NAC DOMAIN CONTAINING PROTEIN53 (NAC053)	1.4 (12.8)	1.6 (13.3)
AT4G01540	NAC DOMAIN CONTAINING PROTEIN68 (NAC068)	1.6 (5.4)	ND (6.3)
AT5G07680	NAC DOMAIN CONTAINING PROTEIN80 (NAC080)	2.0 (32.9)	ND (26.7)
AT5G13180	VND-INTERACTING2 (VNI2)	2.0 (58.4)	1.7 (46.7)
AT5G63790	NAC DOMAIN CONTAINING PROTEIN102 (NAC102)	1.6 (9.9)	ND (7.3)
bHLH (5)			
AT1G05805	-	1.5 (5.2)	1.5 (6.3)
AT2G43060	ILI1 BINDING bHLH1 (IBH1)	1.5 (628.3)	1.3 (890.4)
AT3G07340	-	3.0 (6.8)	2.4 (9.6)
AT4G09820	TESTA TRANSPARENT8 (TT8)	2.6 (109.8)	1.8 (103.0)
AT4G28790	-	1.8 (2.4)	1.6 (1.5)
C2H2 (5)			
AT1G26590	-	1.5 (6.5)	ND (6.3)
AT1G26610	-	2.5 (6.7)	ND (6.9)
AT3G49930	-	1.9 (14.5)	N.D. (23.4)
AT4G16610	-	1.5 (46.7)	N.D. (83.4)
AT5G04340	COLD INDUCED ZINC FINGER PROTEIN2 (CZF2)	2.6 (8.7)	2.9 (20.4)
Homeodomain (4)			
AT1G01520	ALTERED SEED GERMINATION4 (ASG4)	1.9 (1.1)	N.D. (0.8)
AT2G38090	-	1.4 (11.3)	1.7 (17.2)
AT3G01470	ARABIDOPSIS THALIANA HOMEODOMAIN1 (ATHB1)	2.6 (1.9)	3.4 (1.5)
AT5G01380	-	1.9 (2.1)	ND (2.0)
bZIP (2)			
AT1G75390	BASIC LEUCINE-ZIPPER44 (bZIP44)	1.4 (411.0)	1.8 (533.0)
AT2G36270	ABA-INSENSITIVE5 (ABI5)	1.4 (5.9)	1.5 (4.9)
Miscellaneous (10)			
AT1G07640	UAS-TAGGED ROOT PATTERNING3 (URP3)	1.7 (10.9)	ND (5.5)
AT1G30500	NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT A-7 (NF-YA7)	1.4 (13.9)	1.7 (16.8)
AT1G55650	-	1.8 (2.9)	ND (2.0)
AT2G35940	EMBRYO SAC DEVELOPMENT ARREST29 (EDA29)	1.9 (8.4)	ND (10.1)
AT3G17100	-	1.4 (23.2)	ND (35.3)
AT4G34000	ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING FACTOR3 (ABF3)	1.4 (7.7)	1.8 (6.8)
AT5G18000	VERDANDI (VDD)	2.8 (1.8)	4.8 (3.3)
AT5G39230	-	4.0 (2.6)	ND (2.2)
AT5G52510	SCARECROW-LIKE8 (SCL8)	1.5 (49.6)	1.5 (64.9)
AT5G59230	-	7.6 (5.4)	4.0 (3.4)

Table 3.2. Expression pattern of seed coat-expressed TFs that are elevated in *luh* mutants.

The expression of 20 LUH-regulated TFs elevated in *luh* seed coats (see Table 3.1) was further investigated using the eFP browser at the BAR website (<http://bar.utoronto.ca/>) (Belmonte et al., 2013; Winter et al., 2007). Seventeen TFs are primarily expressed in the general seed coat and three are expressed predominantly in the chalazal region of the seed coat (indicated with an asterisk). TFs that were excluded from further analysis are indicated in red (5/20) (see text for details). Stages of SC expression are indicated as follows: Pre Globular (Pre. Glob.; 1 DPA), Globular (Glob.; 2 DPA), Heart (3 DPA), Torpedo (Torp.; 5 DPA), Linear Cotyledon (Lin. Cot.; 7 DPA), Bent Cotyledon (Bent Cot.; 9 DPA) and Mature Green (Mat. Green; 10 DPA).

TF family	Gene	Stages expressed in		eFP				
		SC						
MADS-box (7)	<i>SEP3</i>	Glob. to Heart		Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>SHP2</i>	Glob. to Mat. Green		Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>AP3</i>	Glob. to Heart		Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>SHP1</i>	Glob. to Heart		Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green

<i>STK</i>		Glob. to Heart					
<i>AG</i>		Glob.					
<i>SEP1</i>		Glob. to Heart					
AP2-like/ERF (2)	<i>AT5G60140</i>	Glob. to Lin. Cot.					
<i>AP2</i>		Glob. to Mat.. Green					
MYB (3)	<i>MYB3</i>	Lin. Cot. to Mat. Green					
	<i>MYB5</i>	Glob. to Heart					
	<i>TT2</i>	Glob. to Heart					

WRKY (1)	<i>TTG2</i>	Glob.	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
NAC-domain (3)	<i>NARS2</i>	Heart to Mat. Green	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>NAC080</i>	Mat. Green	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>*VNI2</i>	Glob. to Mat. Green (chalazal)	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
bHLH (2)	<i>IBH1</i>	Lin. Cot. to Mat. Green	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
	<i>TT8</i>	Glob. to Mat. Green	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
bZIP (1)	<i>*bZIP44</i>	Glob. to Lin. Cot. (chalazal)	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green
Miscellaneous (1)	<i>*VDD</i>	Glob. (chalazal)	Pre-globular	Globular	Heart	Linear Cotyledon	Maturation Green

Table 3.3. Comparative analysis of expression patterns of candidate *MUM2* repressors across eFP browser and RNA-seq time-course.

Gene expression values for the fifteen LUH-regulated TFs identified as being expressed exclusively in the general seed coat (see Table 3.2). Top row indicates values the general seed coat (SC) expression values from the eFP database. Mean FPKM values $\pm$ standard error of the mean for both Col-0 and *luh-5* whole seeds at the heart stage (3 DPA), linear cotyledon stage (7 DPA) and mature green stage (10 DPA) are shown in the second and third row for each gene. Microarray data from the eFP browser is for wildtype Ws-0 seeds. Expression levels are colour-coded within each data source according to strength, with the darkest and lightest shade of green corresponding to the highest and lowest expression values, respectively. Asterisk (*) indicates discrepancies between the temporal expression patterns of TFs (8/15) in the eFP and RNA-seq datasets in Ws-0 and Col-0 wildtype backgrounds, respectively.

Seed developmental stage					
	Gene		Heart	Lin. Cot.	Mat. Green
Data source	Name	Locus	(3 DPA)	(7 DPA)	(10 DPA)
MADS-box					
eFP browser (Ws-0 SC)	SEP3	AT1G24260	544.8 ±29.1	281.3 ±7.8	182.4 ±9.6
RNA-seq, FPKM (Col-0)			38.6	28.0	6.8
RNA-seq, FPKM (luh-5)			48.3	33.5	9.1
eFP browser (Ws-0 SC)	SHP2	AT2G42830	1854.3 ±205.8	2398.1 ±140.1	1767.7 ±207.3
RNA-seq, FPKM (Col-0)			81.9	95.6	22.8
RNA-seq, FPKM (luh-5)			102.6	147.1	42.9
eFP browser (Ws-0 SC)	AP3	AT3G54340	2279.3 ±1.6	824.3 ±185.8	798.1 ±393.5

RNA-seq, FPKM (Col-0)			99.7	41.2	38.2
RNA-seq, FPKM (<i>luh-5</i>)			108.8	98.5	45.7
eFP browser (Ws-0 SC)	<i>*SHP1</i>	AT3G58780	1602.9 ±237.0	393.5 ±6.5	443.9 ±12.8
RNA-seq, FPKM (Col-0)			70.6	25.6	13.5
RNA-seq, FPKM (<i>luh-5</i>)			78.1	31.9	15.3
eFP browser (Ws-0 SC)	<i>*STK</i>	AT4G09960	1459.8 ±35.1	270.0 ±96.1	902.8 ±178.3
RNA-seq, FPKM (Col-0)			63.6	19.4	14.6
RNA-seq, FPKM (<i>luh-5</i>)			80.3	38.5	15.0
eFP browser (Ws-0 SC)	<i>AG</i>	AT4G18960	235.4 ±0.8	117.5 ±21.8	32.9 ±26.4
RNA-seq, FPKM (Col-0)			15.1	8.6	2.2
RNA-seq, FPKM (<i>luh-5</i>)			18.4	10.2	3.3
eFP browser (Ws-0 SC)	<i>*SEP1</i>	AT5G15800	657.3 ±90.7	732.7 ±89.3	1199.1 ±208.1
RNA-seq, FPKM (Col-0)			86.2	64.4	30.4
RNA-seq, FPKM (<i>luh-5</i>)			104.2	96.2	36.7
AP2-like /ERF					
eFP browser (Ws-0 SC)	<i>*AP2</i>	AT4G36920	1170.0 ±276.6	583.9 ±162.3	1955.3 ±259.1
RNA-seq, FPKM (Col-0)			29.0	22.6	15.5
RNA-seq, FPKM (<i>luh-5</i>)			40.4	32.3	20.4

MYB					
eFP browser (Ws-0 SC)	<i>*MYB3</i>	AT1G22640	149.3 ±30.4	211.8 ±40.2	228.0 ±36.8
RNA-seq, FPKM (Col-0)			30.1	20.1	19.6
RNA-seq, FPKM (<i>luh-5</i>)			42.1	37.6	35.8
eFP browser (Ws-0 SC)	<i>MYB5</i>	AT3G13540	1016.0 ±185.2	481.2 ±37.8	236.0 ±34.2
RNA-seq, FPKM (Col-0)			74.0	58.0	22.1
RNA-seq, FPKM (<i>luh-5</i>)			84.6	81.7	27.9
eFP browser (Ws-0 SC)	<i>TT2</i>	AT5G35550	857.8 ±103.0	178.2 ±56.1	30.3 ±3.1
RNA-seq, FPKM (Col-0)			67.8	25.1	3.2
RNA-seq, FPKM (<i>luh-5</i>)			60.2	36.9	5.1
WRKY					
eFP browser (Ws-0 SC)	<i>*TTG2</i>	AT2G37260	140.7 ±60.7	104.5 ±45.2	193.5 ±13.8
RNA-seq, FPKM (Col-0)			39.7	20.8	11.1
RNA-seq, FPKM (<i>luh-5</i>)			50.1	35.3	14.1
NAC					
eFP browser (Ws-0 SC)	<i>*NARS2</i>	AT1G52880	695.8 ±119.8	1000.2 ±236.7	1771.0 ±21.1
RNA-seq, FPKM (Col-0)			72.0	97.0	77.1
RNA-seq, FPKM (<i>luh-5</i>)			92.0	116.4	84.7
eFP browser (Ws-0 SC)	<i>NAC080</i>	AT5G07680	25.1 ±0.2	106.9 ±51.4	575.1 ±199.0
RNA-seq, FPKM (Col-0)			12.7	26.1	81.2

RNA-seq, FPKM (<i>luh-5</i>)			33.3	26.7	137.7
bHLH					
eFP browser (Ws-0 SC)	<i>*TT8</i>	AT4G09820	293.5 ±24.7	662.9 ±109.2	821.2 ±226.0
RNA-seq, FPKM (Col-0)			34.8	57.9	20.6
RNA-seq, FPKM (<i>luh-5</i>)			48.5	103.0	33.9

Table 3.4. Expression of LUH-regulated TFs in *luh* seeds at early and late developmental stages.

The activity of the 15 LUH-regulated TFs identified that were denoted as being expressed only in the general seed coat (see Table 3.2) were further examined for differential expression in heart (3 DPA) and mature green (10 DPA) stages. Activity in the linear cotyledon stage (7 DPA) is also shown for comparison. Bold text indicates TFs that are candidates for being *MUM2* repressors (7/15) (see text for details). ND - gene not detected as being DE.

Gene Name	Locus	Fold change		
		Heart (3 DPA)	Lin. Cotyledon (7 DPA)	Mature Green (10 DPA)
MADS-box				
<i>SEP3</i>	AT1G24260	1.2	ND	1.3
<i>SHP2</i>	AT2G42830	1.3	1.5	1.9
<i>AP3</i>	AT3G54340	ND	2.4	ND
<i>SHP1</i>	AT3G58780	ND	ND	ND
<i>STK</i>	AT4G09960	1.3	2.0	ND
<i>AG</i>	AT4G18960	1.2	ND	1.5
<i>SEP1</i>	AT5G15800	1.2	1.7	ND
AP2-like /ERF				
<i>AP2</i>	AT4G36920	1.4	1.4	1.3
MYB				
<i>MYB3</i>	AT1G22640	1.4	1.9	1.8
<i>MYB5</i>	AT3G13540	ND	1.4	ND
<i>TT2</i>	AT5G35550	ND	1.5	ND
WRKY				
<i>TTG2</i>	AT2G37260	1.3	1.7	N.D
NAC				
<i>NARS2</i>	AT1G52880	1.3	ND	ND
<i>NAC080</i>	AT5G07680	2.6	ND	1.7
bHLH				
<i>TT8</i>	AT4G09820	1.4	1.8	1.6

Table 3.5. qRT-PCR analysis of seed coat-specific TFs in *luh-4* mutants.

Fifteen elevated TFs expressed throughout the seed coat with elevated expression in *luh-4* shortlisted from Table 3.2 were further examined using qRT-PCR on isolated seed coat tissue from *luh-4* and Col-0 seeds. TFs detected as elevated in both whole seeds and seed coats are indicated in bold (12/15). TFs that display an inverted pattern of differential expression compared to the RNA-seq analysis (2/15) are indicated in red text. The remaining TF (1/15) is not significantly DE and is denoted by ns (not significant). Shaded in purple are candidates selected for further analysis by promoter-VENUSn (8/15).

Statistical significance was calculated using Student's unpaired *t*-test ($p < 0.05$).

		RNA-seq data		qRT-PCR data	
Gene		(whole seeds, lin. cot. <i>luh-5</i>)		(seed coat, lin. cot. <i>luh-4</i>)	
Name	Locus	Fold Change	p-value	Fold Change	p-value
MADS-box TFs					
<i>SEP3</i>	AT1G24260	1.9	2.17E-26	27.9	p<0.001
<i>SHP2</i>	AT2G42830	1.5	8.85E-15	65.9	p<0.001
<i>AP3</i>	AT3G54340	3.3	1.35E-75	50.7	p<0.05
<i>SHP1</i>	AT3G58780	1.5	2.33E-12	1.6	ns
<i>STK</i>	AT4G09960	1.6	1.28E-16	5.0	p<0.01
<i>AG</i>	AT4G18960	2.1	6.63E-34	40.1	p<0.01
<i>SEP1</i>	AT5G15800	1.7	3.92E-24	564.7	p<0.001
MYB TFs					
<i>MYB3</i>	AT1G22640	1.9	1.32E-25	4.2	p<0.001
<i>MYB5</i>	AT3G13540	1.8	4.96E-34	35.6	p<0.001
<i>TT2</i>	AT5G35550	1.8	1.52E-20	156.0	p<0.002
NAC-domain TFs					
<i>NARS2</i>	AT1G52880	1.6	1.32E-19	929.5	p<0.001
<i>NAC080</i>	AT5G07680	2.0	2.10E-20	24.2	p<0.005
Other TFs					
<i>TTG2</i>	AT2G37260	1.9	7.35E-30	-9.8	p<0.002
<i>TT8</i>	AT4G09820	2.6	6.12E-69	4756.1	p<0.002
<i>AP2</i>	AT4G36920	1.5	2.87E-19	-2.1	ns

Table 3.6. Mucilage extrusion phenotypes of lines transformed with amiRNAs targeting *LUH* or *MUM2* and controlled by various seed coat-specific promoters.

Seed coat-specific promoters *AT1G62000*, *DP1* and *GILT* were placed upstream of amiRNAs targeting *LUH* or *MUM2* and introduced into Col-0 plants by transformation. After selecting T₁ plants, T₂ seeds derived from these lines were examined using Ruthenium Red staining. Results of *35S_{pro}* and *AT1G62000_{pro}* were provided by Dr Nadeeka Jayawardana. Mucilage extrusion classifications are shown by the panel below.

Scale bar = 500µm.

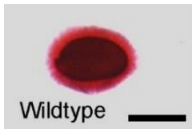
Promoter constructs	Background	Extent of mucilage extrusion defect seen in T ₂ seeds			No. of lines
		Wildtype	Moderate	Severe	
					
		Wildtype	Moderate	Severe	
<i>35S_{pro}</i>					
<i>amiR-LUH</i>	Col-0	5 (19.2%)	6 (23.1%)	15 (57.7%)	26
<i>amiR-MUM2</i>	Col-0	3 (11.1%)	10 (37%)	14 (51.9%)	27
<i>AT1G62000_{pro}</i>					
<i>amiR-LUH</i>	Col-0	29 (100%)	-	-	29
<i>amiR-MUM2</i>	Col-0	10 (16.1%)	31 (50%)	21 (33.9%)	62
<i>DP1_{pro}</i>					
<i>amiR-LUH</i>	Col-0	40 (100%)	-	-	40
<i>amiR-MUM2</i>	Col-0	40 (100%)	-	-	40
<i>GILT_{pro}</i>					
<i>amiR-LUH</i>	Col-0	13 (32.5%)	14 (35%)	13 (32.5%)	40
<i>amiR-MUM2</i>	Col-0	10 (25%)	14 (35%)	16 (40%)	40

Table 3.7. Phenotypes of Col-0 lines transformed with the $35S_{pro}::amiR-AG$ construct.

Three classes of flowers were apparent in T₁ transgenic lines. Aphenotypic flowers looked like wildtype whereas mild floral phenotypes were both sterile and had deformed carpels, with petals replacing stamens. Strong lines have more severe stamen to carpel phenotype compared to mild lines. A total of 40 lines were examined, with approximately 20 seeds per line. Examples of *ag* phenotype are provided below.

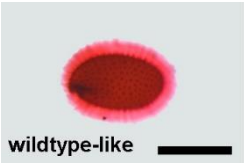
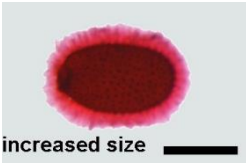
Scale bar = 1mm.

Aphenotypic (wildtype-like)	Mild	Strong	Total no. of lines
11	10	19	40
			

Table 3.8. Phenotypes of seeds derived from transgenic *GILT_{pro}::amiR-AG* lines.

T₂ seeds obtained from transgenic T₁ lines were stained with Ruthenium Red to assess mucilage release. A total of 40 lines were analysed, with approximately 20 seeds per line. No mucilage defects were observed, and instead larger seed sizes of approximately 1.5-fold relative to Col-0 were produced by some lines.

Scale bar = 500µm.

Promoter	Background	Seed size phenotype in T2 seeds		
		No change	Increased	No. of lines
<i>GILT_{pro}</i>	Col-0	28	12	40
		 wildtype-like	 increased size	

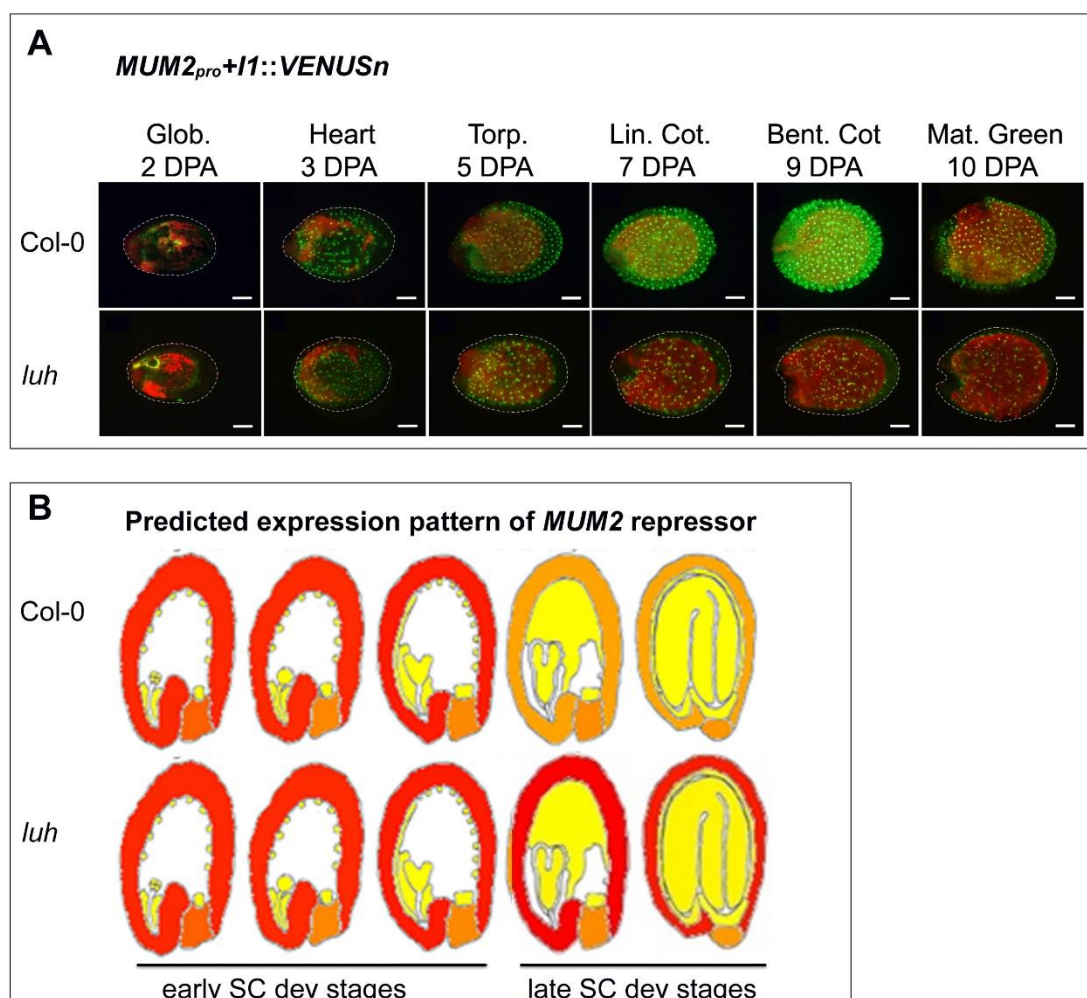


Figure 3.1. *MUM2* promoter (+I1) activity and predicted activity of a *MUM2* repressor.

A: The expression pattern of *MUM2_{pro}+I1* in Col-0 and *luh* seeds. Taken from Jayawardana (2016). Onset of *MUM2* activity appears in the MSCs from 7 DPA onwards in Col-0, while *MUM2* activity is restricted to the endothelial layer in *luh* seeds. Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent Cot.) and Mature Green (Mat. Green). Dotted lines around seed indicate the boundary of the epidermal seed coat layer (MSCs).

Scale bar = 100 μ m.

B: The predicted expression pattern of a *MUM2* repressor in both Col-0 and *luh* backgrounds. In Col-0, the *MUM2* repressor is highly expressed to inhibit *MUM2* activity, and is subsequently repressed by LUH from 7 DPA onwards to allow for

proper *MUM2* expression. In *luh*, activity of the *MUM2* repressor remains high post-7 DPA (late developmental stages). SC: seed coat, dev: developmental. Modified from Winter (2007).

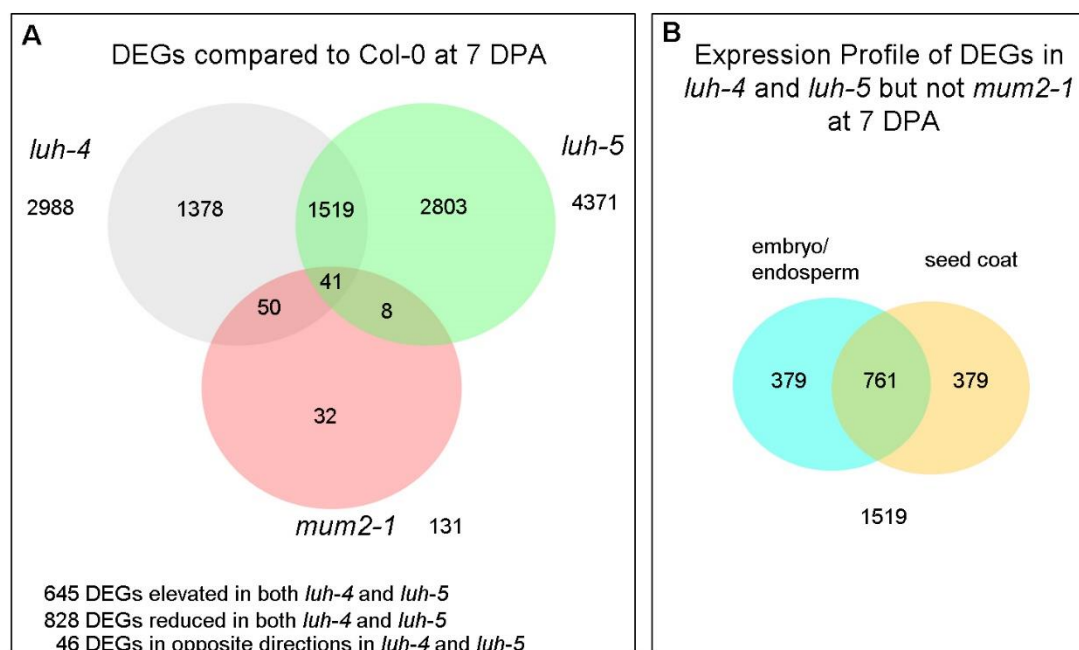


Figure 3.2. Comparison of DEGs present in *luh-4*, *luh-5* and *mum2-1* whole seeds at the linear cotyledon stage of development.

A: The number of DEGs identified in *luh-4*, *luh-5* and *mum2-1* mutants by RNA-seq compared to Col-0 in linear cotyledon (7 DPA) whole seeds are indicated in the purple/grey, green and red circles respectively. The number of DEGs common across backgrounds is indicated in the overlapping regions. The total number of DEGs in *luh* and *mum2* mutant backgrounds are indicated adjacent to each circle.

B: The expression profile (present in embryo/endosperm, seed coat, or both) of DEGs common to both *luh* mutants but not *mum2* seeds was inferred using the the Belmonte dataset (Belmonte et al., 2013).

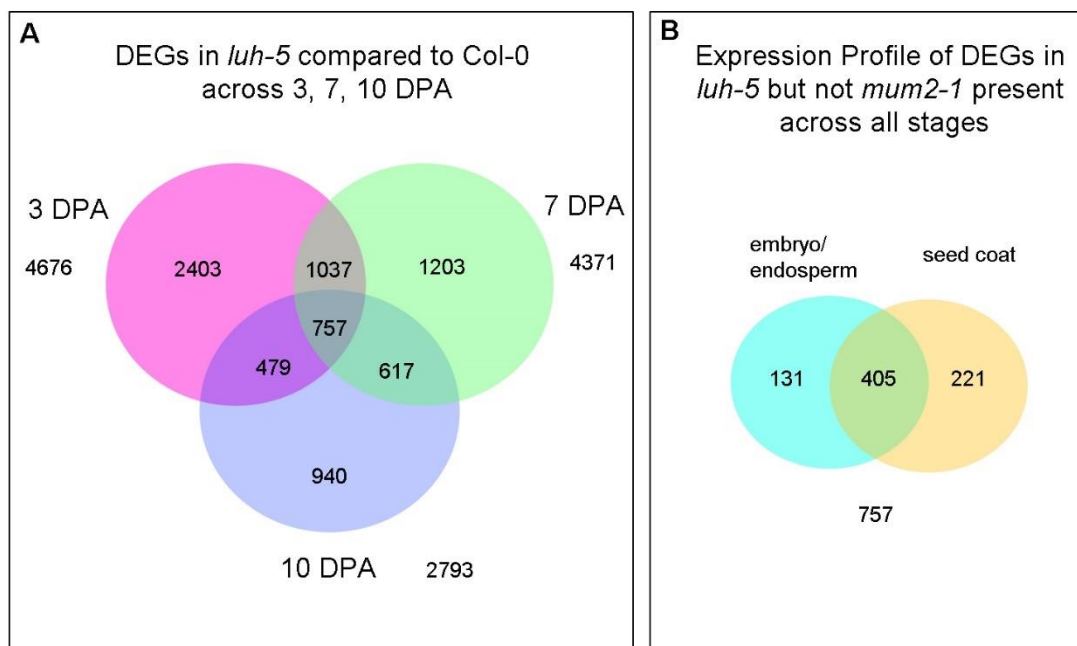


Figure 3.3. Comparison of DEGs present in *luh-5* at different stages of seed development.

A: The number of DEGs identified in *luh-5* mutants by RNA-seq compared to Col-0 in heart (3 DPA), linear cotyledon (7 DPA) and mature green (10 DPA) whole seeds are indicated in the pink, green and blue circles respectively. The number of DEGs common across developmental stages are indicated in respective overlapping regions. The total number of DEGs in respective *luh-5* mutant developmental stages are indicated at adjacent to each circle.

B: The expression profile of DEGs common in all developmental stages of *luh-5* mutants but not *mum2* seeds based on the Belmonte dataset.

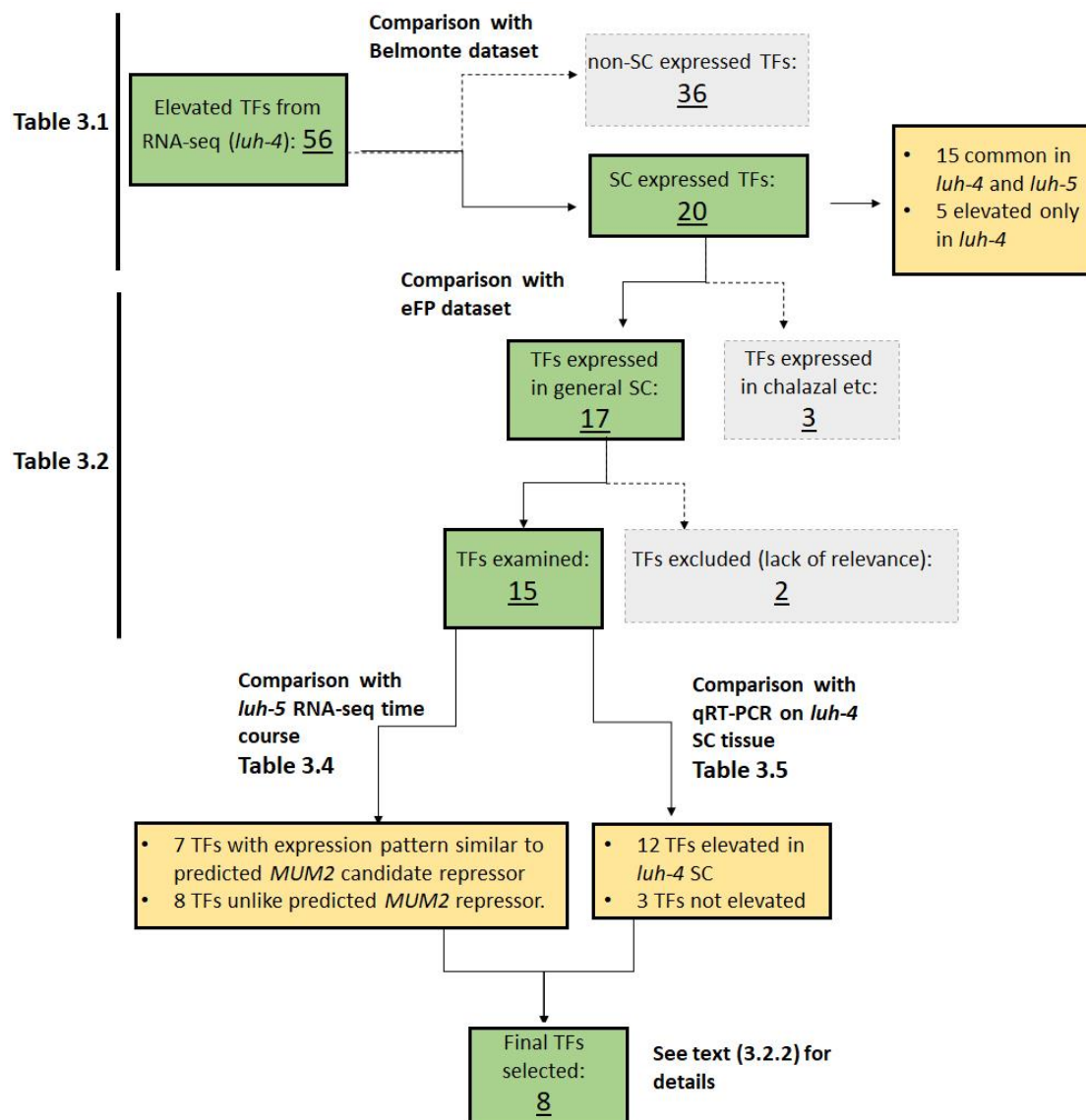


Figure 3.4. Overview of the selection process of TFs elevated in *luh-4* for further examination.

The original 63 TFs identified by Jayawardana, 2016 was put through a more stringent pipeline yielding 56 elevated TFs in *luh-4*. These TFs were then sorted according to expression in, but not limited to the seed coat with the Belmonte dataset (Table 3.1), yielding 20 TFs. The expression pattern in the general seed coat was verified by the eFP dataset (Table 3.2), resulting in 17 TFs, of which 15 were further examined (see text for details). The 15 TFs were verified as robust by RNA-seq time-course on *luh-5* and qRT-PCR on isolated *luh-4* SC tissue. The final 8 TFs were selected from the 15 verified TFs (see **Section 3.2.2** for details). Green boxes indicate

number of TFs carried forward for further examination, grey boxes indicate number of TFs that were eliminated from consideration, and gold boxes indicate additional outcomes of TFs further investigated. SC: seed coat.

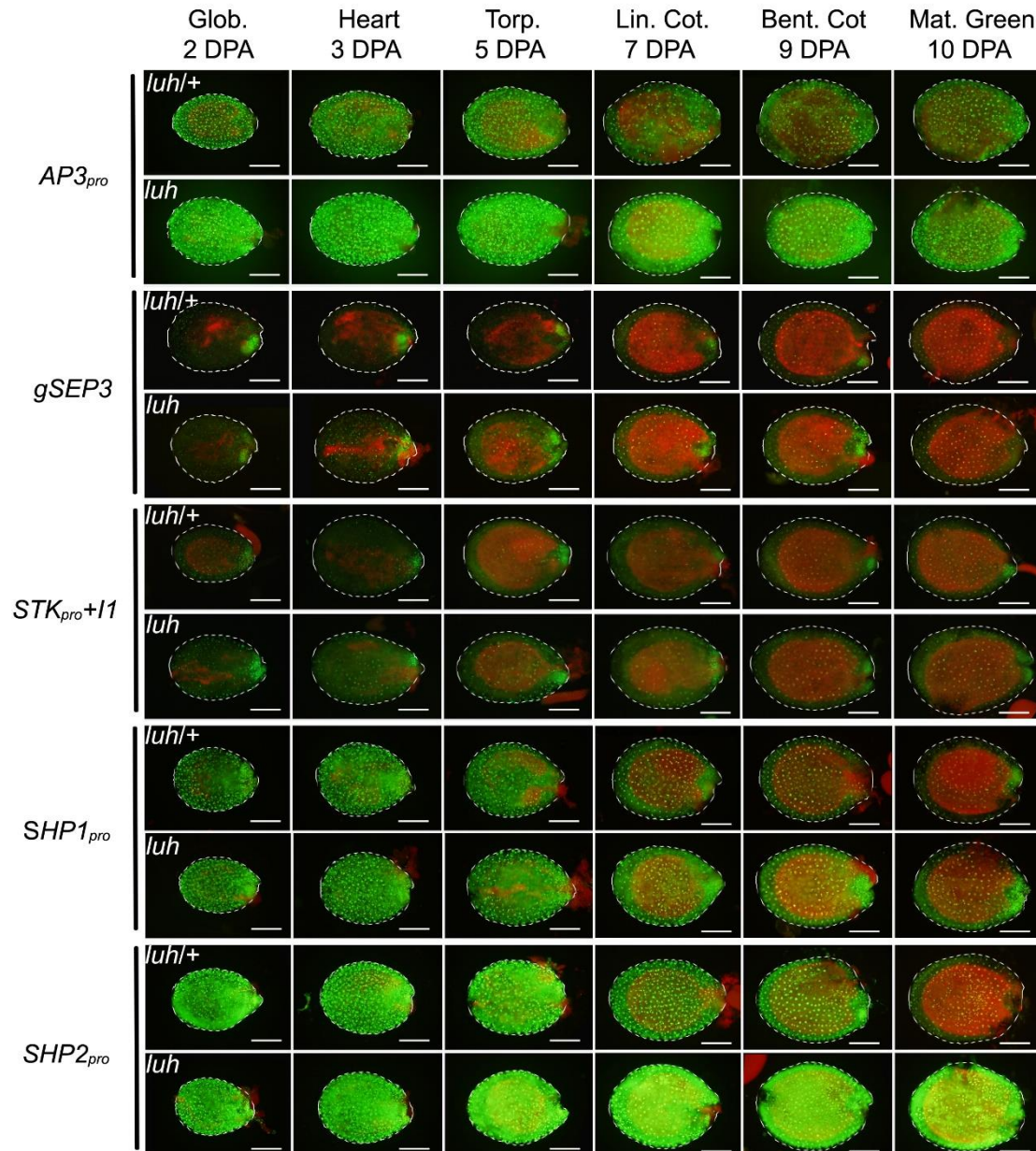


Figure 3.5. Promoter-reporter activity of MADS-box TFs in the developing seed coat.

Promoters of MADS-box TFs were fused to the *VENUSn* reporter and subsequently introduced into *luh* mutants. In the case of *STK*, the large first intron (*I1*) was also included in the construct. *gSEP3* is the entire genomic region of *SEP3* translationally fused to GFP and was kindly provided by Gerco Angenent, Wageningen University and Research. Representative lines were backcrossed to wt and *luh* mutants to generate lines in two different backgrounds (*luh/+* vs *luh/luh*) where activity of the

promoter-reporter is directly comparable. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent. Cot) and Mature Green (Mat. Green). Dotted lines around seed indicate the boundary of the epidermal seed coat layer (MSCs).

Scale bar = 200 μm .

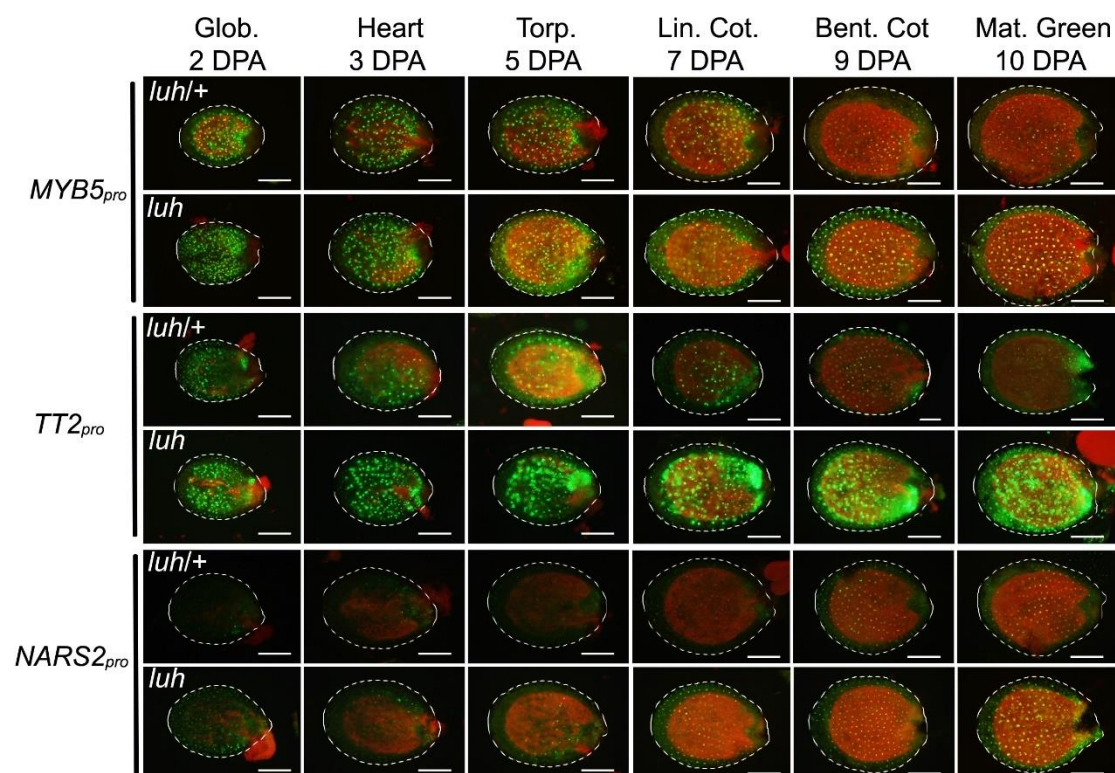


Figure 3.6. Promoter-reporter activity of early differentiation TFs in the developing seed coat.

Promoters VENUSn reporters of TFs associated with the early differentiation of the seed coat were generated as described in Fig. 3.5. Representative lines were backcrossed to Col-0 and *luh* mutants to generate lines in two different backgrounds (*luh/+* vs *luh/luh*) where activity of the promoter-reporter is directly comparable. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent Cot.) and Mature Green (Mat. Green). Dotted lines around seed indicate the boundary of the epidermal seed coat layer (MSCs).

Scale bar = 200 μ m.

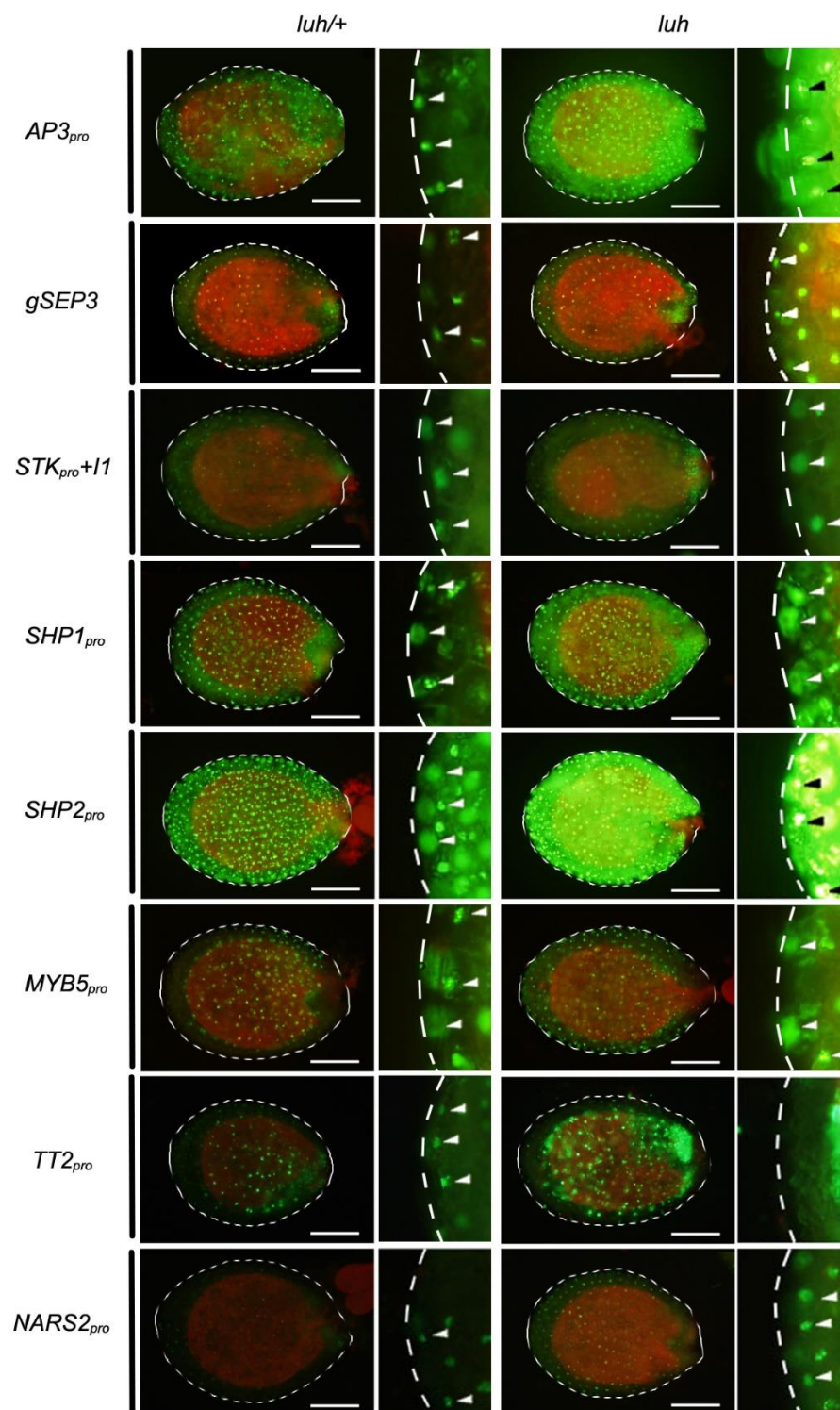


Figure 3.7. Promoter-reporter activity of candidate *MUM2* repressors in the developing seed coat.

Promoters VENUSn reporters of TFs suspected of repressing *MUM2* in the seed coat were generated as described in Fig. 3.5. Included on the right side of each whole seed is a magnification of the seed coat. Shown here is the linear cotyledon (7 DPA) stage, the same stage as samples used in RNA-seq and qRT-PCR. All promoters-VENUSn were expressed in the epidermal layer of the seed coat, except for *TT2 (luh)*. Dotted lines around seed indicate the epidermal seed coat layer (MSCs), with arrowheads indicating *VENUSn* expression in the epidermal layer.

Scale bar = 200 μm .

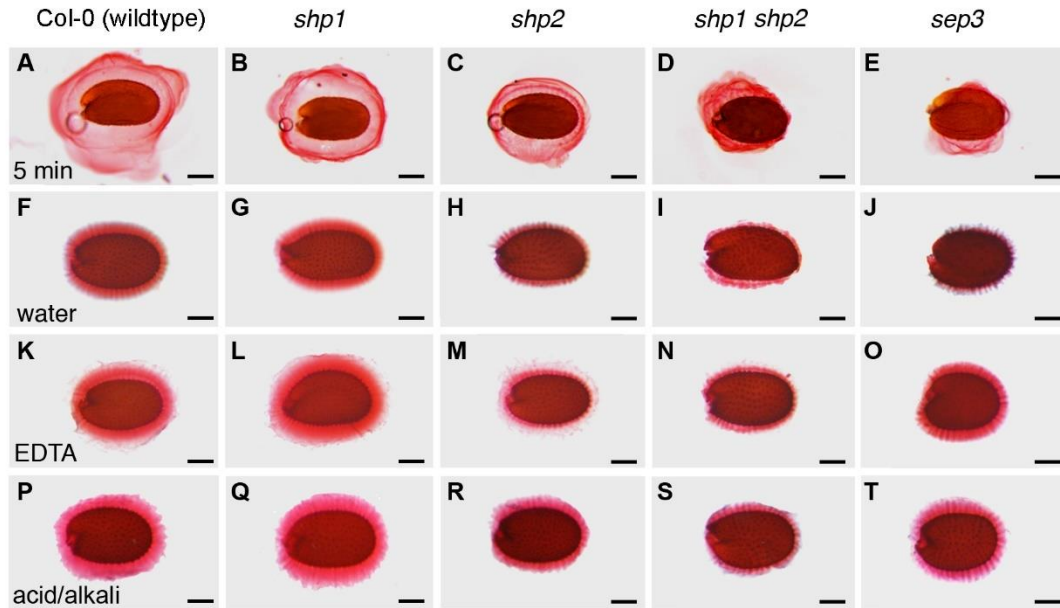


Figure 3.8. Ruthenium Red staining of select MADS-box TF mutant seeds following various treatments.

Seeds from the indicated backgrounds were stained under several conditions: non-shaking incubation for 5 mins for outer mucilage extrusion (A-E), shaking in water to reveal inner adherent layer extrusion (F-J), shaking in EDTA (K-O) and acid/alkali (P-T). Both the EDTA and acid/alkali pre-treatments were used to weaken the cell wall.

Scale bar = 165 μm .

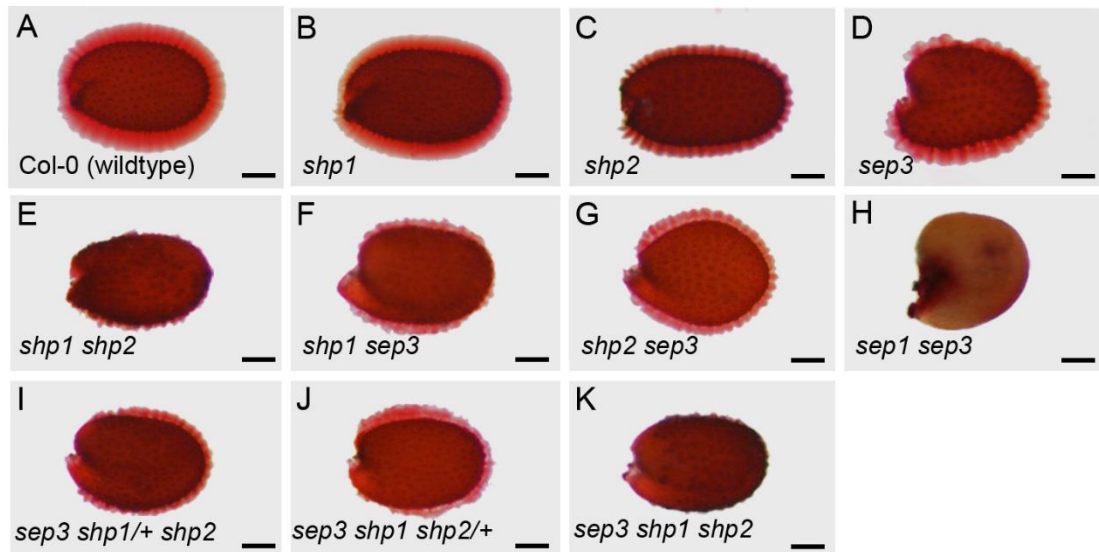


Figure 3.9. Ruthenium Red staining of select MADS-box TF mutant seeds.

Select MADS-box TF mutants were first imbibed in water before staining with Ruthenium Red. Double and triple mutants were generated by crossings.

(A) Col-0 seeds with a broad mucilage halo.

(B-D) Seeds derived from single MADS-box TF mutant plants with genotypes indicated below.

(E-H) Seeds from double MADS-box TF mutant plants. Seeds from double *sep* mutant plants have an enhanced morphological aberration (H).

(I-J) Seeds derived from double mutant plants heterozygous for *shp1* (I) or *shp2* (J).

(K) Seeds derived from triple mutant *sep3 shp1 shp2* plants.

Scale bar = 130 μm.

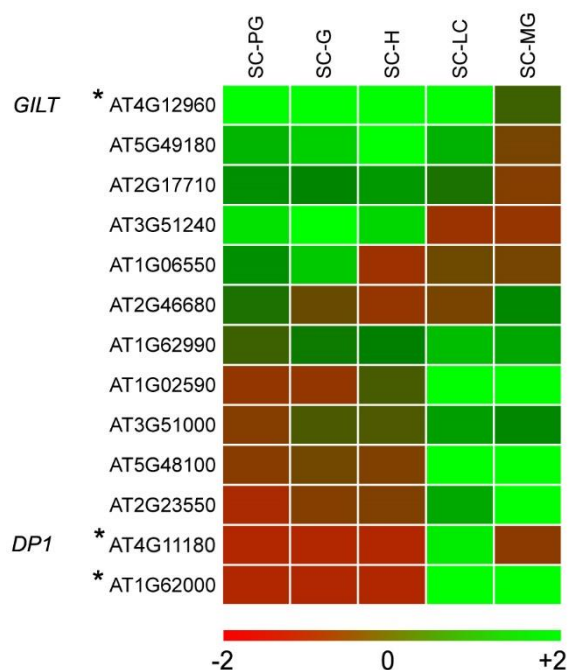


Figure 3.10. Heatmap showing the expression of putative seed coat-specific genes identified in the Belmonte dataset.

A list of candidate seed coat-specific promoters was obtained using methods described in the **Section 3.2.4** that were both expressed in the seed coat and not differentially expressed in *luh* mutants (Belmonte et al., 2013; Winter et al., 2007). A heatmap was generated from this data using the graphical interface tool Protovis (Bostock & Heer, 2009). Colour scale reflects deviation from the overall mean across genes, with red representing low expression values (<-2 S.D) and green representing high expression values ($>+2$ S.D). To accurately assign colours to expression values while maintaining expression patterns within and across genes, a modifier score of 264 was added to each value to prevent saturation. Therefore, gene expressions are directly comparable with each other. The last two genes, *DIRIGENT PROTEIN1* (*DP1*) and *AT1G62000* (*TESTA ABUNDANT*; *TBA1*) are seed coat-specific genes not identified through this analysis but were included for comparisons to the 11 identified candidate seed coat-specific genes (Esfandiari et al., 2013; Tsai et al., 2017). Promoters of genes that were carried forward for further analysis are denoted with an asterisk.

Developmental stages are denoted as following: SC-PG: Seed Coat Pre-Globular; SC-G: Seed Coat Globular; SC-H: Seed Coat Heart; SC-LC: Seed Coat Linear Cotyledon; SC-MG: Seed Coat Mature Green.

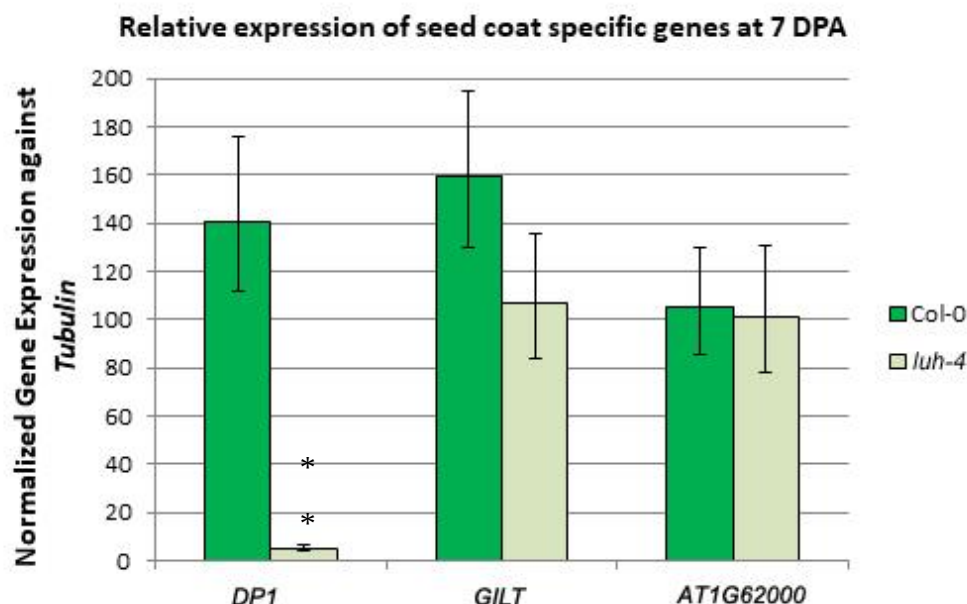


Figure 3.11. Relative expression of putative seed coat-specific genes in Col-0 and *luh-4* mutant seed coats at the linear cotyledon stage.

Expression of putative seed coat-specific genes *DP1*, *GILT* and *AT1G62000* were assessed in Col-0 and *luh-4* seed coat material at linear cotyledon stage using qRT-PCR. Error bars indicate standard error of mean (SEM) across three biological replicates. Statistical significance was calculated using Student's unpaired *t*-test and identified *DP1* as being significantly reduced in *luh-4* seeds (** $p < 0.001$).

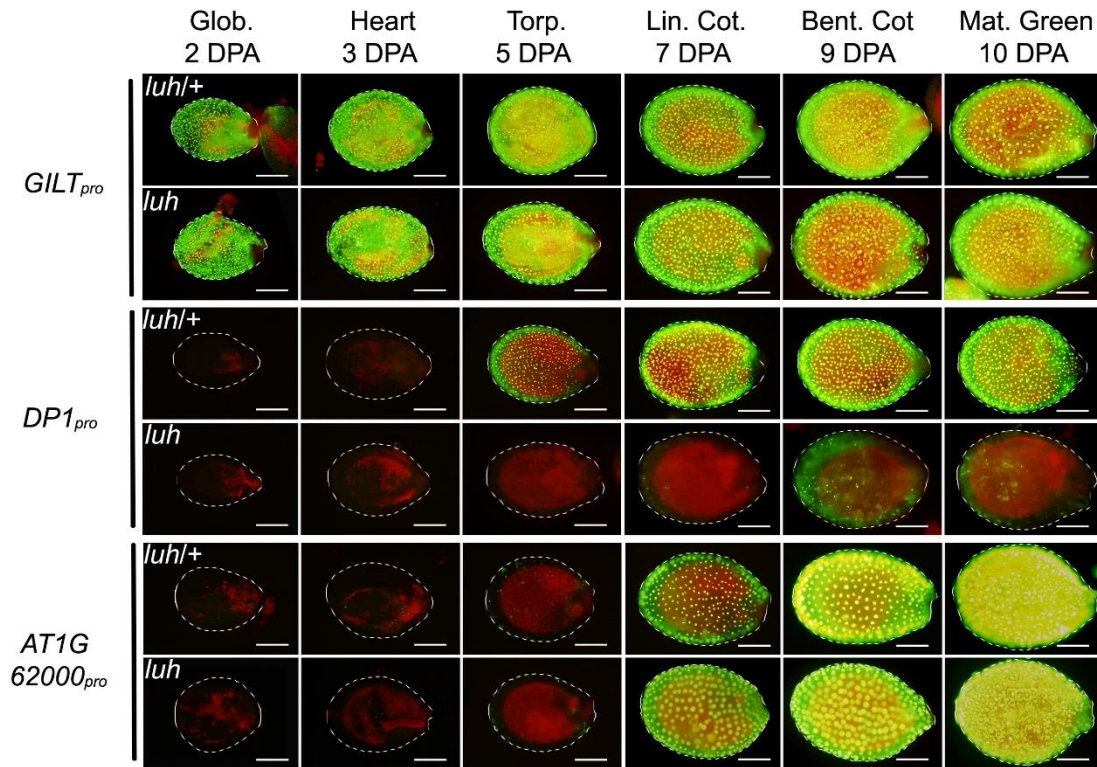


Figure 3.12. Promoter-reporter activity of putative seed coat-specific genes at various stages of seed coat development.

GILT, *DP1* and *AT1G62000* promoters driving the *VENUSn* reporter were introduced into *luh-5* mutants and a representative line was subsequently backcrossed to Col-0 and *luh* mutant to generate lines in two different backgrounds (*luh/+* vs *luh/luh*) where the activity of the promoter-reporter is directly comparable. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent Cot) and Mature Green (Mat. Green). Dotted lines around seed indicate the boundary of the outer seed coat layer. *GILT_{pro}* is active throughout all six stages, while *DP1_{pro}* and *AT1G62000_{pro}* are predominantly active from linear cotyledon stage onwards. Dotted lines around seed indicate the boundary of the outer seed coat layer.

Scale bar = 200 μ m.

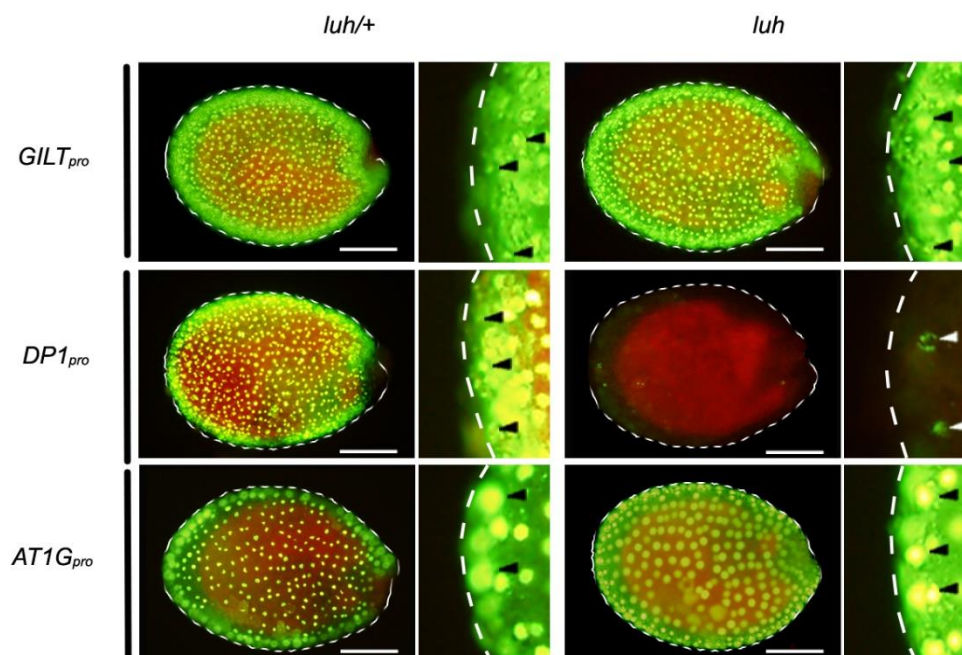


Figure 3.13. Magnified view of promoter-reporter activity in the seed coat at the linear cotyledon stage.

GILT, *DP1* and *ATIG62000* promoter lines fused to *VENUSn* were generated as described in Fig. 3.12. Included on the right side of each whole seed is a magnification of the seed coat. Dotted lines around seed indicate the outer seed coat boundary, with arrowheads indicating the *VENUSn* expression in the epidermal layer.

Scale bar = 200 μm .

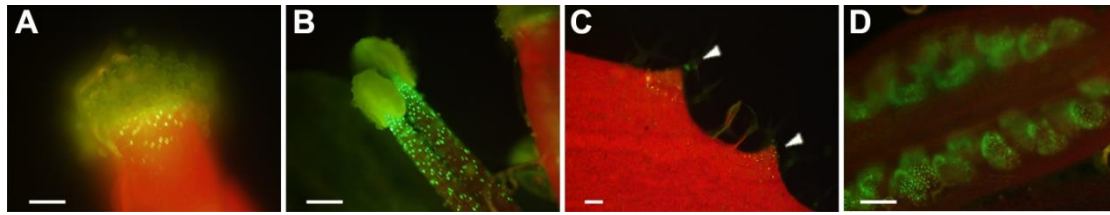


Figure 3.14. Activity of the *GILT* promoter in other plant tissues.

The *GILT*_{pro}::*VENUSn* shows activity in carpel tips (**A**), stamen (**B**), trichomes (**C**, indicated by white arrow heads) and in ovules (**D**) in both Col-0 and *luh* (not pictured).

Scale bar = 100 μ m.

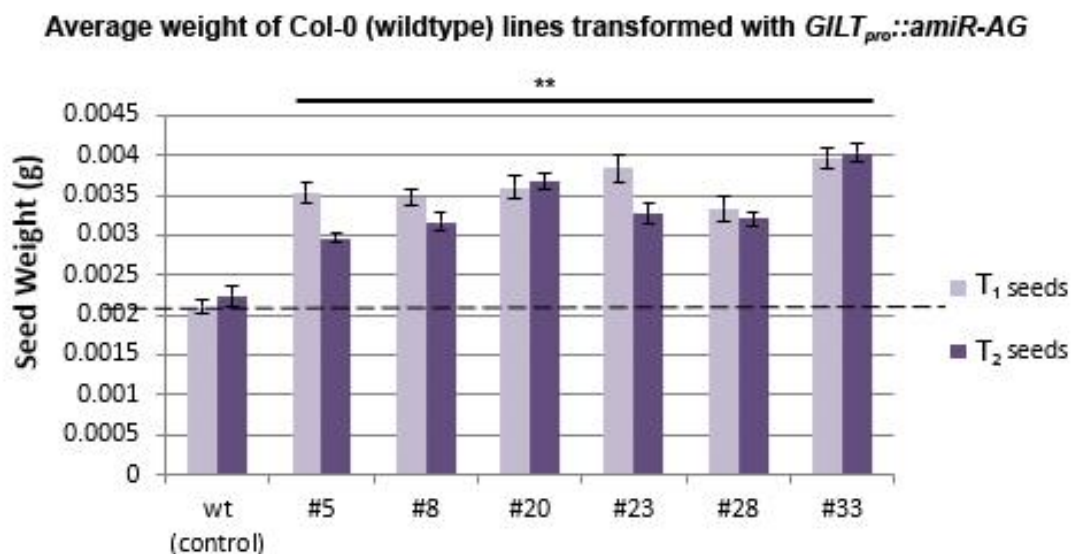


Figure 3.15. Average weight of seeds derived from plants expressing the *GILT_{pro}::amiR-AG* construct.

Segregation analysis on 12 *GILT_{pro}::amiR-AG* T₁ lines scored as having larger seeds on the basis of Ruthenium Red staining revealed that six had single inserts and were brought forward to the T₂ stage. One hundred T₂ seeds per sample were counted and weighed in triplicate (full details in text) with error bars indicating standard deviation. Dotted line indicates baseline weight of Col-0 seeds.

All T₁-T₂ seeds from these six lines were significantly heavier than the Col-0 controls (indicated by **, $p < 0.01$), calculated by Student's paired *t*-test. $n=3$.

Chapter 4

Characterisation of identified MADS-box TF mutants

Characterisation of identified MADS-box TF mutants

4.1 Introduction

In the previous chapter, 15 LUH-responsive TFs were identified as being elevated in *luh* mutants, and thus candidates for the proposed *MUM2* repressor. The function of several of these TFs have been well studied and shown to regulate mucilage polysaccharides production indirectly by promoting differentiation of MSCs such as MYB5, TT2, TT8, NARS2 TTG2 and AP2 (see **Chapter 1, Section 1.2**). These genes were therefore considered unlikely to function directly as *MUM2* repressors according to the model proposed in **Chapter 1, Section 1.3.1** (Fig. 1.9C) and were not considered further. Interestingly, analysis of MADS-box TF mutants *shp2*, double *shp1 shp2* and *sep3* suggested that these genes play a role in mucilage extrusion defects, but that the nature of this role is not known. Therefore, the aim of the work reported in this chapter is to investigate the function of the MADS-box TFs in regulating mucilage extrusion and determine whether this role involves *MUM2*.

Mucilage extrusion defects can arise from disruptions to several processes occurring in the MSCs. These include defects in the specification of MSCs as seen in *ttg1*, *ttg2* and *ap2* mutants (Koornneef, 1981; Western et al., 2001) (described in **Chapter 1, Section 1.2**), decreased mucilage polysaccharides production such as in the *mum4* mutant (Western et al., 2004), alterations to the biochemical structure of mucilage polysaccharides as seen in *luh* and *mum2* mutants (Dean et al., 2007; Macquet et al., 2007b; Walker et al., 2011; Western et al., 2001) or changes to mucilage attachment to the cellulosic rays as exemplified by *sos5* mutants (Griffiths et al., 2014; Voiniciuc et al., 2013). Given the variety of factors impacting on mucilage extrusion, it is necessary to establish which applies to the mucilage extrusion defects observed in *shp2*, *sep3* and *shp1 shp2* double mutants. This is particularly relevant as STK, SHP1, SHP2, SEP3 are known to function redundantly to promote ovule identity and hence disrupting their individual functions may also impact on MSC differentiation

(Pinyopich et al., 2003), and hence poor mucilage extrusion might simply be an indirect consequence of this patterning defect.

Several techniques can be used to study MSC differentiation. These include scanning electron microscopy (SEM) to examine MSC morphology at high magnification, and histological analysis of the developing seed coat (Beeckman et al., 2000; Bozolla, 2007; Dean et al., 2007). Both techniques have their advantages and disadvantages with the main advantage of histological analysis being the ability to view internal structures of cells and thus the extent of the mucilage polysaccharides production and secretion in MSCs can be assessed. In contrast, SEM only provides a surface view of a cell. Despite not being able to assess the extent of mucilage polysaccharides production, the extent of columella formation and radial cell wall thickening, together with a hexagonal cellular morphology are all features of MSC differentiation that can be visualised by SEM analysis. Another important difference between these two techniques is that histological analysis is the more technically challenging as it requires tissue fixation, embedding of tissue in resin and then extensive sectioning. For this reason, SEM was first used to assess MSC morphology of mutant seeds.

As MADS-box TFs were initially identified as candidate *MUM2* repressors, examination of elevated *MUM2* levels in identified MADS-box TF mutants will indicate if *MUM2* is repressed by these TFs. Alternatively, as MADS-box TF mutants identified in **Chapter 3**, particularly triple mutant *sep3 shp1 shp2* reminiscent of *mum2* mutants mentioned in **Section 3.2.4**, a decrease in *MUM2* may potentially account for mucilage defects in identified MADS-box TF mutants. Changes in *MUM2* levels in either scenario can be examined using previously described techniques such as qRT-PCR and promoter-VENUSn reporters in *shp1 shp2* and *sep3* backgrounds relative to their wildtype counterpart. In addition, heterologous expression of *MUM2* under CaMV 35S_{pro} may reveal if loss of *MUM2* is responsible for mucilage defects seen in *shp1 shp2* and *sep3*.

Lastly, a combination of carbohydrate linkage analysis (providing a view of the polysaccharide composition of seed mucilage) and RNA-seq can be used to investigate changes to biosynthesis and maturation/remodelling processes affecting mucilage polysaccharides production in these mutant lines. For instance, the polysaccharide composition of mucilage extracted from wildtype and mutant seeds can be determined and compared by carbohydrate linkage analysis (Lau & Bacic, 1993; Shea et al., 1989). For this procedure, mucilage from both the outer soluble layer and inner adherent layer must first be extracted and then carboxyl-reduced, methylated, acid-hydrolysed, reduced and finally acetylated before the derived permethylated peracetylated alditol acetates are separated, identified and quantified using GC-MS (Kim & Carpita, 1992; Pettolino et al., 2012). Comparisons between wildtype and mutant polysaccharide profiles estimated from this analysis can identify changes in the composition of mucilage and provide insight to the type of gene(s) (e.g. synthase and/or hydrolase) that may have altered activity in these mutant seeds. In addition, RNA-seq, as detailed in **Chapter 3, Section 3.2.1**, can be used to identify DEGs in MADS-box TF mutants relative to wildtype. DEGs identified in MADS-box TF mutants can be subsequently filtered to only include genes either possessing or predicted to possess enzymatic functions associated with the changes in polysaccharide composition. This strategy can produce a list of DEGs that are likely to be associated with mucilage defects seen in the MADS-box TF mutants.

Overall, this chapter describes the verification and characterisation of MADS-box TF mutants, *shp1 shp2* and *sep3*, identified in **Chapter 3** through characterisation of MSC morphology, a series of genetic complementation experiments, examination of the potential loss of *MUM2* in these mutants, and further investigation of the cause(s) of the mucilage defects through a two-pronged approach looking at mucilage composition and transcriptional differences in the mutant lines.

4.2 Results

MADS-box TF mutants *shp2*, *shp1 shp2* and *sep3* were identified in **Chapter 3, Section 3.2.4**, as having mucilage extrusion defects. As *shp1 shp2* double mutants have a more pronounced mucilage extrusion defect than either *shp1* or *shp2* single mutants, characterisation was largely confined to the double mutants unless otherwise stated.

4.2.1 Complementation of MADS-box TF mutants

Prior to investigating the nature of the *shp1 shp2* and *sep3* mucilage phenotypes, it was first important to establish that these defects are due to a loss of their respective MADS-box TF activity. This can be achieved through complementing mutants with their respective full-length coding sequence (CDS) and subsequently assaying for the restoration of normal mucilage extrusion using Ruthenium Red staining. Therefore, to confirm that mucilage defects arising in *shp1 shp2* and *sep3* were due to loss of *SHP* and *SEP3* activity respectively, molecular genetic complementation experiments were performed. *SHP1* was not included in these experiments as based on the analysis of the single *shp1* mutant, this gene plays a less prominent role in regulating mucilage extrusion compared to *SHP2* (**Chapter 3, Section 3.2.4**). Complementation may be achieved through either the use of heterologous promoters such as *35S_{pro}*, or through native promoters inclusive of regulatory introns. A characteristic shared by many members of the MADS-box TF family is the presence of large introns (> 200 bp), which have the potential to contain *cis* regulatory elements necessary for their correct expression (Parenicova et al., 2003). For instance, *AGAMOUS* (*AG*) has *cis* regulatory elements present within the 4-kb second intron that are required for its proper expression in the inner whorls of the developing flower (Sieburth & Meyerowitz, 1997). As both *SEP3* and *SHP2* have large introns, it is possible that *cis* regulatory elements within these introns are also required for their full endogenous expression. Consistent with this view, it has previously been reported that *SEP3* requires *cis* regulatory elements located within the first intron for correct spatial expression in

Arabidopsis in the floral meristem cells (de Folter et al., 2007). However as described in **Chapter 3**, the *SHP2* promoter-VENUSn reporter produced robust fluorescence in the developing seed coat, suggesting that the regulatory elements required for seed coat expression may be contained within the promoter alone (**Chapter, Section 3.2.2**). While characterisation of the native *SHP2* promoter was underway (see **Chapter 3, Section 3.2.3**), it was decided to examine whether over-expression of *SHP2* and *SEP3* using the *35S_{pro}* in their respective mutant backgrounds could restore normal mucilage extrusion (Table 4.1, Fig. 4.1). Constructs were generated using the CDS of *SHP2* and *SEP3* and introduced into *shp1 shp2* and *sep3* mutant plants, and T₂ seeds were analysed for complementation by Ruthenium Red staining of the mucilage. Construct details can be found in Appendix 4, Supplementary Table 2.

The degree of complementation observed in both *shp1 shp2* and *sep3* mutants was classified as either strong (wildtype-like), weak or none following Ruthenium Red staining of T₂ seeds as illustrated in Fig. 4.1. Based on this classification system, 7/40 (17.5%) of the *35S_{pro}::SHP2* lines produced T₂ seeds displaying strong complementation, while the remaining lines produced seeds displaying either weak complementation (21/40; 52.5%) or no complementation (12/40; 30%) (Table 4.1). In contrast, *35S_{pro}::SEP3* lines produced T₂ seeds displaying strong complementation 36/40 (90%) of both mucilage extrusion defects and the seed shape deformity of *sep3* mutants, while the remaining lines produced T₂ seeds with no complementation (4/40; 10%) (Table 4.1, Fig. 4.1). These results confirm that the extrusion defects in *shp1 shp2* and *sep3* mutants are due to the loss of *SHP2* and *SEP3* activity, respectively.

Once seed coat activity of the *SHP2* promoter had been established (see **Chapter 3, Section 3.2.3** for details), it was then used to drive expression of the *SHP2* coding sequence translationally fused to a VENUS tag (Appendix 4, Supplementary Table 2). These constructs were introduced into *shp1 shp2* mutant plants and the resultant T₂ seeds screened for mucilage release by Ruthenium Red staining. T₂ seeds were scored according to mucilage extrusion phenotypes described previously (Table 4.1). *SHP2_{pro}::SHP2-VENUS* in *shp1 shp2* produced strong complementation in 11/22

(50%) lines, weak complementation in 5/22 (22.7%) lines, and 6/22 (27.3%) lines with no complementation (Table 4.1, Fig 4.2). This analysis not only confirms that the loss of *SHP2* activity is responsible for the severe mucilage extrusion defects seen in *shp1 shp2* mutant seeds, but that the large intron is not required for *SHP2* activity in the developing seed coat.

As a previous study had indicated the necessity of *SEP3_{pro}* and its first intron (*I1*) for proper spatial expression of *SEP3* *in planta* (de Folter et al., 2007), an attempt to generate a *SEP3_{pro}I1::SEP3* construct was made but was not successful due to cloning issues. Col-0 seeds containing the promoter and genomic region of *SEP3* driving GFP (*gSEP3*-GFP) were, however, obtained from Prof. Gerco Angenent (Wageningen University, The Netherlands) and subsequently crossed into a *sep3* background. Following this cross, the *sep3*/+ line hemizygous for the *gSEP3*-GFP transgene was subsequently backcrossed to *sep3* mutants. The progeny of this cross produced a population of either *sep3* mutants or *sep3*/+ hemizygous plants with approximately half carrying the *gSEP3*-GFP transgene. A total of six transgenic *sep3* lines carrying *gSEP3*-GFP were identified through genotyping and assayed for mucilage extrusion, which produced 6/6 (100%) *sep3* mutants displaying strong complementation (wildtype-like mucilage and normal seed morphology), indicating *gSEP3*-GFP was functional and sufficient for complementation (Table 4.1, Fig. 4.2). This result, along with those reported for *35S_{pro}::SEP3* (see section above) confirms that the mucilage extrusion defects in *sep3* mutants are caused by a loss of *SEP3* expression and that the first intron is required for seed coat expression.

4.2.2 Examination of the epidermal layer in MADS-box TF mutants

Having established that the loss of *SHP2* in *shp1 shp2* and *SEP3* in *sep3* mutants are responsible for their mucilage defects, mature seeds derived from *shp1 shp2* and *sep3* were examined by SEM alongside wildtype Col-0 seeds, with five seeds per line examined for phenotypic consistency. The outer surface of wildtype seeds is

characterised by the presence of hexagonal-shaped MSCs with a prominent columella in the centre of the cell (Fig. 4.2, Col-0). While *shp1 shp2* double mutants have hexagonal MSCs and columella, the radial cell wall of their MSCs appeared less distinct, suggesting a flatter surface with reduced radial wall and columella heights compared to MSCs of Col-0, *shp1* and *shp2* (Fig. 4.2). This phenotype is reminiscent of *exocyst subunit family protein A1 (exo70A1)* mutants (Kulich et al., 2010), and to a more severe degree, *mum4* mutants (Arsovski et al., 2009b) which may indicate subtle defects downstream of MSC patterning, such as an impact on mucilage polysaccharides synthesis or remodelling.

4.2.3. Expression of *MUM2* in MADS-box TF mutants

To determine whether the MADS-box TF mucilage defects involve disruptions to *MUM2* expression, *MUM2* expression in Col-0, *shp1 shp2*, *sep3*, and *luh-5* mutant whole seeds at the linear cotyledon stage was characterised by qRT-PCR, with normalisation against *ACTIN2 (AT3G18780, ACT2)*. The same RNA samples extracted from age-matched Col-0, *luh-5* and *shp1 shp2* whole seeds at the linear cotyledon stage utilised in RNA-seq experiments were used to synthesise cDNA for qRT-PCR, while *sep3* seeds at the linear cotyledon stage were freshly harvested and RNA extracted. A full list of qRT-PCR primers used for this analysis can be found in Appendix 4, Supplementary Table 3. Interestingly, a statistically significant decrease in *MUM2* in *shp1 shp2* and *sep3* mutants relative to Col-0 (Fig. 4.3) was found. A 4-fold decrease of *MUM2* expression was observed in *shp1 shp2* seeds and a 1.6-fold decrease in *sep3* seeds compared to a 10.7-fold decrease of *MUM2* in *luh-5* seeds (Fig. 4.3). Intriguingly, the fold decrease of *MUM2* in these mutants is correlated with their mucilage extrusion phenotype; the more severe the mucilage extrusion phenotype, the higher the decrease in *MUM2* expression observed (Fig. 4.3). However, as RNA used for the qRT-PCR analysis was extracted from whole seeds, it is possible that the decrease in *MUM2* expression arises in tissue other than the seed coat.

To determine whether *MUM2* activity is reduced in the seed coat of MADS box TF mutants, the *MUM2* promoter *II*-VENUS n reporter was introduced into both *shp1 shp2* and *sep3* mutants using floral dip transformation (Appendix 4, Supplementary Table 2). To generate two different backgrounds (Col-0 and mutant) with the same transgene for direct comparison of VENUS n fluorescence, a crossing approach was utilised as described in **Chapter 3, Section 3.2.3**. Seeds obtained from these crosses were all hemizygous for the transgene and were either *shp1*/ $+$ *shp2*/ $+$ or *sep3*/ $+$ (arising from a cross to Col-0) or *shp1 shp2* or *sep3* mutant (arising from a cross to respective MADS-box TF mutants). Fluorescence in these two different backgrounds therefore arose from the same transgene and hence could be directly compared. This analysis resulted in a visible reduction of fluorescence observed in the MSC of *shp1 shp2* mutants compared to *shp1*/ $+$ *shp2*/ $+$, which has phenotypically wildtype seed coats (Fig. 4.4A & B). Similarly, reduced fluorescence was observed in MSC of *sep3* mutants compared to the phenotypically wildtype *sep3*/ $+$ seed coat (Fig. 4.4A & B).

The analysis of the *MUM2* reporter demonstrates that reduced *MUM2* expression detected in *shp1 shp2* and *sep3* whole seeds is likely to arise from reduced activity of the *MUM2* promoter in the seed coat. This observation is consistent with the MADS-box TFs functioning as *MUM2* activators in the seed coat, and not as a repressor depicted in Fig. 1.9C.

4.2.3.1 Heterologous expression of *MUM2* in MADS-box TF mutants

Reduced *MUM2* expression in the developing seed coats of both *shp1 shp2* and *sep3* mutants raises the possibility that the mucilage release defects observed in these lines is due to loss of *MUM2* activity. To test this possibility, a transgenic approach was taken to elevate *MUM2* levels in developing seeds to see what effect this would have on mucilage release. A previously generated 35S $_{pro}$::*MUM2* construct was introduced into these mutant lines, and T $_2$ seeds derived from transgenic T $_1$ individuals scored for

mucilage release by Ruthenium Red staining. The degree of mucilage extrusion was scored relative to the untransformed mutant phenotype and categorised as being either strong, weak, reduced (reduction in mucilage extrusion relative to untransformed controls) or equivalent (Table 4.2). Images of these phenotypes are shown in Fig. 4.5. *mum2* mutants transformed with *35S_{pro}::MUM2* were used as a positive control so that the range of phenotypes arising from use of the 35S promoter could be assessed. This revealed that 70% transgenic *mum2* mutant lines produced T₂ seeds displaying some form of mucilage release; 14/40 (35%) with strong complementation, 14/40 (35%) with weak complementation and 12/40 (30%) with no complementation (Table 4.2, Fig. 4.5). Having established the capacity of the *35S_{pro}::MUM2* to complement *mum2*, *shp1 shp2* and *sep3*, mutants transformed with *35S_{pro}::MUM2* were generated and their T₂ seeds scored for mucilage release. Strong complementation was observed in 13/40 (32.5%) *shp1 shp2* lines (Table 4.2, Fig. 4.5). As *sep3* mutants have subtle mucilage extrusion defects, mucilage complementation observed was scored as weak relative to the baseline mucilage present in *sep3*. Weak complementation was seen in 16/40 (40%) and 19/40 (47.5%) *shp1 shp2* and *sep3* lines, respectively. Finally, no complementation was seen 11/40 (27.5%) and 11/40 (27.5%) *shp1 shp2* and *sep3* lines respectively (Table 4.2, Fig. 4.5). Interestingly, 11/40 (27.5%) transgenic *sep3* lines displayed reduced mucilage release relative to the untransformed *sep3* control and were therefore categorised as “reduced” (Table 4.2, Fig. 4.5). It is possible that this phenotype arises from transgene-induced silencing of the endogenous *MUM2* gene in the *sep3* background, which would be expected to enhance defects in mucilage release. Overall, these data indicate that a loss of *MUM2* may partially but not fully account for mucilage defects observed in *shp1 shp2* and *sep3*.

4.2.4. Expression of *LUH* in MADS-box TF mutants

Previous work has shown that *LUH* expression is elevated in *stk* seeds, raising the possibility that other MADS-box TFs might regulate *LUH* expression in the MSCs (Ezquer et al., 2016). To test whether the mucilage extrusion defects observed in *shp1*

shp2 and *sep3* mutant seeds might arise from reduced *LUH* activity (**Section 4.2.3**), the expression of *LUH* in *shp1 shp2*, *sep3* and Col-0 whole seeds was examined by qRT-PCR, along with *luh-5* whole seeds as a control for the reduction of *LUH* expression using the same samples used for *MUM2* expression analysis (Fig. 4.6). Although RNA-seq results of *luh-5* later revealed that *LUH* was only reduced by 2.2-fold (Appendix 3, Supplementary Table 2), these results were not available at the time of this experiment. Given that *luh-4* and *luh-5* are phenotypically indistinguishable, it was reasonable to assume at that point in time that *luh-5* is a reasonable control for the loss of *LUH* expression. These experiments were performed in the same manner as detailed in **Section 4.2.3**, with a full list of primers used for qRT-PCR listed in Appendix 4, Supplementary Table 3. Surprisingly, this analysis not only revealed a statistically significant ($p < 0.005$) decrease in *LUH* in *luh-5* (1110-fold), but also *shp1 shp2* (11.5-fold) and *sep3* (5.1-fold) relative to Col-0 (Fig. 4.6). Again, the severity of mucilage extrusion defects in MADS-box TF mutants corresponded to the level of reduction in *LUH* expression, in a similar manner to the loss of *MUM2* described in **Section 4.2.3** (Fig. 4.6). The observed decrease in *LUH* expression, a regulator of *MUM2*, in *shp1 shp2* and *sep3* seeds, might therefore account for the observed mucilage extrusion defects seen in these mutant lines.

4.2.4.1 Heterologous expression of *LUH* in MADS-box TF mutants

To test if the decrease in *LUH* is responsible for the mucilage extrusion defects observed in *shp1 shp2* and *sep3* mutants, *LUH* expression was elevated in *shp1 shp2* and *sep3* mutant seeds through the introduction of a *35S_{pro}::LUH* construct. Following floral dip transformation, Basta resistant T₁ *shp1 shp2* and *sep3* lines were identified and derived T₂ seeds assayed for mucilage extrusion by Ruthenium Red staining. Introduction of the *35S_{pro}::LUH* construct into the *luh-5* background produced approximately 60% of lines exhibiting complementation to some extent; (16/40; 25%) with strong complementation and 13/40 (32.5%) with weak complementation (Table 4.2, Fig. 4.5). In contrast, the *35S_{pro}::LUH* construct was unable to restore the

mucilage extrusion defects of *shp1 shp2* (n=40) and *sep3* (27/40; 67.5%) mutants when introduced into these backgrounds (Table 4.2, Fig. 4.5). Surprisingly, reduced mucilage extrusion was observed in 13/40 (32.5%) *sep3* mutants (Table 4.2, Fig. 4.5). This phenotype was also observed in a small portion of *sep3* mutants transformed with *35S_{pro}::MUM2* (**Section 4.2.3.1**) and again may arise from transgene-induced silencing.

Given that heterologous expression of *MUM2*, but not of *LUH* was capable of partially restoring mucilage extrusion to *shp1 shp2* and *sep3* mutant lines, a reasonable explanation for these observations is that reduced expression of *MUM2*, but not *LUH* is responsible for the observed mucilage extrusion defects seen in these mutant lines. To confirm that elevated expression of *LUH* in these mutant backgrounds did not restore *MUM2* expression levels back to wildtype, three independent transgenic lines with indistinguishable mucilage extrusion phenotypes (biological replicates) of *35S_{pro}::LUH* in *shp1 shp2* were selected, along with three biological replicates each of untransformed Col-0 and *luh* as controls for *MUM2* expression in these backgrounds. Unfortunately, due to time constraints transgenic *sep3* lines were not included in this analysis. The housekeeping gene *ACT2* was used for normalisation of *MUM2* expression as detailed in **Section 4.2.3**, with primers used for the qRT-PCR analysis listed in Appendix 4, Supplementary Table 3. Similar levels of *MUM2* expression between *shp1 shp2* and *35S_{pro}::LUH; shp1 shp2* were observed, with a 4.6-fold and a 4.2-fold decrease in *MUM2* in *shp1 shp2* and *35S_{pro}::LUH shp1 shp2*, respectively, compared to Col-0 (Fig. 4.7A). Therefore, heterologous expression of *LUH* in *shp1 shp2* was unable to increase *MUM2* expression. Given that heterologous expression of *LUH* in *sep3* also failed to rescue mucilage (Table 4.2), it is reasonable to assume that there was no increase in *MUM2* expression in this *sep3* transgenic background.

4.2.5 Carbohydrate analysis of Col-0, *shp1 shp2* and *sep3* mucilage

As the qRT-PCR and promoter/intron-reporter analysis indicated that *MUM2* expression was reduced in *shp1 shp2* and *sep3* mutant seeds, the next step was to determine whether this was also associated with an increase in *t*-Gal residues attached to the RG1 backbone as reported for *mum2* and *luh* mutants (Dean et al., 2007; Macquet et al., 2007b; Walker et al., 2011). To do this, soluble polysaccharides from intact mature seeds were extracted sequentially with an acid and alkali treatment (to remove the outer soluble layer) and sonication (to remove the inner adherent layer) as detailed in **Chapter 2, Section 2.9.1**. These fractions were then subjected to carboxyl reduction and methylation analysis as detailed in **Chapter 2, Section 2.9.2**. Data are presented as mol % (Table 4.3) and re-calculated as a polysaccharide composition in Table 4.4.

As reported previously, the predominant RG1 linkages present in mucilage extracted from Col-0 and *mum2* seeds were 2-linked Rha (2-Rhap) and 4-linked GalA (4-GalAp) with a small quantity of 2,4- and 2,3- linked Rha (2,4- and 2,3-Rhap) as part of the pectic polysaccharide RG1 (Table 4.3; Penfield et al., 2001; Western et al., 2004; Walker et al., 2011). This was also the case for mucilage extracted from both *shp1 shp2* and *sep3* mutant seeds for both the inner and outer layers. Overall, the polysaccharide composition between Col-0 and *mum2* mucilage (inner and outer) were similar with differences mostly residing in the substituted:unsubstituted Rhap residues, i.e. the degree of RG1 branching (comparing Tables 4.3 and 4.4 with Tables I and II of Walker et al., (2011), respectively). In general, the results for the degree of RG1 substitution for *shp1 shp2* and *sep3* seeds did not differ substantially from that of Col-0, but did differ significantly between *mum2* and Col-0 in both inner and outer fractions, with the substituted (2,4- + 2,3-Rhap): unsubstituted (2-Rhap) Rha ratio being approximately 1:30 and approximately 1:5 for Col-0 and *mum2*, respectively (Table 4.4; Walker et al., 2011). These results are similar to those of Walker et al., (2011), with the exception being that in the current study, the ratio of substituted: unsubstituted Rhap residues, for both inner and outer fraction of Col-0 were similar at

approximately 1:30, while ratios of 1:42.3 and 1:14.2 were reported in equivalent fractions by Walker et al., (2011).

The substituted (2,4- + 2,3-Rhap): unsubstituted (2-Rhap) Rhap ratios for both *shp1* *shp2* and *sep3* mutants in the inner and outer layers were comparable to Col-0 (i.e. approximately 1:20-30), in stark contrast to the reduced levels in *mum2* (Table 4.4). The minor polysaccharides (arabinans/Type II AGs/ glucuronoxylans/ heteromannans/ cellulose) were all within the 1-10% range, which means there is little meaningful change between mutants and Col-0 except that the arabinan linkages (t-Araf, 1,5-Araf, and 2,5-Araf) appear to be slightly elevated in the *shp1 shp2* mutant mucilage (Table 4.4). The levels of unassigned residues were of the magnitude previously reported, except for the higher levels in the *mum2* mucilage (both inner and outer) due to an unexplained higher levels of *t*-Galp residues. Based on these results, it seems that reduced *MUM2* expression levels that were detected in *shp1 shp2* and *sep3* mutants do not obviously correlate with increased *t*-Gal substitution as seen in *luh* and *mum2* mutants (Walker et al., 2011). This may mean that although *MUM2* expression is reduced, there may still be sufficient *MUM2*-encoded enzyme to ensure that an equivalent amount of *t*-Gal residues compared to wildtype are removed from the RG1 backbone. Alternatively, it is possible that there is a subtle increase in *t*-Gal residues attached to the RG1 polymers present in *shp1 shp2* and *sep3*, but this is below the level of detection by carbohydrate linkage analysis.

4.2.6 Differentially expressed genes in *shp1 shp2* and Col-0 whole seeds

While mucilage extrusion defects in both *shp1 shp2* and *sep3* seeds appear phenotypically reminiscent of *mum2* mutants, carbohydrate analysis revealed that mucilage extracted from *shp1 shp2* and *sep3* seeds do not have significantly elevated *t*-Gal residues unlike *mum2*. As stated above, an explanation for this may be that there is a subtle change in *t*-Gal substitution of RG1 that is sufficient to alter mucilage extrusion but is not detectable using carbohydrate linkage analysis. Another

conceivable explanation is that factors other than MUM2 may be responsible for mucilage defects seen in *shp1 shp2* and *sep3*. For example, there may be additional changes in different mucilage components, again slight, such as the minor increase in arabinans observed in *shp1 shp2* mutants compared to Col-0, although no significant differences between mucilage extracted from *sep3* and Col-0 seeds were observed.

Therefore, to further examine the causative factors behind the mucilage defects of *shp1 shp2* mutants, RNA-seq was performed on RNA extracted from age-matched *shp1 shp2* mutants and Col-0 seeds at the linear cotyledon stage in three biological replicates as detailed in **Chapter 2, Section 2.10**. RNA-seq was not extended to *sep3* due to a combination of a lack of time and the absence of significant differences observed between mucilage extracted from *sep3* and Col-0 in the carbohydrate analysis. Only DEGs above the expression threshold of 1 count-per-million (CPM), with a lower log-fold change (LFC) of 1.1 across all three biological samples were considered, resulting in a total of 2234 DEGs in *shp1 shp2* mutants relative to Col-0 (Appendix 4, Supplementary Table 4). As expected for T-DNA mutants, the transcript abundance (FPKM) of both *SHP1* and *SHP2* was very low at 1.9 and 4.5 respectively, compared to their respective levels of 25.6 and 95.6 in Col-0 (Appendix 4, Supplementary Table 4).

Next, to look for downstream factors that may account for the mucilage defects seen in *shp1 shp2*, cell wall-related enzymes were specifically identified from the initial list of 2234 DEGs. Given that the onset of mucilage polysaccharides production is from the linear cotyledon stage and is produced in MSCs, the outermost layer of the seed coat, only enzymes present in the seed coat at the linear cotyledon stage were selected. This resulted in 787 DEGs identified as encoding enzymes which were subsequently examined for expression in, but not limited to the seed coat, using the Belmonte microarray dataset (Belmonte et al., 2013), yielding 184 DEGs (Appendix 4, Supplementary Table 5). Finally, genes whose encoded products are involved in housekeeping roles such as ubiquitin, DNA-, RNA-, lipid- associated enzymes were further excluded as they were unlikely to have relevance to mucilage polysaccharides

regulation. Collectively, this analysis produced 39 DEGs in *shp1 shp2* mutants that are expressed in the seed coat at the linear cotyledon stage and are therefore likely to be involved in mucilage defects apparent in *shp1 shp2* lines (Table 4.5).

Examination of the 39 DEGs revealed seven MSC tool box genes, which are genes associated with mucilage mutants and are known to have roles in downstream mucilage polysaccharides regulation such as *GALACTURONOSYLTRANSFERASE-LIKE 5 (GATL5)* (Kong et al., 2013), *GALACTURONOSYLTRANSFERASE11 (GAUT11)* (Francoz et al., 2015), *MUM4* (Western et al., 2004), *MUCILAGE-RELATED10 (MUCI10)* (Voiniciuc et al., 2016), *PEROXIDASE56 (PRX56)* (Ranocha et al., 2014), *CELLULOSE SYNTHASE-LIKE A2 (CSLA02)* (Francoz et al., 2015) and *MUM2* (Walker et al., 2011). Additionally, comparisons of these DEGs in *luh* and *mum2* are made in Table 4.5 to facilitate comparisons and discussion between the profile of *shp1 shp2*, *luh* and *mum2*. The differential expression of these known mucilage polysaccharides production/modification genes in *shp1 shp2* mutants is consistent with the possibility that other cell wall-related changes, in addition to reduced *MUM2* levels, may account for the mucilage extrusion defects seen in this mutant line. Of the 39 DEGs, two genes are associated with arabinan modification; *ALPHA-L-ARABINOFURANOSIDASE1 (ASD1)* and *ASD2* (Table 4.5). *ASD1* and *ASD2* are members of Glycosyl-hydrolase (GH) family 51 and are considered bifunctional arabinofuranosidase/ β -D-xylosidases that can hydrolyse both L-Ara and D-Xyl residues from pectic arabinans and have been reported to produce an increase in arabinans in cell walls of *Arabidopsis* roots and stems (Chavez Montes et al., 2008). Both *ASD1* and closely related *ASD2* are repressed in *shp1 shp2* mutants at 2.0-fold and 1.7-fold respectively, and therefore may be responsible for the slight increase in arabinans observed in the extracted mucilage derived from *shp1 shp2* mutant seeds. In addition to the MSC tool box genes (7/39) and arabinan modification genes (2/39), other DEGs potentially causing alterations to mucilage polysaccharides structure were identified and include genes involved in pectin modification (14/39), pectin synthesis (3/39), xylan modification (1/39) and miscellaneous functions (12/39) (Table 4.5). Given that mucilage predominantly consists of pectin in the form of RG1,

followed by homogalacturonan (HG) and the large number of misexpressed genes involved in pectin modification and synthesis in *shp1 shp2*, it is not unreasonable to infer that the misregulation of these genes might lead to subtle changes in mucilage polysaccharides that are below the detection level of carbohydrate analysis but are collectively sufficient to cause mucilage defects.

Besides enzyme effectors, several cell wall components were found to be differentially expressed in *shp1 shp2* mutants. For example, several arabinogalactans (AGPs) such as fasciclin-like arabinogalactan proteins (FLAs) such as *FLA7*, *FLA11* and *FLA15* were found to be both differentially expressed and present in high amounts in the seed coat (Appendix 4, Supplementary Table 5). Specifically, *FLA7* is decreased by 1.4-fold (229.6 FPKM), *FLA11* by 2.1-fold (16.1 FPKM), while *FLA15* is elevated by 2.5-fold (165.4 FPKM) (Appendix 4, Supplementary Table 4). Of these, only *FLA11* is detected as differentially expressed in *luh* mutants at -12.9-fold (2.5 FPKM), while none of the FLAs are differentially expressed in *mum2* mutants (Appendix 3, Supplementary Table 2 & 4). Given the differential expression of *FLA7*, *FLA11* and *FLA15* in *shp1 shp2* but not *mum2*, it is possible that changes in these cell wall components may account for the differences in the carbohydrate analysis profile in both backgrounds and therefore represent factors needed for proper mucilage extrusion.

4.3 Discussion

4.3.1 Mucilage defects in *shp1 shp2* and *sep3* are due to downstream factors

The characteristic hallmarks of proper seed epidermal differentiation such as the hexagonal shape of MSCs and the presence of a columella were observed across all *shp1*, *shp2*, *shp1 shp2* and *sep3* seeds by SEM (Fig. 4.2). Therefore, it is unlikely that upstream epidermal factors such as MYB5, TT2 etc are causative of the mucilage extrusion defects observed in these MADS-box TF mutants. While *shp1*, *shp2* and *sep3* seeds appeared indistinguishable from Col-0 seeds, the MSC walls and columella of *shp1 shp2* mutants appeared flatter, with less defined radial walls and columella (Fig. 4.2). This phenotype is reminiscent of *exo70A1* mutants (Kulich et al., 2010), and to a more severe degree, *mum4* mutants (Arsovski et al., 2009b). Both *exo70A1* and *mum4* mutants are associated with decreased mucilage polysaccharides production; *exo70A1* results in decreased accumulation of pectins in the MSCs through disruption of pectin deposition into cell walls (Kulich et al., 2010) while *mum4* produces less pectin due to the loss of RHAMNOSE BIOSYNTHESIS 2, (RHM2), a biosynthetic enzyme necessary for RG1 synthesis, the major pectin found in mucilage (Western et al., 2004). Interestingly, RNA-seq on *shp1 shp2* mutants revealed that *MUM4* expression is reduced by 1.9-fold (Table 4.5), suggesting that a decrease in RG1 may be responsible for the flatter MSCs seen in *shp1 shp2*. However, this proposition is not supported by the carbohydrate analysis of mucilage extracted from *shp1 shp2* seeds, which showed similar levels of RG1 to Col-0 (Table 4.4). Unlike *exo70A1*, mucilage defects in *shp1 shp2* do not appear to be related to secretory defects caused by the exocyst complex, consisting of components *EXOCYST COMPLEX COMPONENT SEC3A* (*SEC3A*), *SEC5*, *SEC6*, *SEC8*, *SEC10*, *SEC15*, *EXO70A1*, *EXO84* and *ROH1* (Kulich et al., 2010), as none of these components were detected as differentially expressed in *shp1 shp2* (Appendix 4, Supplementary Table 4).

Having reasonably excluded defects in epidermal differentiation as causing the mucilage extrusion defects seen in *shp1 shp2* and *sep3*, further insight into other

potential downstream factors is presented in the following sections largely through RNA-seq, and to a limited extent carbohydrate analysis. In particular, given that MADS-box TFs SHP1, SHP2 and SEP3 were initially identified as candidate *MUM2* repressors in **Chapter 3**, the relationship between *MUM2*, these MADS-box TFs and their mucilage extrusion defects are discussed and evaluated.

4.3.2 *MUM2*, but not *LUH* expression levels correlates well across RNA-seq and qRT-PCR

Prior to evaluating the relationship between *MUM2* and MADS-box TFs *SHP1*, *SHP2* and *SEP3*, the accuracy of their transcriptomic results must be evaluated and established. As both RNA-seq and qRT-PCR were performed on exactly the same RNA samples extracted from Col-0, *luh-5* and *shp1 shp2* mutants, comparisons between the results obtained by the two techniques can be directly compared for validation and extrapolated to *sep3*. Examination of *MUM2* levels in both *luh-5* and *shp1 shp2* relative to Col-0 appeared to be closely correlated, as both RNA-seq and qRT-PCR recorded similar decreases of *MUM2*; *luh-5* at 10.2-fold (Appendix 3, Supplementary Table 2) and 10.7-fold (Fig. 4.3), respectively and *shp1 shp2* at a 2.2-fold (Fig. 4.3) and 4.0-fold (Fig. 4.3), respectively. These results improve confidence that the decrease of *MUM2* in *sep3* is real and is further supported by increased *MUM2* promoter activity seen in both *shp1 shp2* and *sep3* (Fig. 4.4).

While RNA-seq and qRT-PCR expression data for *MUM2* correlated well in both *luh-5* and *shp1 shp2* mutants, noticeable discrepancies of *LUH* expression levels in these mutants were observed. These discrepancies include the large differences in the fold change of *LUH* detected in *luh-5* mutants, with a 2.2-fold decrease observed in RNA-seq (Appendix 3, Supplementary Table 2), in contrast to the 1110-fold decrease observed in qRT-PCR (Fig. 4.6). In *shp1 shp2* mutants, *LUH* was not detected as being differentially expressed by RNA-seq (Appendix 4, Supplementary Table 4), while qRT-PCR showed a 11.5-fold decrease (Fig. 4.6). One caveat identified in *luh-5*

was that *ACT2*, the normalisation gene used for calculation of normalised expression via the $2^{-\Delta\Delta Ct}$ method, was reduced by 1.2-fold compared to Col-0 (Appendix 3, Supplementary Table 2). Unfortunately, the RNA-seq data derived from *luh-5* were not available at the time qRT-PCR was performed. Although the reduction of *ACT2* expression can impact the relative expression values of *LUH* and other genes detected in *luh-5*, a reduction in the Ct values of *ACT2* in *luh-5* compared to Col-0 was not observed. Ct values in qRT-PCR experiments appeared stable across all mutant and Col-0 samples with values within ± 1 Ct (Appendix 4, Supplementary Table 6). In this context, misexpression of *ACT2* cannot account for the variance between RNA-seq and qRT-PCR data in fold changes obtained for *LUH* expression. Instead, deviations of *LUH* and *ACT2* are likely due to the technical differences between RNA-seq and qRT-PCR experiments, in addition to the analysis approaches which translates to differences in output. In this study RNA-seq provided absolute (in this instance when normalised to FPKM values) while qRT-PCR provided relative gene quantitation (Conesa et al., 2016; Li et al., 2015). Such technical differences include bias that can be introduced into the RNA-seq workflow such as library preparation, intrinsic bias such as sequencing depth and the number of mismatches allowed in target sequences (Fang & Cui, 2011; Li et al., 2015). With regards to qRT-PCR, primers may not necessarily detect all isoforms of a gene (Camacho Londono & Philipp, 2016) which can be detected in RNA-seq due to its ability to account for mismatches (Hatem et al., 2013).

Given that the loss of *LUH* in *shp1 shp2* and by extension, *sep3* could not entirely be established by transcriptional methods (i.e. only qRT-PCR detected loss of *LUH*, but not RNA-seq on *shp1 shp2* mutants), *LUH* promoter-*VENUSn* can be introgressed into these backgrounds and examined for reduced *LUH* promoter activity. However, this appears unnecessary as heterologous expression of *LUH* in *shp1 shp2* did not rescue either mucilage extrusion or *MUM2* levels, indicating that expression of *LUH* at the linear cotyledon stage has no significant impact on *MUM2* expression in *shp1 shp2*. These implications are further discussed in the next section.

4.3.3 MADS-box TFs SHP1, SHP2 and SEP3 may act as *MUM2* activators

MADS-box TFs SHP1, SHP2 and SEP3 were initially identified as candidate *MUM2* repressors in **Chapter 3**, with the expectation that mutants lacking these TFs should produce an increase in *MUM2* expression. However, decreases in *MUM2* expression as revealed by both RNA-seq and qRT-PCR on *shp1 shp2* and qRT-PCR on *sep3* mutants, is inconsistent with SHP1, SHP2 and SEP3 being *MUM2* repressors and instead suggests that they function as *MUM2* activators (Fig. 4.3). This possibility is further supported by the observed decrease in *MUM2* promoter activity in *shp1 shp2* and *sep3* mutants (Fig. 4.4). Additionally, partial rescue of mucilage extrusion in *shp1 shp2* and *sep3* mutants following heterologous expression of *MUM2* (Table 4.2, Fig. 4.5) demonstrates that loss of *MUM2* in these mutants is likely responsible for their mucilage extrusion defects to some extent. However, despite partial mucilage rescue, an expected increase in *t*-Gal attached to the RG1 backbone in *shp1 shp2* or *sep3* mutant mucilage was not observed (Table 4.3 & 4.4). Accounting for these results are two possible scenarios, one where *MUM2* is promiscuous and with heterologous expression, may act on and cleave other pectic substrates that may be implicated in mucilage defects, such as Gal-containing polysaccharides such as FLA15, which was increased in *shp1 shp2* mucilage, therefore artificially rescuing mucilage extrusion. Another possibility is that increases in *t*-Gal residues in *shp1 shp2* and *sep3* are too subtle to be detected by carbohydrate linkage analysis but may be sufficient to cause mucilage extrusion defects – this may be addressed by extending carbohydrate analysis on mucilage extracted from a more severe mutant such as *sep3 shp1 shp2*. While it is possible that other seed coat-expressed BGALs such as BGAL3, which is elevated in *shp1 shp2* mutants (Table 4.5) can compensate for the loss of *MUM2* activity, this is unlikely as *BGAL3* is moderately expressed in *luh* mutants and does not rescue any mucilage extrusion in this background (Jayawardana, 2016). Additionally, probing for possible subtle changes in the composition/organisation of mucilage with a combination of immuno-TEM and glycan microarrays, using the bank of epitope-defined polysaccharide antibodies (Ruprecht et al., 2017) should be performed on *shp1 shp2*, *sep3* and *shp1 shp2* complemented with *MUM2*. These

analyses may reveal subtle changes in mucilage composition/structure/organisation that are affecting mucilage extrusion of *shp1 shp2* and *sep3*. Moreover, comparisons between MUM2 complemented and uncomplemented *shp1 shp2* that have been probed in this manner may reveal if MUM2 is able to act promiscuously on Gal-containing polysaccharides, and if so, can be subjected to further enzyme-substrate testing for confirmation.

Interestingly, despite the similarity of the mucilage extrusion phenotype seen in *shp1 shp2* and *mum2*, of the 39 DEGs identified in Table 4.5, only *MUM2* is differentially expressed in both mutants relative to Col-0 in the same direction, with a 2.2-fold reduction in *shp1 shp2* and an 8.9-fold reduction in *mum2* mutants at the linear cotyledon stage. In contrast, comparisons of DEGs between *shp1 shp2* and *luh-5* produce 21/39 common DEGs, with 9 DEGs in the same direction and 12 DEGs differentially expressed in the opposite direction (Table 4.5). This suggests that *shp1 shp2* and *luh-5* are more similar on a transcriptional level than *mum2* at the linear cotyledon stage, and that LUH, SHP1 and SHP2 may have roles in activating *MUM2* in the same pathway.

Initial examination of *LUH* in *shp1 shp2* and *sep3* compared to Col-0 whole seeds at the linear cotyledon stage revealed significant decreases in *LUH* in these mutants (Fig. 4.6). Given that LUH is well established as a positive regulator of *MUM2* (Huang et al., 2011; Walker et al., 2011), this prompted the question if the loss of LUH in these MADS-box TF mutants is responsible for the corresponding decreases in *MUM2*. While there is a lack of convergence between qRT-PCR and RNA-seq on *LUH* expression levels in *luh-5* and *shp1 shp2* mutants relative to Col-0 (**Section 4.3.2**), neither rescue of mucilage extrusion (Table 4.2 & Fig. 4.5) nor increased *MUM2* levels were observed in *shp1 shp2* and *sep3* lines transformed with *35S_{pro}::LUH* (Fig. 4.7). This outcome clearly shows that lack of LUH is not responsible for decreased *MUM2* expression in *shp1 shp2* and *sep3* and that LUH-induced elevation of *MUM2* is dependent on MADS-box TF activity. Accommodating these results gives rise to two possible models of *MUM2* regulation by SHP1, SHP2 and SEP3 – one is that

MUM2 is activated by MADS-box TFs independently of *LUH*, and the other is that activation of *MUM2* by MADS-box TFs is dependent on *LUH*. In the latter case, the formation of this activation complex would be predicted to occur before the linear cotyledon stage, the stage which marks the onset of *MUM2* expression (Jayawardana, 2016). If the first scenario is correct, the misexpression of *LUH* in either *shp1 shp2* or *sep3* at any stage of seed development will be inconsequential. If the second scenario is correct, misexpression of *LUH* is unlikely to be of consequence in *shp1 shp2* and *sep3* at the linear cotyledon stage but *LUH* must be properly expressed at early seed developmental stages, such as the heart stage. This can be examined with a *LUH* promoter-*VENUSn* reporter that has been introgressed into both *shp1 shp2* and *sep3* backgrounds. However, insight into the mode of *MUM2* activation by these MADS-box TFs, if either independent or dependent on *LUH*, can be better studied through protein-protein interaction studies and *in planta* transactivation assays to determine if such a complex can form and activate *MUM2*. The dependency of MADS-box TFs on *LUH* for *MUM2* activation are further examined and discussed in **Chapter 5**.

4.3.4 Issues associated with carbohydrate analysis of *shp1 shp2* and *sep3* mutants

The data generated from the carbohydrate linkage analysis in *shp1 shp2* and *sep3* offered little insight to the potential composition/structural/organisational differences that may produce the observed mucilage extrusion defects. Given that the *mum2* control produced increased *t*-Gal residues (Table 4.3) as expected, it is possible that carbohydrate linkage analysis lacked the resolution to detect subtle changes in mucilage, particularly of *t*-Gal residues that are likely present in *shp1 shp2* and *sep3*, since they can be partially rescued with heterologous *MUM2* expression (Table 4.2 & Fig. 4.5). The detection of subtle differences in mucilage extracted from *shp1 shp2* and *sep3* may therefore be addressed by further experiments discussed in **Section 4.3.3**.

Another consideration is that the mucilage defects in both *shp1 shp2* and *sep3* might be due to disruptions to downstream processes involved in the secretion, synthesis and/or modification of mucilage polysaccharides as both mutants have relatively normal MSCs. Therefore, the lack of obvious differences between mucilage extracted from Col-0, *shp1 shp2* and *sep3* despite the obvious differences in their respective Ruthenium Red staining (**Chapter 3, Section 3.2.4**) raises the issue if errors in the extraction/processing of the samples could have led to these results. While this scenario is not likely as the *mum2* control produced the expected increase in *t*-Gal, the interpretations arising from this carbohydrate analysis should still be treated with caution. One such example is the increase in total arabinans detected in *shp1 shp2* compared to Col-0. A cursory inspection reveals that the total arabinan content in outer and inner layer fractions amounts to 20.2 mol% for *shp1 shp2* (12.2 mol% outer layer, 8.0 mol% inner layer) and 6.9 mol% for Col-0 (4.5 mol% outer layer, 2.4 mol% inner layer) (Table 4.3). However, comparisons with published data (Walker et al., 2011) revealed that the total arabinan content of Col-0 amounted to 18.3 mol% (4.7 mol% outer layer, 13.6 mol% inner layer), similar to the 20.2 mol% of total arabinans detected in *shp1 shp2* by this study. This suggests that the increased arabinans may not be causative of the mucilage defects observed in *shp1 shp2*. Possible reasons for such discrepancies between the polysaccharide profile of Col-0 in this study and in Walker (2011) include the differences in mucilage extraction methods, such as the differences in acid concentration (0.2 M vs 0.05 M HCl), base concentration (0.05 M vs 1.0 M) and the use of a sonicator in this study. Thus, the reliability of high total arabinan content in *shp1 shp2* requires resolving by either repeating the mucilage extraction in the same manner as Walker (2011), or by antibody staining with LM6 to detect for arabinans.

4.3.5 Other factors involved in mucilage polysaccharides biosynthesis/modification that could be contributing to mucilage defects seen in *shp1 shp2* mutants

Considering that heterologous expression of *MUM2* only partially rescued mucilage extrusion from *shp1 shp2* and *sep3* mutants (Table 4.2, Fig. 4.5), it is worthwhile examining if other downstream factors that are mis-regulated are contributing to the mucilage extrusion defects. A total of 39 candidates encoding enzymes involved in, or otherwise closely related to enzymes involved in polysaccharide regulation that were misexpressed in *shp1 shp2* and present in the seed coat was identified through RNA-seq analysis on *shp1 shp2* and Col-0 (Table 4.5). Unfortunately, the carbohydrate linkage analysis (**Section 4.3.4**) did not assist in the matching of misexpressed candidates to the abnormalities in the mucilage of *shp1 shp2*. The only noticeable deviation in the carbohydrate analysis was the slightly increased arabinan content in mucilage extracted from *shp1 shp2* (Table 4.4) and is questionable as discussed in **Section 4.3.4**. While carbohydrate analysis on mucilage needs to be repeated to address some of the issues raised in **Section 4.3.4**, if increased arabinan content in *shp1 shp2* is taken to be true, then this may be attributed to decreased expression of *ASD1* and *ASD2* (Table 4.5). A reduction in *ASD1*, also known as *ARAF1*, has been reported to produce an increase in arabinans in cell walls of *Arabidopsis* roots and stems, resulting in an increase of anti-arabinan antibody (LM6) signal intensity (Chavez Montes et al., 2008). It is reasonable to expect that the 2.0-fold decrease in *ASD1* and a 1.7-fold decrease in *ASD2* in *shp1 shp2* may result in increased arabinan content in mucilage (Appendix 4, Supplementary Table 4). One caveat is that the RNA-seq was performed on *shp1 shp2* whole seeds, and hence seed coat expression of both *ASD1* and *ASD2* inferred from the Belmonte dataset requires further confirmation by techniques such as promoter-*VENUSn* reporters in Col-0 and *shp1 shp2*. Additionally, if *asd1* and/or *asd2* mutants are found to have a mucilage extrusion defect, this would support the hypothesis that mucilage defects in *shp1 shp2* are caused by a loss of *ASD1* and *ASD2*. Therefore, an important future experiment is to examine *asd1* and *asd2* mutants for mucilage defects by Ruthenium Red as this has not been reported in current published literature. Besides *ASD1* and *ASD2*, seven MSC tool box genes have been identified as misexpressed in *shp1 shp2* and are likely to have some impact on mucilage extrusion, given their roles in mucilage polysaccharides regulation (Table 4.5). Although misexpression of these genes, do not

correlate with abnormalities in *shp1 shp2* mucilage using carbohydrate analysis (**Section 4.3.4**), it is important to consider that there are many different causes of mucilage extrusion defects (i.e. disruptions to mucilage synthesis, export, secretion, maturation, organisation of mucilage polysaccharides etc.) and that genes involved in these processes have not been fully identified. These may include AGP substrates such as FLAs (FLA5, FLA11 and FLA15) that appear to be present in abnormal levels due to differential expression of their respective transcripts detected by RNA-seq. Moreover, loss of SALT-OVERLY SENSITIVE5 (SOS5)/FLA4 has been demonstrated to cause mucilage defects due to changes in cell wall architecture, leading to a loss of pectin adherence (Basu et al., 2016; Griffiths et al., 2014). Differences in the presence of FLA5, FLA11 and FLA15 may therefore impact on the overall organisation and adherence of mucilage polysaccharides to the seed, leading to mucilage defects in *shp1 shp2*. If true, this would partly explain the difficulty in identifying causative factors in *shp1 shp2* by carbohydrate analysis and should be tested through antibody staining to detect FLAs and determine if these FLAs have a role in mediating mucilage structure and organisation.

4.3.6 Conclusion

Mucilage extrusion defects observed in *shp1 shp2* and *sep3* mutants do not appear to be due to disruptions in the upstream formation of MSCs, but rather to disruptions in downstream factors that impact secretion/synthesis/modification of mucilage. While misexpression of genes encoding enzymes such as *MUM2*, *ASD1* and *ASD2* may be causative of these mucilage defects, further experimentation is required to determine the underlying molecular mechanism. Nevertheless, several lines of evidence such as the decrease of *MUM2* in *shp1 shp2* and *sep3* mutants and the subsequent partial rescue of mucilage extrusion in these mutants by heterologous expression of *MUM2* indicate that the loss of *MUM2* is partly responsible for the mucilage defects. These observations are consistent with MADS-box TFs acting as *MUM2* activators, despite being previously identified as candidate *MUM2* repressors in **Chapter 3**.

Further work includes the confirmation of MADS-box TFs as *MUM2* activators and whether this proposed *MUM2* activation is independent or dependent on LUH. These issues are addressed through the use of repressor motif fusion techniques, transactivation assays and protein-protein interaction assays that are described in **Chapter 5**.

Table 4.1. Phenotypes of transgenic *shp1 shp2* and *sep3* mutant lines.

MADS-box TF mutants *shp1 shp2* and *sep3* were complemented with constructs listed below. Ruthenium Red staining was performed on Col-0 (wildtype control) and T₂ seeds from transgenic lines. The degree of complementation exhibited in the T₂ seed was categorised as strong, weak, or none as shown in Fig. 4.1.

Construct	Background	Complementation phenotype of T ₂ seeds			
		Strong	Weak	None	No. of lines
<i>35S_{pro}::SHP2</i>	<i>shp1 shp2</i>	7	21	12	40
<i>35S_{pro}::SEP3</i>	<i>sep3</i>	36	-	4	40
<i>SHP2_{pro}::SHP2- VENUSn</i>	<i>shp1 shp2</i>	11	5	6	22
<i>gSEP3-GFP</i>	<i>sep3</i>	6	-	-	6

Table 4.2 Phenotypic rescue of MADS-box TF mutants by *MUM2* or *LUH*.

Due to variance in mucilage defects of MADS-box TF mutants, strength of complementation phenotypes were scored relative to the characteristic mucilage defect of each background. Examples of each complementation phenotype are shown in Fig. 4.5. *35S_{pro}* driving *MUM2* in *mum2-1*, and *35S_{pro}* driving *LUH* in *luh-5* were used as positive controls.

Construct	Transgenic background	Complementation phenotype of T ₂ seeds				
		Strong	Weak	None	Reduced	No. of lines
<i>35S_{pro}::MUM2</i>	<i>mum2-1</i>	14	14	12	-	40
	<i>shp1 shp2</i>	13	16	11	-	40
	<i>sep3</i>	-	19	11	11	40
<i>35S_{pro}::LUH</i>	<i>luh-5</i>	16	13	11	-	40
	<i>shp1 shp2</i>	-	-	40	-	40
	<i>sep3</i>	-	-	27	13	40

Table 4.3 Sugar linkage composition of extracted mucilage from Col-0, *mum2*, *shp1 shp2* and *sep3* seeds.

Soluble polysaccharides from mature seeds were extracted sequentially with acid and alkali treatment (outer soluble layer) and sonication (inner adherent layer). Samples were then neutralised and following carboxyl reduction and methylation, the partially methylated alditol acetates were identified and quantified by GC-MS (see **Chapter 2, section 2.9.2**). Data are reported as mol % and are the average of duplicates. ND, Not Detected; tr, Trace (<0.5 mol %)

		<i>shp1 shp2</i>		<i>sep3</i>		<i>mum2</i>		Col-0	
		Outer	Inner	Outer	Inner	Outer	Inner	Outer	Inner
Rhap									
	t-Rhap	1.5	1.0	1.3	1.2	2.6	1.4	1.7	1.2
	1,2-Rhap	40.9	46.2	51.5	57.0	30.7	28.1	47.2	54.8
	1,2,3-Rhap	0.4	0.5	0.7	0.3	0.1	0.1	0.4	0.4
	1,2,4-Rhap	1.0	1.4	1.7	1.6	5.8	7.7	1.2	1.5
	Total	43.9	49.2	55.1	59.9	39.2	37.3	50.4	58.0
Araf									
	t-Araf	4.1	2.3	2.0	0.7	1.3	2.8	1.7	0.8
	1,2-Araf	0.8	0.7	0.3	0.1	0.6	0.5	0.1	0.1
	1,3-Araf	1.0	0.9	0.5	0.3	-	0.4	0.2	0.5
	1,5-Araf	5.5	2.9	1.9	0.7	0.2	3.3	2.3	0.7
	1,2,5-Araf	0.8	1.3	0.2	0.1	-	0.5	0.2	0.2
	Total	12.2	8.0	4.9	1.9	2.2	7.5	4.5	2.4
Fucp									
	t-Fucp	0.1	-	0.1	-	-	0.2	-	-
	Total	0.1	-	0.1	-	-	0.2	-	-
Xylp									
	t-Xylp	3.6	3.0	2.9	3.3	6.5	2.6	4.0	3.4
	1,2-Xylp	0.2	0.2	0.2	0.2	0.3	0.2	0.3	0.3
	1,4-Xylp	4.1	4.3	3.2	3.7	5.3	2.9	5.3	4.9
	1,2,4-Xylp	1.7	2.1	2.3	2.7	1.2	0.8	1.8	2.8
	1,2,3,4-Xylp	-	0.3	0.2	0.6	0.1	1.4	0.3	0.2
	Total	9.6	9.9	8.7	10.6	13.4	7.8	11.7	11.5
Manp									
	1,4-Manp	0.2	0.1	0.4	0.1	0.1	0.6	0.2	0.2
	1,4,6-Manp	0.1	-	0.1	0.1	-	-	-	-
	Total	0.3	0.1	0.5	0.1	0.1	0.6	0.2	0.2
Glc p									

	t-Glcp	0.4	0.6	0.2	0.1	0.3	0.4	0.5	0.1
	1,4-Glcp	2.6	1.9	1.9	1.6	2.6	2.0	3.3	1.4
	1,2,4-Glcp	0.2	0.2	0.4	0.3	-	-	0.1	0.1
	1,4,6-Glcp	0.1	-	0.1	0.1	tr	0.1	0.2	0.1
	Total	3.3	2.7	2.7	2.1	2.9	2.5	4.2	1.8
Galp									
	t-Galp	1.5	0.7	1.4	0.9	18.5	16.7	1.0	0.8
	1,3-Galp	0.3	0.1	0.9	-	-	-	-	-
	1,6-Galp	0.2	-	0.2	-	-	0.5	0.1	-
	1,3,6-Galp	0.6	0.3	1.1	0.1	-	0.7	0.2	0.1
	Total	2.6	1.2	3.6	1.0	18.5	17.9	1.3	0.9
GlcAp									
	t-GlcAp	-	-	tr	-	-	-	tr	-
	Total	-	-	tr	-	-	-	tr	-
GalAp									
	1,4-GalAp	28.1	29.0	24.3	24.4	23.7	26.1	27.6	25.3
	Total	28.1	29.0	24.3	24.4	23.7	26.1	27.6	25.3

Table 4.4 Calculated polysaccharide composition of extracted mucilage from Col-0, *mum2*, *shp1 shp2* and *sep3* seeds.

Sugar linkage composition data (mol %; from Table 4.3) was assigned to relevant polysaccharides found in mucilage and their corresponding composition is summarised as below.

	<i>shp1 shp2</i>		<i>sep3</i>		<i>mum2</i>		Col-0	
	Outer	Inner	Outer	Inner	Outer	Inner	Outer	Inner
Arabinan	7.1	5.4	2.3	0.9	0.2	4.2	2.7	1.2
Type II AG	2.4	1.5	3.6	0.4	0.6	2.4	0.5	0.3
RG I (substituted: unsubstituted ratio)	70.4 (1:29.2)	77.1 (1:24.3)	78.1 (1:21.5)	83.1 (1:30.0)	60.2 (1:5.2)	62.0 (1:3.6)	76.4 (1:29.3)	82.0 (1: 28.8)
Glucuronoxylan	6.7	7.5	6.2	7.3	6.6	5.5	7.7	8.3
Heteromannan	0.8	0.2	1.2	0.5	0.3	1.5	0.9	0.4
Xyloglucan	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5
Cellulose	2.4	1.8	1.5	1.5	2.4	1.4	3.1	1.1
Unassigned	10.1	6.5	6.9	6.3	29.5	23.0	8.7	6.1

Table 4.5 List of differentially expressed genes encoding cell wall-related enzymes in *shp1 shp2* seeds at the linear cotyledon stage.

RNA-seq between *shp1 shp2* and Col-0 seeds yielded 2234 DEGs at the linear cotyledon stage (7 DPA). Thirty-nine DEGs encoding enzymes were identified as expressed in the seed coat at the linear cotyledon stage according to the Belmonte dataset (Belmonte et al., 2013) (see text for details). ND: not detected as differentially expressed. FPKM values indicated in parenthesis.

Locus		Gene	Fold change at lin. cot. (7 DPA)		
			<i>shp1 shp2</i>	<i>mum2</i>	<i>luh-5</i>
MSC tool box genes (7)					
<i>AT1G02720</i>	<i>GALACTURONOSYLTRANSFERASE-LIKE 5 (GATL5)</i>		-2.3 (87.5)	ND	ND
<i>AT1G18580</i>	<i>GALACTURONOSYLTRANSFERASE11 (GAUT11)</i>		-1.4 (20.5)	ND	1.4 (22.5)
<i>AT1G53500</i>	<i>MUCILAGE-MODIFIED4 (MUM4)</i>		-1.9 (359.7)	ND	ND
<i>AT2G22900</i>	<i>MUCILAGE-RELATED10 (MUCI10)</i>		-1.6 (74.0)	ND	1.5 (103.8)
<i>AT5G15180</i>	<i>PEROXIDASE 56 (PRX56)</i>		-2.4 (104.1)	ND	-1.5 (164.3)
<i>AT5G22740</i>	<i>CELLULOSE SYNTHASE-LIKE A2 (CSLA02)</i>		-1.6 (24.1)	ND	ND
<i>AT5G63800</i>	<i>MUCILAGE-MODIFIED2 (MUM2)</i>		-2.2 (34.0)	-8.9 (3.3)	-10.2 (7.4)
Pectin modification (13)					
<i>AT1G09550</i>	<i>PECTIN ACETYLESTERASE1 (PAE1)</i>		-27.3 (7.4)	ND	-11.9 (22.2)
<i>AT1G11590</i>	<i>PECTIN METHYLESTERASE19 (PME19)</i>		-4.9 (5.1)	ND	ND
<i>AT1G20550</i>	-		1.7 (26.3)	ND	-1.3 (36.9)
<i>AT1G23200</i>	<i>PECTIN METHYLESTERASE69 (PME6)</i>		-1.7 (709.1)	ND	-4.0 (805.7)
<i>AT1G57590</i>	<i>PECTIN ACETYLESTERASE1 (PAE2)</i>		2.2 (27.8)	ND	-15.3 (1.1)

AT1G70370	POLYGALACTURONASE 2 (PG2)	-1.8 (24.5)	ND	-1.3 (32.4)
AT2G22620	-	1.4 (148.2)	ND	-1.3 (56.1)
AT2G28470	BETA-GALACTOSIDASE8 (BGAL8)	-1.5 (28.9)	ND	1.9 (61.6)
AT2G43050	PECTIN METHYLESTERASE PCR D (PMEPCRD)	-2.8 (58.2)	ND	-165.4 (12.4)
AT4G03930	PECTIN METHYLESTERASE42 (PME42)	-3.9 (5.5)	2.7 (7.6)	1.6 (7.6)
AT4G36360	BETA-GALACTOSIDASE3 (BGAL3)	1.8 (26.0)	ND	ND
AT5G62360	PECTIN METHYL-ESTERASE INHIBITOR 13 (PMEI13)	13.8 (14.1)	ND	ND
AT5G63180	-	1.6 (17.8)	ND	-1.7 (16.5)
Pectin synthesis (4)				
AT1G70090	GALACTURONOSYLTRANSFER ASE-LIKE 9 (GATL9)	-1.4 (26.5)	ND	ND
AT3G14790	RHAMNOSE BIOSYNTHESIS3 (RHM3)	-1.3 (24.5)	ND	ND
AT3G28340	GALACTURONOSYLTRANSFER ASE-LIKE 10 (GATL10)	-1.4 (28.3)	ND	ND
AT5G47780	GALACTURONOSYLTRANSFER ASE4 (GAUT4)	-1.2 (70.0)	ND	ND
Xylan modification (1)				
AT2G37090	IRREGULAR XYLEM 9 (IRX9)	-8.0 (5.1)	ND	ND
Arabinan modification (2)				
AT3G10740	ALPHA-L- ARABINOFURANOSIDASE1 (ASD1)	-2.0 (48.2)	ND	-2.6 (2.1)
AT5G26120	ALPHA-L- ARABINOFURANOSIDASE2 (ASD2)	-1.7 (45.8)	ND	ND
Miscellaneous (12)				
AT1G29050	TRICHOME BIREFRINGENCE-	1.4	ND	1.5

	<i>LIKE 38 (TBL38)</i>	(834.9)		(881.2)
<i>AT1G61810</i>	<i>BETA GLUCOSIDASE45</i>	-1.5	ND	-11.1
	<i>(BGLU45)</i>	(3.4)		(0.3)
<i>AT1G63140</i>	-	7.1	ND	
		(5.8)		
<i>AT2G38180</i>	-	2.0	ND	-1.9
		(50.0)		(13.0)
<i>AT2G44450</i>	<i>BETA GLUCOSIDASE15</i>	1.4	ND	ND
	<i>(BGLU15)</i>	(55.9)		
<i>AT3G21190</i>	<i>MANNAN SYNTHESIS</i>	-1.3	ND	ND
	<i>RELATED1 (MSR1)</i>	(123.2)		
<i>AT3G47590</i>	-	-1.3	ND	1.3
		(21.9)		(26.1)
<i>AT4G21960</i>	<i>PEROXIDASE XR1 (PRX1)</i>	1.3	ND	ND
		(1963.2)		
<i>AT4G34480</i>	-	-1.3	ND	ND
		(37.9)		
<i>AT5G18460</i>	-	-1.3	ND	-1.6
		(8.3)		(6.0)
<i>AT5G45650</i>	<i>SUBTILISIN-LIKE SERINE</i>	1.7	ND	-2.4
	<i>PROTEASE5.6 (SBT5.6)</i>	(28.2)		(12.8)
<i>AT5G48100</i>	<i>TRANSPARENT TESTA 10 (TT10)</i>	-1.3	ND	ND
		(589.4)		

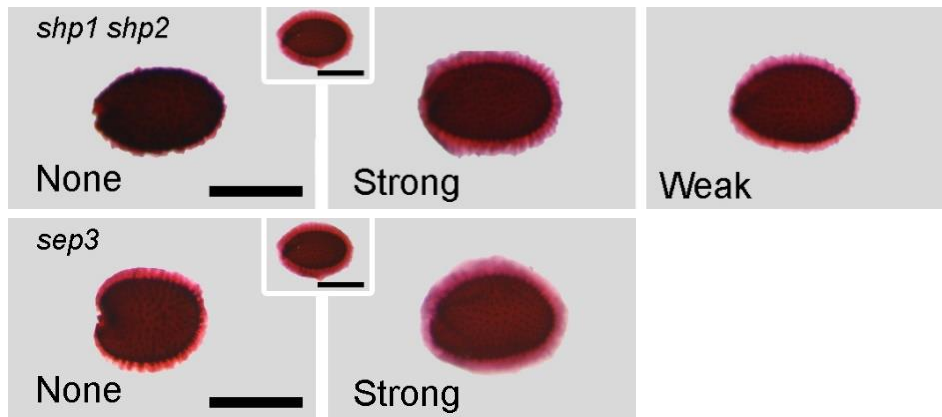


Figure 4.1. Categorisation of T₂ seeds for complementation analysis.

The strength of complementation phenotypes in T₂ seeds was scored relative to the characteristic mucilage defect of each mutant background. Shown here are examples of complementation phenotypes (strong, weak and none) of *shp1 shp2* complemented with either *35S_{pro}::SHP2* or *SHP2_{pro}::SHP2-VENUSn* and of *sep3* with *35S_{pro}::SEP3* and gSEP3-GFP as detailed in Table 4.1, with Col-0 wildtype control as inset.

Scale bar = 500 μ m.

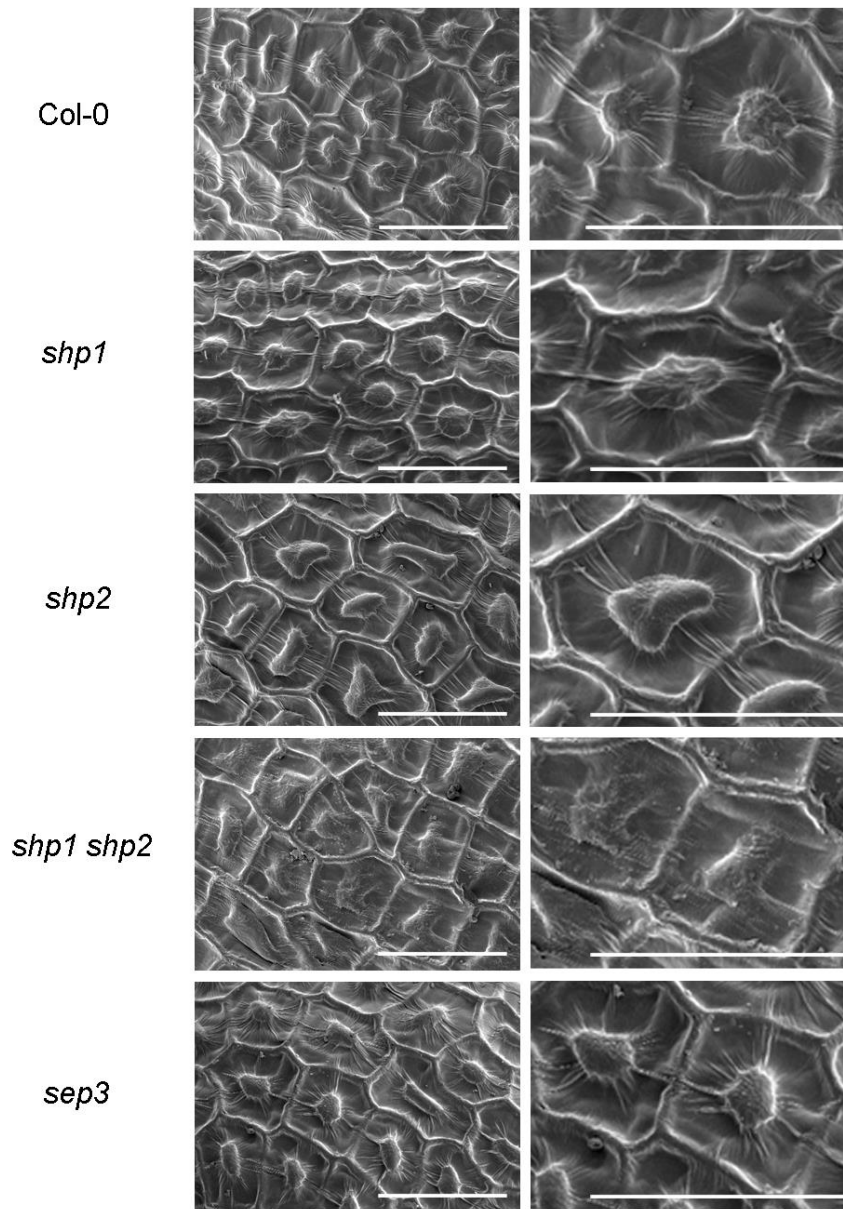


Figure 4.2. Seed coats of mature MADS-box TF mutant seeds as seen by SEM.

Gold coated wildtype (Col-0), *shp1*, *shp2*, double *shp1 shp2* and *sep3* mutant seeds were imaged by SEM. The left column shows close-ups of individual MSCs. All samples except for *shp1 shp2*, have wildtype appearance, with normal formation of the columellae observed in the MSC. In some instances, *shp1 shp2* have a less prominent columella relative to wildtype.

Scale bar = 50 μm, n = 5.

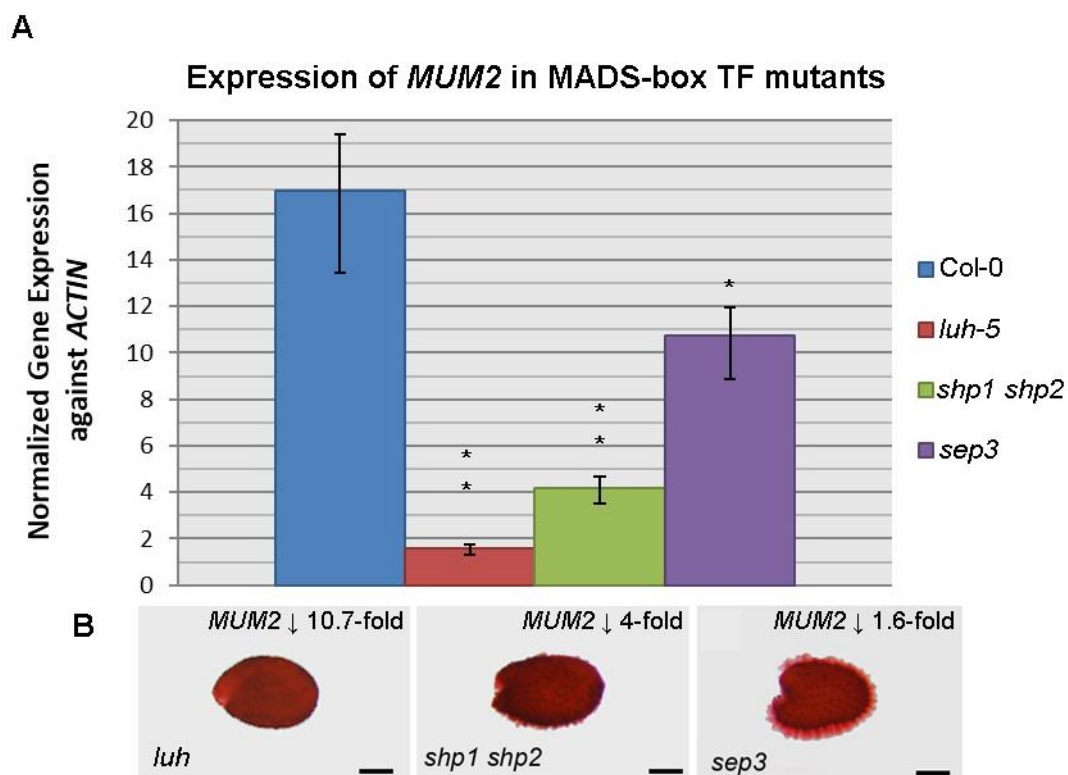


Figure 4.3. Expression of *MUM2* in MADS-box TF mutants.

A: Expression of *MUM2* in Col-0, *luh-5*, *shp1 shp2* and *sep3* mutant whole seeds at linear cotyledon stage (7 DPA) was examined by qRT-PCR. Error bars represent +/- standard error (S.E) of mean for three biological replicates and statistical significance was calculated using Student's unpaired *t*-test (* $p < 0.05$ & ** $p < 0.001$).

B: Panel of mucilage mutants *luh-5*, *shp1 shp2* and *sep3*, with fold change of *MUM2* expression relative to Col-0 indicated.

Scale bar = 200 μ m.

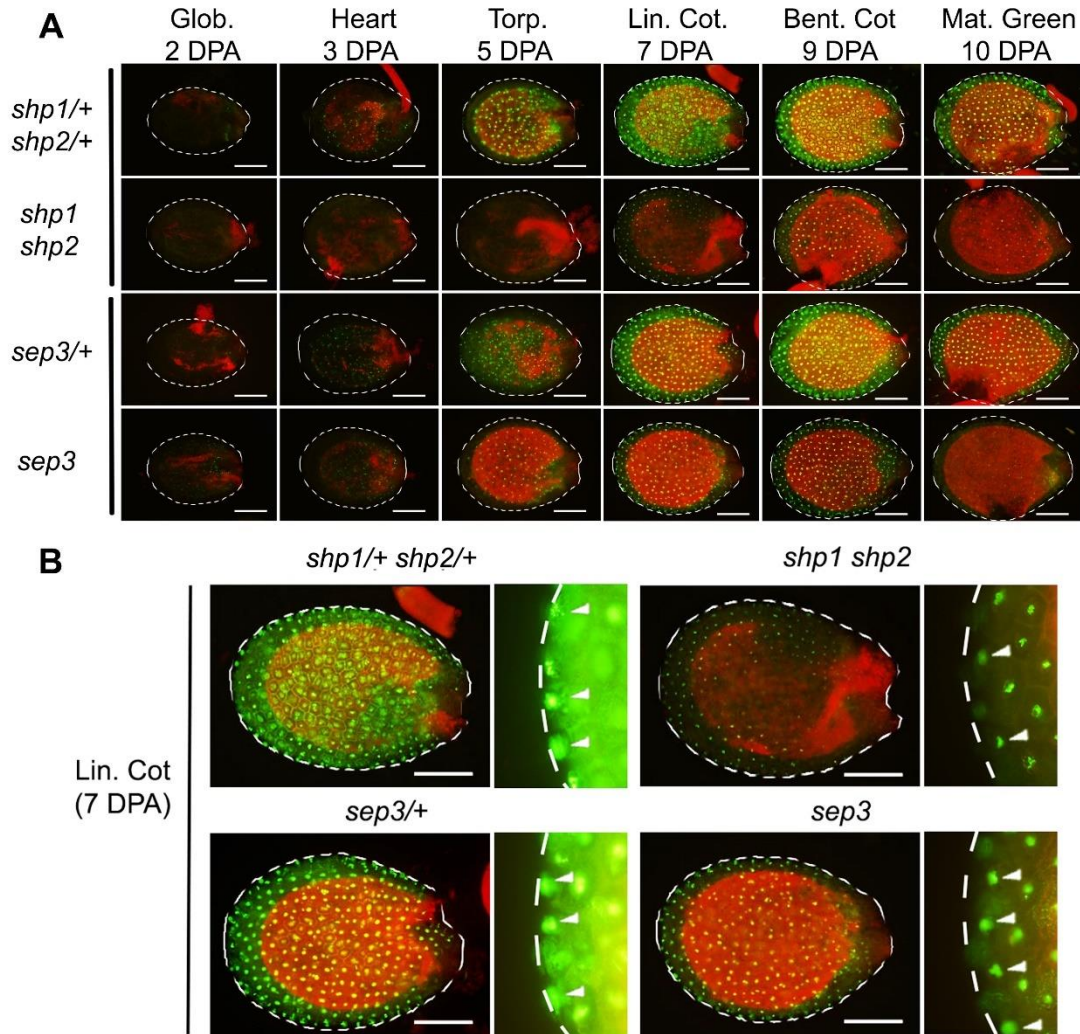


Figure 4.4. *MUM2* promoter I1-*VENUSn* reporter activity in seed coats of phenotypically wildtype, *shp1 shp2* and *sep3* mutants.

MUM2 promoter I1-*VENUSn* reporter constructs were transformed into *shp1 shp2* and *sep3* backgrounds independently and a representative T_1 plant crossed to Col-0 and either *shp1 shp2* or *sep3* so that comparisons could be made between the same transgene in two different backgrounds.

A: The expression of the *MUM2*_{proI1}::*VENUSn* reporter in early to late stages of *shp1/+ shp2/+*, *shp1 shp2*, *sep3/+* and *sep3* seeds. Dotted lines indicated the boundary of MSCs.

B: Enlarged image of *MUM2_{proI1}::VENUSn* in MADS-box TF mutants at the linear cotyledon stage. White arrows indicate VENUSn activity in MSCs, while dotted lines indicated the boundary of MSCs.

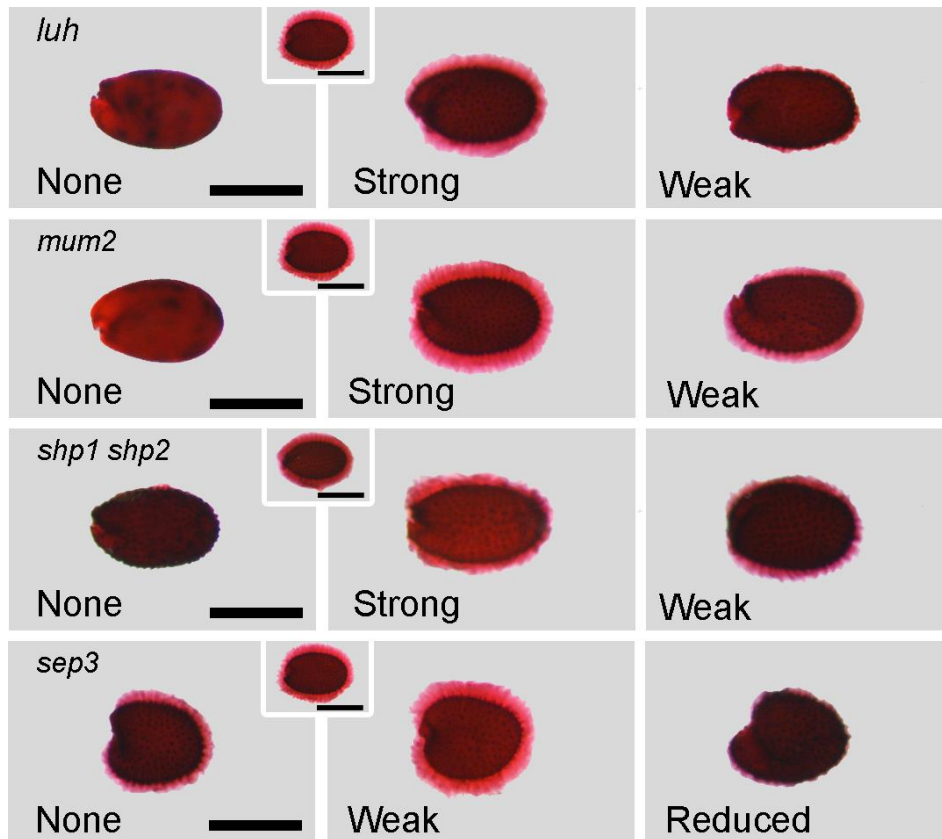


Figure 4.5. Mucilage release phenotypes of seeds derived from transgenic MADS-box TF mutants with elevated levels of *MUM2* or *LUH*.

Examples of complementation phenotypes (strong, weak, none and reduced) of *shp1 shp2*, *sep3* and respective transformation controls as noted in Table 4.2 are shown where applicable, with Col-0 as wildtype control shown in the inset.

Scale bar = 500 μ m.

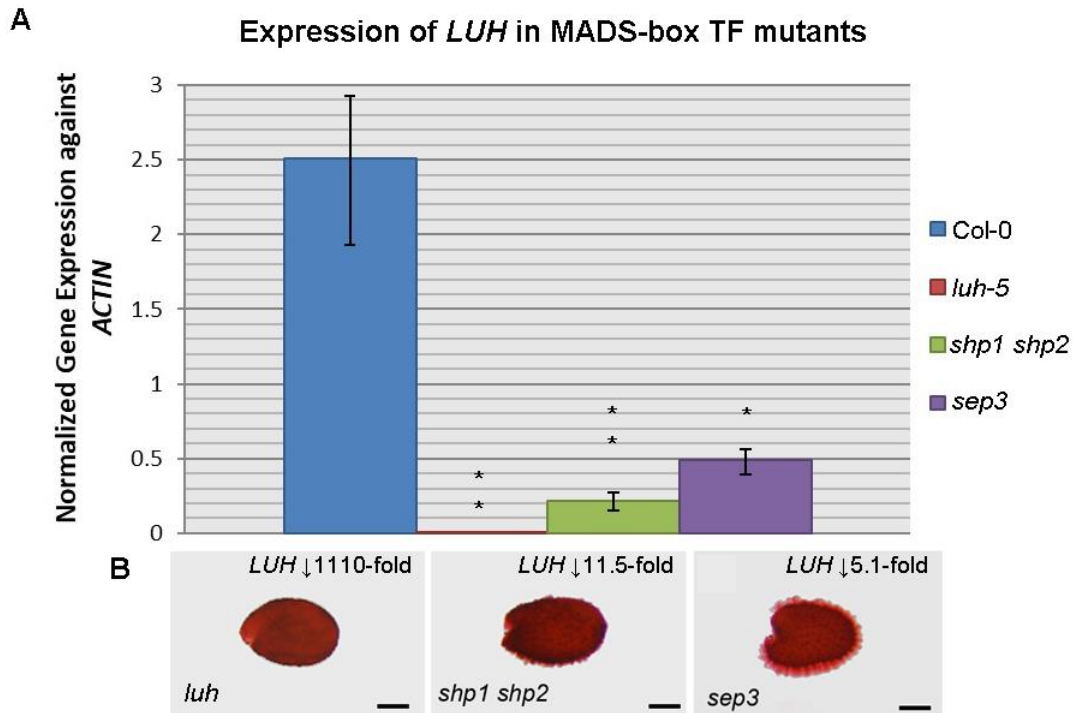


Figure 4.6. Expression of *LUH* in MADS-box TF mutants.

A: Expression of *LUH* in Col-0, *luh-5* (as a control for *LUH* expression), *shp1 shp2* and *sep3* mutant whole seeds at linear cotyledon stage was examined by qRT-PCR. Error bars represent $\pm$ S.E for three biological replicates and statistical significance was calculated using Student's unpaired *t*-test (* $p < 0.05$ & ** $p < 0.001$).

B: Panel of mucilage mutants *luh-5*, *shp1 shp2* and *sep3*, with fold change of *LUH* expression relative to Col-0 indicated.

Scale bar = 200 μ m.

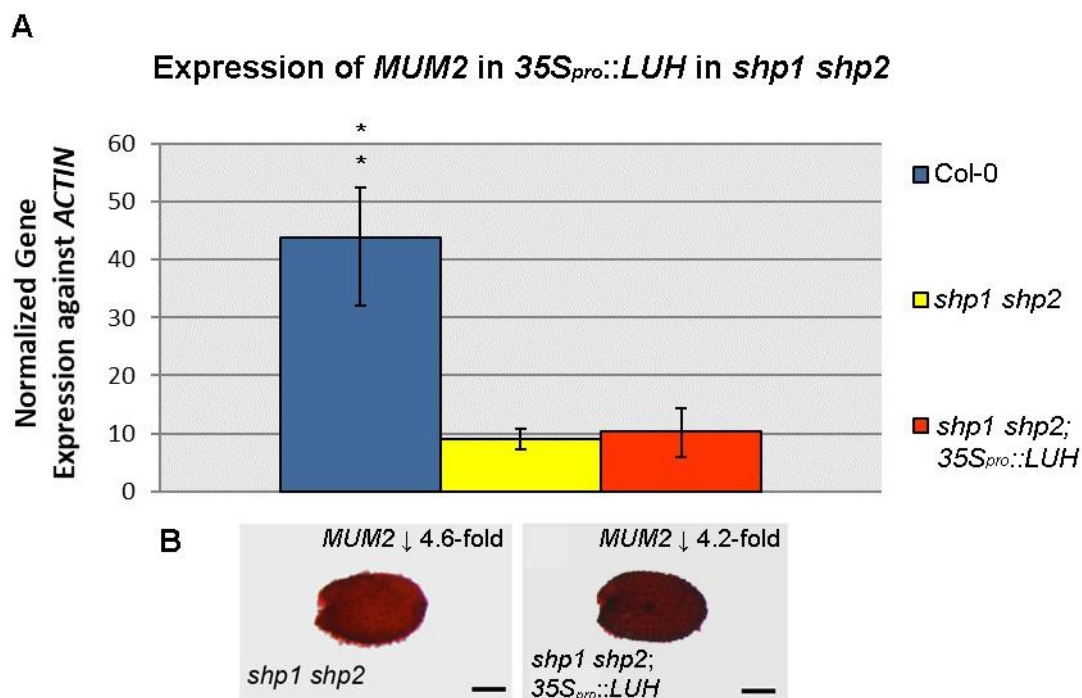


Figure 4.7. Expression of *MUM2* in *shp1 shp2* lines carrying $35S_{pro}::LUH$

A: Expression of *MUM2* in whole seeds derived from Col-0, *shp1 shp2* and a representative $35S_{pro}::LUH$ *shp1 shp2* line at the linear cotyledon stage was examined by qRT-PCR. Error bars represent $\pm$ -S.E for three biological replicates and statistical significance was calculated using Student's unpaired *t*-test (** $p < 0.001$).

B: Panel of mucilage mutants *shp1 shp2* and $35S_{pro}::LUH$ in *shp1 shp2*, with fold change of *MUM2* expression relative to Col-0 indicated.

Scale bar = 200 μ m.

Chapter 5

Defining the mechanism by which MADS-box transcription factors regulate *MUM2*

Defining the mechanism by which MADS-box transcription factors regulate *MUM2*

5.1 Introduction

Although **Chapter 4** established that SHP1, SHP2 and SEP3 are positive regulators of *MUM2*, the mechanism of regulation (direct vs indirect; Fig. 1.9A & B, respectively) was not determined. Thus, this chapter seeks to address this by investigating the nature of the *MUM2* regulatory mechanism and how it relates to LUH. A key characteristic of MADS-box TFs is that they bind to conserved sequences called CArG box motifs, with canonical sequences defined as CC(A/T)₆GG (Tang & Perry, 2003). These motifs typically arise in pairs within 250 bp of each other and bind tetrameric MADS-box complexes, resulting in the looping of DNA and the recruitment of additional regulatory elements (Fig. 1.13.) (Smaczniak et al., 2012; Theissen et al., 2016; Theissen & Saedler, 2001). Therefore, a simple way to determine if MADS-box TFs may directly regulate *MUM2* is to look for the presence of CArG box motifs within the regulatory sequence of this gene. A key expectation is that at least two CArG box motifs should be present within a 250 bp of each other if they are to be considered functional (Smaczniak et al., 2012; Theissen et al., 2016; Theissen & Saedler, 2001). Previous work by Dr Nadeeka Jayawardana, a former PhD student in the Golz group, established that the 1.8 kb first intron of *MUM2* (*I1*) along with the promoter is necessary for robust reporter activity in the MSCs of the seed coat, and that deletion of a 200 bp region at the 5' end of *I1* (referred to as region A) abolishes expression in the MSCs (Jayawardana, 2016). Subsequent computational analysis of region A identified several putative CArG box motifs and that removal of these elements affects the activity of the *MUM2* promoter/intron-reporter in MSCs. Thus, this analysis indicates that MADS-box TFs may regulate *MUM2* through these elements.

The first section of this chapter describes how recently published binding site data for the MADS-box TFs AP1 and SEP3 identified new CArG box and CArG box-like

(CL) binding sites within region A of the *MUM2* I1. In this study, these CArG box/CL motifs are inactivated to investigate their roles on *MUM2* promoter/intron-reporter activity in the seed coat. An important consideration for these experiments is whether the inactivation of the elements should be done through either deletions or by simply mutating the motifs. Of these two approaches, deletions have the potential to alter activity of nearby regulatory elements due to their removal (if they overlap with the motif being removed), or by changing their DNA conformation/spacing relative to other elements. Given the potential complications arising from the deletion approach, both approaches are investigated in this chapter. In parallel, the transcriptional activity of 107 reported members of the MADS-box family in *Arabidopsis* (Parenicova et al., 2003) is assessed using a publicly available microarray dataset of genes expressed in specific regions and sub-compartments of the seed (Belmonte et al., 2013). This enables the identification of MADS-box TFs that are expressed in the seed coat and thus potentially involved in *MUM2* regulation.

There are several different approaches that can be taken to provide additional experimental support for SHP1, SHP2 and SEP3 functioning as *MUM2* activators. The first involves the use of chimeric repressor silencing technology to convert an activating TF into a dominant repressor by fusing it to the repressive EAR motif (Ethylene-responsive element binding factor-associated amphiphilic repression, also known as SRDX). When such a chimeric TF is introduced into plants, the resulting dominant negative activity induces a strong loss of function phenotype by compromising the activity of matching endogenous TF(s), as well as closely related redundantly-acting TFs (Hiratsu et al., 2003). For example, the SRDX motif fused downstream of CUP-SHAPED COTYLEDON1 (*CUC1*) represses both *CUC1* and *CUC2* to produce cup-shaped cotyledons in wildtype plants, while MYB23-SRDX reduced both *MYB23* and *GL2*, resulting in leaves with less trichomes (Hiratsu et al., 2003). This feature is of particular importance for the MADS-box TFs as they function as tetramers and display a high degree of functional redundancy (Smaczniak et al., 2012; Theissen, 2001; Theissen & Saedler, 2001). It is therefore expected that fusion of an SRDX motif to SHP2, for instance, will not only reduce activity of the

endogenous SHP2 but also redundantly acting SHP1. In addition, the chimeric SHP2 may also inactivate other MADS-box TFs that are part of the SHP2 complex (e.g. SEP3, Immink et al., 2009). Thus, the use of the chimeric repressor silencing technology not only has the potential to show that the MADS-box TFs function as activators but may also produce a phenotype that is more pronounced than *shp1 shp2* due to multiple MADS-box TFs being inactivated.

Another approach that can be used to investigate SHP1, SHP2 and SEP3 activity is a transactivation assay. Transactivation assays are routinely used to evaluate the responses of particular promoters to TFs when expressed transiently in heterologous systems, such as in the study of relative promoter strengths (Hellens et al., 2005; Hernandez-Garcia et al., 2009; Nassuth et al., 2014). The transient transactivation system developed by Hellens (2005) is attractive because it uses both a firefly luciferase (FLUC) reporter gene to provide a read out of promoter activity, as well as a Renilla luciferase (RLUC) reporter gene under the control of the constitutive promoter CaMV 35S_{pro}, as an internal control that provides a measure of transformation efficiency (Hellens et al., 2005). By calculating the ratio of FLUC to RLUC activity, promoter activity is able to be expressed relative to the transformational efficiency and thus provides a normalised value that enables direct comparison between different experiments. This system is therefore ideal to test whether SHP1, SHP2 and SEP3 act directly on the *MUM2* promoter/intron and whether this activity is dependent on CArG box/CL motifs that are present in region A.

Lastly, as previous work established that LUG physically interacts with SEP3 via SEU (Sridhar et al., 2006), it is possible that SEP3, SHP1 and SHP2 may also be part of the LUH regulatory complex. This is especially relevant given that single *shp2*, double *shp1 shp2* and triple *shp1 shp2 sep3* mutants have increasingly severe mucilage extrusion defects (**Chapter 3, Section 3.2.4**), the latter of which are reminiscent of *mum2*. To investigate this possibility, yeast two-hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) assays were performed. While

Y2H allows for the rapid screening of interactions between proteins, it is a heterologous system that suffers from a high rate of false positives (Hengen, 1997; Xing et al., 2016). To address this concern, potential interactions can also be tested *in planta* using the BiFC assay. Although this approach also has the advantage of being both simple and rapid, it can still produce false positives as fluorescent protein fragments fused to the proteins being tested can self-assemble independently of protein-protein interactions (Horstman et al., 2014). Despite these issues, combined use of both Y2H and BiFC is considered a robust way to investigate potential protein-protein interactions.

5.2 Results

5.2.1 Identification of CArG box and CL motifs within the regulatory regions of the *MUM2* intron

Previous work has suggested that two CArG box motifs, CArG I and II, located within the 5' region of *II* function as activators because their removal from the *MUM2* promoter/intron-reporter by deletion resulted in reduced activity in the seed coat (Jayawardana, 2016) (Fig. 5.1). To confirm that these previously identified CArG box motifs are bound by MADS-box TFs, published AP1 and SEP3 ChIP-seq data from floral tissue was examined (Kaufmann et al., 2009; Pajoro et al., 2014). Surprisingly, in both instances, the TFs bound to sequences other than CArG I and II, suggesting either the presence of other CArG box elements in the intron or alternatively, the lack of binding resolution in these datasets. Further analysis of these ChIP-seq data identified additional CArG box and CL motifs that have been annotated as CArG III and CL I and CL II in Fig. 5.1, with SEP3 binding to a both CArG III and CL I (Kaufmann et al., 2009; Pajoro et al., 2014) and AP1 binding to CL II (Pajoro et al., 2014).

Next, the significance of these CArG box elements for *MUM2* regulation was investigated by generating *MUM2* promoter/intron-reporter constructs lacking a combination of these CArG box motifs and their activity subsequently assayed in both Col-0 and *luh* backgrounds. The first approach described in **Section 5.2.1.1**, investigates the consequences of removing these elements from the *MUM2* promoter/intron-reporter using deletions. In the second approach described in **Section 5.2.1.2**, the CArG box motifs were not removed but mutated such that they are no longer functional. This approach was taken because it is possible that deletions in an important regulatory region may either inadvertently remove adjacent motifs or alter their spacing and hence their activity.

5.2.1.1 Analysis of CArG box, CL and MYB binding site deletions

CArG I and II deletions previously generated affected activity of the promoter/intron-reporter (*MUM2_{pro}+I1*) in the seed coat. To determine whether the CArG box motifs CArG III, CL I and CL II are also required for *MUM2* activity in the seed coat, deletion analysis was extended to these elements (Fig. 5.1). Given that MADS-box TFs are thought to bind to two CArG motifs concurrently, it was decided that removing multiple motifs simultaneously would more likely generate noticeable changes in *MUM2* reporter activity. However, computation analysis revealed the presence of two putative MYB binding sites within region A with one (MYB II) overlapping with the CArG III and CL II motifs (Jayawardana, 2016; Fig 5.1). This meant that deletions of CArG III and CL II inadvertently removed the MYB II binding site (Table 5.1).

Deletion constructs of CArG box, CL motifs and MYB binding sites in region A were generated using gBlocks as outlined in **Chapter 2, Section 2.2**. A list of plasmids associated with the construction of the various TF binding sites deletions is listed in Appendix 5, Supplementary Table 1. In addition, the full length *MUM2* promoter/intron-reporter constructs in both Col-0 and *luh-5* (Jayawardana, 2016) were included as controls to facilitate comparison to the TF binding site deletion constructs (indicated with an asterisk in Table 5.2).

A total of 16 independent lines per background were generated for each deletion construct, with VENUSn fluorescence in the developing seed coats scored according to strength to facilitate comparisons across backgrounds (Table 5.1). Deletion of both CArG III and MYB II motifs (*MUM2_{pro}+I1^{ΔCArGIII,MYBII}*; double deletion) in region A decreased fluorescence intensity from early (3 DPA) to late (10 DPA) stages in 3/16 lines (19%) of both Col-0 and *luh-5* developing seed coats, with the remaining 13/16 lines (81%) lines having very low detectable fluorescence (Table 5.1, Fig. 5.2). These results suggest that either CArG III, MYB II or both elements are required for *MUM2* activity in the seed coat. Examination of developing embryos in lines displaying some fluorescence in both Col-0 and *luh-5*, revealed decreases in fluorescence activity

compared to *MUM2_{pro}+I1* from early stages (3 DPA) (Appendix 5, Supplementary Fig. 1), indicating that these putative *MUM2* activator elements are also likely to function in the embryo (Table 5.2)

Surprisingly, Col-0 and *luh-5* lines transformed with a *MUM2_{pro}+I1^{ΔCArGIII,CLII,MYBII}* reporter construct (triple deletion) produced seeds that had increased fluorescence relative to the double deletion reporter, but with levels reduced compared to the intact *MUM2_{pro}+I1* reporter. In addition, VENUSn fluorescence was also present in the epidermal layer of *luh-5* seeds from the linear cotyledon stage (7 DPA) onwards (Fig. 5.2 & 5.3). This contrasts with the *MUM2_{pro}+I1* reporter which is active in the inner, but not MSC layer of the *luh-5* seed coat (Fig. 5.2 & 5.3). Increased fluorescence activity in the seed coat of the triple deletion reporter lines compared to the double deletion reporter lines was observed in both backgrounds; only the triple deletion, but not the double deletion construct, produced any lines with a strong to intermediate level of fluorescence in both Col-0 (8/16; 50%) and *luh-5* seed coats (9/16; 56%), respectively (Table 5.1). The key difference between the double and triple motif deletions is the additional deletion of CL II, suggesting that its removal is responsible for the overall increase in VENUSn expression and the expansion of VENUSn expression in *luh-5* seeds to the MSCs (Fig. 5.3). This observation is consistent with CL II being a repressor element that is bound by the hypothesised *MUM2* repressor, presumably active from 7 DPA onwards. Therefore, removal of CL II renders the reporter construct sufficiently insensitive to the *MUM2* repressor and partially restores VENUSn expression in the MSC of *luh* seeds (Fig. 5.3). Lastly, examination of embryos in seeds derived from triple deletion reporter lines in both Col-0 and *luh-5* backgrounds revealed a similar level of fluorescence when compared to double deletion lines in embryos (Appendix 5, Supplementary Fig. 1), suggesting that loss of CL II has no effect in the embryo (Table 5.2).

Examining the *MUM2_{pro}+I1^{ΔCArGIII,CLI,CLII,MYBII}* reporter (quadruple deletion) in both Col-0 and *luh-5* backgrounds revealed broadly similar expression levels to those of the triple deletion reporter, but with a noticeable loss of fluorescence in the MSC layer

of *luh-5* from the linear cotyledon stage onwards (Fig. 5.2 & 5.3). Similarly, fluorescence activity of the quadruple deletion reporter in developing embryos was similar to the triple deletion reporter (Appendix 5, Supplementary Fig. 1). Taken together, a reasonable explanation for these observations is that the CL I motif is bound by a putative activator necessary for the expression of *MUM2* in the MSC layer but when absent does not alter activity in the embryo (Table 5.2).

Lastly, to control for the effects of deleting the MYB II binding site, both MYB I and MYB II motifs were removed leaving CArG box and CL motifs intact (Fig. 5.2 & 5.3). This analysis showed that both MYB binding sites are necessary for *MUM2* activity in the seed as there was a decrease of VENUSn fluorescence in Col-0 that is comparable to both the triple and quadruple deletions and a complete loss of fluorescence in *luh-5* seeds (Table 5.1, Fig. 5.2). Likewise, the deletion of MYB motifs resulted in a drastic decrease in VENUSn signal in Col-0 compared to *MUM2_{pro+I1}*, with a complete loss of fluorescence in *luh-5* lines (Appendix 5, Supplementary Fig. 1). This finding shows that MYB binding sites are required for *MUM2* activity in the seed coat and embryo (Table 5.2). As such, removal of the MYB II motif in the double, triple and quadruple deletion constructs complicates the interpretation of CArG box and CL motifs in *MUM2* regulation. Because of this, the impact of CArG III and CL II on *MUM2* activity using a mutational approach is examined in the next section.

5.2.1.2 Analysis of CArG box and CL motif mutations

Selected CArG box and CL motifs (shaded in purple, Table 5.2) were mutated by converting the internal A/T sequences into G/C sequences (Mendes et al., 2013). A key strength of this approach is the ability to target CArG box and CL motifs without affecting the sequence of the overlapping MYB II binding site (Fig. 5.1). Based on deletion analysis (Table 5.2), CArG I, II, III and CL I have been inferred to act as *MUM2* activator elements, whereas CL II would appear to function as a *MUM2*

repressor element to which the hypothesised *MUM2* repressor binds. Two sets of experiments were performed to assess how the behaviour of the *MUM2* promoter/intron-reporter is altered following mutations in the CArG box and CL elements. In the first experiment, CL II was mutated to determine whether this results in VENUSn activity being detected in the MSCs of the *luh* seed coat as suggested by the deletion counterpart (Fig. 5.2 & 5.3). In the second set of experiments, mutations in CArG I, II and III were combined so as to determine whether they work co-operatively as *MUM2* activators.

Constructs harbouring mutations in the CArG box and CL motifs of region A were generated by seamless cloning of gBlocks. A list of plasmids associated with the making of the various CArG box and CL mutations can be found in Appendix 5, Supplementary Table 1. The constructs were then introduced into both Col-0 and *luh-5* backgrounds using the floral dip transformation procedure and transformed lines identified by Basta screening. Sixteen independent lines per background were selected for analysis and categorised according to fluorescence strength (Table 5.3). In addition, to facilitate direct comparison of promoter-reporter activity between a Col-0 and *luh-5* background, a representative *luh-5* line for each construct was subsequently crossed to *luh* and Col-0 to generate either *luh* homozygous mutants or heterozygous (*luh/+*) lines each with exactly the same promoter-reporter (Fig. 5.4 & 5.5). This comparison eliminates variation in promoter-reporter activity that arises from differences in insertion sites.

Interestingly, mutations in CL II resulted in reduced fluorescence activity in both Col-0 and *luh-5* backgrounds at all developmental stages (2-10 DPA), with the reduction being further exacerbated in the *luh-5* background (Fig. 5.4 & 5.5). Changes in fluorescence were consistently observed with 14/16 (87.5%) lines in *luh/+* and 12/16 (75%) lines in *luh-5* displaying strong to intermediate fluorescence while the remainder had either weak or no fluorescence (Table 5.3). Examination of the CL II mutation in developing embryos in *luh/+* and *luh-5* revealed a decrease in overall fluorescence activity, indicating that the TFs binding to this motif are active in the

embryo (Appendix 5, Supplementary Fig. 2). These results contradict those obtained from the deletion experiments mentioned above (**Section 5.2.1.1**) and suggests that CL II is a *MUM2* activator element in both seed coat and embryo (Table 5.2). The difference between the deletions and mutation experimental data may stem from disturbances arising from the removal of sequences in region A, which lead to the spacing of other regulatory elements being altered.

Mutations in CArG I, II and III resulted in decreased VENUSn fluorescence in both *luh/+* and *luh-5* backgrounds compared to the full length *MUM2* promoter/intron-reporter (Table 5.3, Fig. 5.4 & 5.5). The strength of fluorescence in *luh/+* was slightly stronger than *luh-5*; 3/16 (19%) and 2/16 (12%) lines produced intermediate strength fluorescence, 10/16 (62%) and 7/16 (44%) lines with weak fluorescence and 3/16 (19%) and 7/16 (44%) with no fluorescence in *luh/+* and *luh-5* backgrounds respectively (Table 5.3). The triple mutation also appears to produce weaker fluorescence signal in the seed coats compared to the single mutation of CL II, as expected with the inactivation of the increased number of CArG box motifs (Table 5.3, Fig. 5.4 & 5.5). Similar to the deletion constructs, fluorescence activity in the corresponding developing embryos appears weaker to the full length *MUM2* promoter/intron-reporter, indicating that TFs binding to these motifs are also active in the embryo (Appendix 5, Supplementary Fig. 2). Collectively, these observations are consistent with CArG I, II and III acting as putative *MUM2* activator elements that are active in both the seed coat and embryo (Table 5.2).

Given that CL II appears to function as a *MUM2* activator element, a further CArG I, II, III and CL II mutation (quadruple mutation) reporter construct was generated and introduced into the *luh/+* and *luh-5* backgrounds. As expected from the inactivation of a putative *MUM2* activation site, there was a marked decrease in VENUSn fluorescence in both *luh/+* and *luh-5*, with either no (13/16 lines; 81%) or weak (3/16 lines; 19%) fluorescence observed in both backgrounds (Table 5.3). The quadruple mutation produced such weak fluorescence signal in both *luh/+* and *luh-5* that it was necessary to increase the exposure 4-fold (ISO 800) to enable fluorescence to be

compared to both single and triple mutations (ISO 200) (Fig. 5.4 & 5.5). Subsequently, to enable comparisons across all CArG box and CL mutation sets, full length *MUM2_{pro}+I1* control, single and triple mutation images were processed post-capture using Adobe Photoshop (ver. 12.0) and their overall exposure was increased 4-fold (Fig. 5.4 & 5.5). Comparisons between the quadruple mutation against both single and triple mutations reveal that inactivation of these four motifs severely diminishes *VENUSn* activity, and by extension *MUM2* expression, in both seed coat (Fig. 5.4, 5.5) and embryo (Appendix 5, Supplementary Fig. 2). Thus, this supports their role as putative *MUM2* activator sites in the seed coat and embryo (Table 5.2).

Collectively, the results of the mutation of CArG box and CL motifs largely corroborate findings from the deletions of CArG box and CL motifs, with these motifs acting as putative *MUM2* activator sites (Table 5.2). However, while deletion analysis indicated CL II was a *MUM2* repressor binding site active only in the seed coat, mutagenesis analysis showed that it acts as a *MUM2* activation site in both seed coat and embryo (Table 5.2).

5.2.2 Candidate MADS-box TFs binding partners to CArG box and CL motifs in the *MUM2* intron

Having established in **Section 5.2.1** that region A contains numerous CArG box elements that function as activation sites of *MUM2* expression, a bioinformatics approach was used to generate a list of potential MADS-box TFs that may bind to these motifs. Given that *MUM2* is expressed in the seed coat, potential MADS-box TFs that may bind to these CArG box and CL motifs must be expressed in the seed coat. Further evidence supporting this is the change in *VENUSn* expression within the seed coat, but not the embryo, when CArG I, II and III motifs and CLII are mutated, therefore indicating that participating MADS-box TFs must be present in the seed coat (**Section 5.2.1.2**). Therefore, expression patterns for the 107 reported members of the MADS-box family in *Arabidopsis* (Parenicova et al., 2003) were examined using

a publicly available microarray dataset (Belmonte et al., 2013), to specifically look for MADS-box TFs that are expressed in, but not limited to the seed coat. As noted in **Chapter 3, Section 3.2.1**, this extensive microarray dataset was derived from seed samples processed by laser capture microdissection and includes seven sub-regions of the wildtype seed (embryo proper, suspensor, micropylar endosperm, peripheral endosperm, chalazal endosperm, chalazal seed coat and general seed coat) across five different seed developmental stages (pre-globular, globular, heart, linear cotyledon and mature green) and is referred to as the Belmonte dataset (Belmonte et al., 2013).

Expression of a total of 86 out of the 107 MADS-box TFs was detected in the Belmonte dataset and the expression values extracted for both embryo (E) and seed coat (SC) tissues across early to late developmental stages: Pre-Globular (PG); Globular (G); Heart (H); Linear Cotyledon (LC); Mature Green (MG). Next, only MADS-box TFs present in, but not limited to the seed coat (denoted as a PP in the Belmonte dataset) were further investigated. This yielded a total of 10 MADS-box TFs, and their expression values were then used to generate a heat map using the graphical tool Protovis (Bostock & Heer, 2009), which enabled the visualisation of these MADS-box TF across developmental stages in both embryo and seed coat tissues (Fig. 5.6). Interestingly, the majority of the MADS-box TFs with seed coat expression were also identified in **Chapter 3** as differentially expressed in *luh* relative to Col-0 whole seeds with the exception of *SEP2*, *PHERES1* (*PHE1*) and *AGAMOUS-LIKE18* (*AGL18*) (Fig. 5.6).

Overall, this analysis identified 10 MADS-box TFs as potential candidates for being *MUM2* regulators in the seed coat, with work described in other chapters already confirming that *SHP1*, *SHP2* and *SEP3* perform this function. For example, mutants lacking these TFs are associated with mucilage defects (**Chapter 3, Section 3.2.4**) and decreased *MUM2* expression (**Chapter 4, Section 4.2.3**).

5.2.3 SHP2 regulates *MUM2* via CArG box/CL motifs in *I1*

Although SHP1, SHP2 and SEP3 are strong candidates for binding to the CArG box and CL motifs identified in **Section 5.2.1**, time constraints meant that it was not possible to thoroughly investigate the capacity of all three TFs to regulate *MUM2*. Therefore, experiments described in this section focus on SHP2, as several lines of evidence suggest that this MADS-box TF was likely to be a potent *MUM2* regulator. For instance, SHP2 has high expression in the seed coat, and *shp1 shp2* mutants have a marked decrease in *MUM2* expression which is attributable mainly to the loss of SHP2 activity as *shp2*, but not *shp1*, mutants display mucilage extrusion defects (**Chapter 3, Section 3.2.4**).

5.2.3.1 SHP2-SRDX fusions induce mucilage extrusion defects in transgenic lines

As mentioned in the introduction to this chapter, chimeric repressor silencing technology can be used to convert a TF from an activator into a repressor through fusions to the SRDX motif. Expression of these chimeric repressors in plants results in a dominant negative effect whereby activity of the homologous endogenous TF and closely related and redundantly acting TFs are suppressed (Hiratsu et al., 2003). A complicating factor in applying this approach to SHP2 is that this TF also regulates ovule formation and the development of dehiscence zones in maturing siliques, in addition to mucilage extrusion (Ehlers et al., 2016; Liljegren et al., 2000). Strong loss of function phenotypes arising from the introduction of the SHP2-SRDX may therefore not only affect mucilage extrusion, but other developmental processes.

Despite this possibility, a strong constitutive promoter was used to express SHP2-SRDX to observe the effects of its conversion to a repressor. The *35S_{pro}::SHP2::SRDX* construct was generated by Gateway recombination, and transformed to Col-0 plants via floral dip transformation. A list of primers and plasmids associated with the making of this construct is provided in Appendix 5, Supplementary Table 1 & 2. Transgenic plants were identified following Basta screening and seeds derived from

the resulting 36 T₁ lines assayed for mucilage defects following Ruthenium Red staining.

Seed staining analysis revealed a range of mucilage defects amongst *35S_{pro}::SHP2::SRDX* transformants, with 5/36 (14%) lines producing no mucilage release, 10/36 (28%) lines with weak mucilage release (*shp2* phenocopy), 15/36 (41%) lines with moderate mucilage release and 6/36 (17%) lines with strong (wildtype-like) mucilage release (Fig. 5.7). Additionally, poor seed set was observed in seeds which produced little mucilage extrusion, while lines with moderate and strong mucilage release were unaffected. In addition to these 36 lines, several other lines displayed severe (split carpels, no petals) to weak (shortened stamens) floral phenotypes and were therefore unable to produce seeds (Fig. 5.8). Collectively, the observed mucilage and floral defects reflect dominant negative phenotypes that are consistent with the conversion of SHP2 from an activator to a repressor.

Next, quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was performed on transgenic lines containing *35S_{pro}::SHP2::SRDX* to determine if poor mucilage extrusion was associated with a decrease in *MUM2* expression. Three independent transgenic lines with weak mucilage release were selected and sown, along with three biological replicates of *shp1 shp2* double mutants and three independent transgenic lines containing *35S_{pro}::SHP2* in a Col-0 background as mutant and wildtype controls, respectively. The use of *35S_{pro}::SHP2* as a control enables the potential for co-suppression associated with a *35S_{pro}::SHP2* transgene to be assessed.

Whole seeds at the linear cotyledon stage (7 DPA) were collected as this stage coincides with the onset of *MUM2* expression in the MSCs (Jayawardana, 2016). *MUM2* transcripts were normalised against *ACTIN2* (*AT3G18780*, *ACT2*) and a list of qRT-PCR primers used for this analysis is provided in Appendix 5, Supplementary Table 3. To allow for statistical analysis, the three independent transgenic *35S_{pro}::SHP2::SRDX* and *35S_{pro}::SHP2* lines were treated as biological replicates

respectively. This analysis revealed a 4.5-fold decrease in *MUM2* in the *shp1 shp2* control ($p < 0.001$) and a 2.3-fold decrease in *MUM2* in *35S_{pro}::SHP2::SRDX* ($p < 0.05$) when compared against the wildtype control (Fig. 5.9). No significant difference between either *MUM2* levels (Fig. 5.9) or mucilage extrusion phenotype (data not shown) was observed in *35S_{pro}::SHP2* and Col-0, therefore excluding the possibility of potential co-suppression associated with the *35S_{pro}::SHP2* transgene.

Taken together, the dominant negative phenotypes and the decrease in *MUM2* expression in lines transformed with *SHP2::SRDX* provides compelling support for SHP2 acting as a *MUM2* activator in *Arabidopsis*.

5.2.3.2 MADS-box TFs activate *MUM2* in transactivation assays

Transactivation assays were next used to determine whether SHP2 functions as a *MUM2* activator. This technique involves the fusing of a promoter of interest to a luciferase reporter gene and co-infiltrating this construct with another construct containing a TF of interest into the same *N. benthamiana* leaf for examination at 3-6 days post infiltration (DPI) (Cazzonelli & Velten, 2006; Hellens et al., 2005). The *Agrobacterium*-mediated transient transactivation system developed by Hellens (2005) is an attractive system due to its ability to provide a quantitative measurement of promoter and TF activity through the use of dual luciferase reporters. Specifically, pGREENII cloning vectors that were developed to carry both FLUC reporter genes for expression under a target promoter, and RLUC reporter genes for expression under a control promoter such as the Cauliflower mosaic virus promoter (CaMV *35S_{pro}*), allows for the normalised quantification of promoter activity in *N. benthamiana* leaves, which is expressed as a ratio of FLUC/RLUC (Hellens et al., 2005). The normalised expression of FLUC against RLUC is important as RLUC acts as a loading/transformation control, accounting for variations in FLUC output associated with transformation efficiencies in *N. benthamiana* leaves, their collection and

processing (Hellens et al., 2005; Nassuth et al., 2014). Normalised FLUC/RLUC ratios thereby allow direct comparison between different combinations of co-infiltrated promoters and TFs (Hellens et al., 2005; Nassuth et al., 2014). The pGREENII vector, together with the commercially available dual luciferase assay system, can therefore be used to determine if MADS-box TFs such as SHP2 can bind to target promoters such as *MUM2*, including appropriate regulatory elements such as the first intron of *MUM2* (*MUM2 I1*). Successful binding of SHP2 to the *MUM2* promoter and/or intron elements is expected to produce an increase in its FLUC/RLUC ratio compared to a paired experiment lacking SHP2. In addition, mutations of CArG box motifs in *MUM2 I1* can be included in this system to test whether SHP2 activates *MUM2* by binding to these motifs.

Having presented evidence in **Section 5.2.3.1** that SHP2 acts as a *MUM2* activator *in planta*, dual luciferase transactivation assays were next used to investigate the extent of *MUM2* activation by SHP1, SHP2 and SEP3. This involves fusing the promoter of interest to a luciferase reporter gene and introducing this reporter construct along with another construct containing a TF that is thought to target this promoter into *N. benthamiana* leaves (Cazzonelli & Velten, 2006; Hellens et al., 2005). An *Agrobacterium*-mediated transient transactivation system developed by Hellens (2005) as described above was adopted for these experiments. Some changes were made to the vectors used for these experiments to improve consistency of the assay. These include the reversal of the internal control *35S_{pro}::RLUC* cassette relative to the FLUC expression cassette such that they are divergently transcribed (Fig. 5.10). All reporter constructs used in this transactivation assay were generated in binary vector pGREENII. A list of associated primers and plasmids is found in Appendix 5, Supplementary Table 4 & 5. Collection and processing of infiltrated leaves were performed at 6 DPI in triplicates and is described in detail in **Chapter 2, Section 2.7**.

Co-infiltration of the *MYB75* promoter driving *FLUC* (*MYB75_{pro}*) and the *35S* promoter driving *FILAMENTOUS FLOWER* (*FIL*) was used as a positive control for these experiments, as *FIL* has been shown to be a direct activator of *MYB75_{pro}* (Boter

et al., 2015). A 1.7-fold increase in normalised FLUC/RLUC expression from co-infiltrations of *MYB75_{pro}* and *FIL* compared to baseline FLUC/RLUC expression of *MYB75_{pro}* (i.e. expression arising from the empty vector pGREENII) was observed (Fig. 5.11, in purple), confirming the utility of the assay to detect transactivation. Consistent with SHP2 activating *MUM2*, co-infiltrations of *MUM2* promoter/intron FLUC reporter (*MUM2_{pro}+I1::FLUC*) and *SHP2* produced a 2.0-fold increase in FLUC/RLUC when compared to its corresponding baseline (*MUM2_{pro}+I1::FLUC* with empty vector) (Fig. 5.11, in blue). Both fold-increases were statistically significant at $p < 0.05$. Lastly, no statistically significant fold changes were observed in co-infiltrations of *MUM2_{pro}+I1::FLUC* containing quadruple CArG I, II, III and CL II mutations (aka *MUM2_{pro}+I1^{-CArGs}*) either with or without *SHP2* (Fig. 5.11, in yellow). These findings are consistent with SHP2 binding to the CArG box/CL motifs to activate *MUM2*.

Once the utility of the transactivation assay to detect *MUM2* activation was established, the system was next extended to SHP1 and SEP3 to determine whether they too have the capacity to activate *MUM2*. Infiltration of *MUM2_{pro}+I1::FLUC* with empty vector (*MUM2_{pro}+I1::FLUC* baseline) and with *SHP2* served as positive controls for this experiment and yielded a 2.9-fold increase ($p < 0.001$) in FLUC/RLUC ratio (Fig. 5.12, in blue). This result demonstrates that there is some experimental variation between transactivation assays suggesting that data should not be compared between assays performed at different times (Fig. 5.11 & 5.12, in blue). Thus, to avoid inaccuracies associated with experimental variation, all infiltrations were performed and analysed simultaneously on a single microtitre plate. Co-infiltration sets of *MUM2_{pro}+I1* with either *SEP3* or *SHP1* were both able to produce increased FLUC/RLUC ratios of 1.8-fold at $p < 0.01$ and $p < 0.05$, respectively, when compared to the *MUM2_{pro}+I1* baseline (Fig. 5.12, in blue). When *MUM2_{pro}+I1* was co-infiltrated with either *SEP3* and *SHP2* or *SEP3* and *SHP1*, a 3.2-fold ($p < 0.001$) and a 2.5-fold ($p < 0.05$) increase in FLUC/RLUC ratios were observed, respectively, compared to the *MUM2_{pro}+I1* baseline (Fig. 5.12, in blue). As expected, the exact same set of TF combinations co-infiltrated with *MUM2_{pro}+I1^{-CArGs}* failed to produce

any statistically significant increase in FLUC/RLUC ratio when compared to the *MUM2_{pro}+I1^{-CArGs}* baseline (with empty vector) (Fig. 5.12, in yellow). Collectively, these results are consistent with MADS-box TFs SEP3, SHP1 and SHP2 functioning as *MUM2* activators that bind to the CArG box and CL motifs within the *MUM2 I1*.

5.2.4 Interactions of SHP1, SHP2 with LUH/SEU

The possibility of the LUH regulatory complex activating *MUM2* expression through MADS-box TFs SEP3, SHP1 and SHP2 (Fig. 1.9A) is dependent on the presence of all these components in the epidermal layer of the seed coat (MSCs), particularly at the linear cotyledon stage, that coincides with the onset of *MUM2* expression in the MSCs (Jayawardana, 2016). Given that the promoters of SEP3, SHP1 and SHP2 were all found to be active in the MSC (**Chapter 3, Section 3.2.3**), this analysis was extended to include the promoter activity of LUG, LUH and the co-regulators SEU, SLK1 and SLK2, in the developing seed coat. These constructs were previously made by other members of the Golz lab and a list of associated plasmids can be found in Appendix 5, Supplementary Table 1. These promoter-reporter (*LUG*, *LUH*, *SEU*, *SLK1*) and promoter-reporter gene fusion (*SLK2*) constructs were introduced into a Col-0 background by floral dip transformation, and representative lines for *LUH_{pro}::VENUSn*, *LUG_{pro}::VENUSn*, *SEU_{pro}::VENUSn*, *SLK1_{pro}::VENUSn* and *SLK2_{pro}::YFP-SLK2* selected for imaging from early to late seed development. *LUH_{pro}::VENUSn* produced noticeable fluorescence in the MSCs, with peak expression at the linear cotyledon stage, while of the co-adaptors, only *SEU_{pro}::VENUSn* showed weak fluorescence in the MSCs, noticeable from the linear cotyledon stage onwards, with peak expression at the bending cotyledon stage (Fig. 5.13 & 5.14). The diffuse nature of fluorescence from *SLK1_{pro}::VENUSn* and *SLK2_{pro}::YFP-SLK2* made it difficult to conclude if these constructs were active in the MSCs (Fig. 5.13), even under 400-fold magnification (data not shown). *VENUSn* fluorescence associated with the *LUG_{pro}* reporter was not detected in the seed coat, consistent with published data showing that LUG is not active in the developing seed

coat (Walker et al., 2011) (Fig. 5.13). Taken together, these results indicate the possibility that LUH and SEU may be capable of forming a complex with SEP3, SHP1 and SHP2 to activate *MUM2* in the MSCs. Given this, investigating potential interactions between LUH, SEU, SEP3, SHP1 and SHP2 was warranted.

5.2.4.1 Physical interactions between MADS-box TFs and components of the LUG/LUH regulatory complex

The possibility of a *MUM2* activator complex composed of MADS-box TFs SEP3, SHP1, SHP2 and the LUG/LUH regulatory complex was investigated through the use of Y2H assays. The Y2H assay involves fusing proteins of interest to either the DNA-binding domain (BD) or the activation domain (AD) of, for example, GAL4 from *Saccharomyces cerevisiae*, to produce a ‘bait’ and ‘prey’ construct, respectively (Fields & Song, 1989). These constructs can be co-expressed in yeast through either mating or co-transformation, where they are expressed in the nucleus. Positive protein interaction between the bait and prey proteins brings the attached BD and AD domains in close proximity and thus activates transcription of the downstream reporter gene (Fields & Song, 1989). Reporter genes in this assay are typically biosynthetic genes such as *HIS3*, which encodes for the production of histidine and therefore allows *HIS3*⁻ auxotrophic yeast strains to grow on media lacking histidine (Mehla et al., 2017). While Y2H allows for the rapid screening of interactions between proteins, the high rate of false positives is a concern, along with the potential differences in protein folding between *S. cerevisiae* and *Arabidopsis*. One method to address the high false positive rate is to incorporate 3-Amino-1,2,4-triazole (3-AT) in the selection media, a *HIS3* competitive inhibitor, which raises the stringency of *HIS3* selection by allowing only robust interactions that are able to express sufficient *HIS3* to grow (Mehla et al., 2017).

Previous studies have reported interactions between AP1/SEP3 and SEU, so these were used as positive interaction controls in the Y2H assay (Sridhar et al., 2006). In

addition, the empty prey vector *pGADT7* was used to detect auto-activation of bait constructs. Prey constructs consisted of *pGADT7* with either *AP1/SEP3* or other MADS-box TFs *SHP1* and *SHP2*, while bait constructs consisted of *pGBKT7* with either *LUG*, *LUH*, *SEU*, *SLK1* or *SLK2*. A list of associated primers and plasmids can be found in Appendix 5, Supplementary Table 6 & 7. Selection for positive yeast growth was performed on quadruple dropout media (SD/-Trp/-Leu/-His/-Ade) following the co-transformation of bait and prey plasmids into *S. cerevisiae* Y2HGold and duplicated onto quadruple dropout media containing 2 mM 3-AT.

Y2H assays confirmed that AP1 and SEP3 interact with SEU, SLK1 and SLK2, but not with LUG/LUH, while SHP1 and SHP2 did not interact with any of the components of the LUG/LUH regulatory complex (Fig. 5.15, left panel). Furthermore, AP1 interacted strongly with SEU and SLK1, while SEP3 formed strong interactions with SEU based on growth assays performed on the quadruple dropout media containing 2 mM 3-AT (Fig. 5.15, right panel). These results indicate that of the MADS-box TFs of interest (SHP1, SHP2 and SEP3), only SEP3 can interact directly with SEU, SLK1 and SLK2, with stronger interactions seen with SEU, while SHP1 and SHP2 are unable to interact directly with any components of the LUG/LUH regulatory complex.

Given that MADS-box TFs typically form tetramers consisting of either homodimers or heterodimers (de Folter et al., 2005; Honma & Goto, 2001; Immink et al., 2003), SEP3 may mediate the binding of SHP1 and SHP2 to the LUH regulatory complex via the co-regulator SEU to generate a higher order complex that can regulate *MUM2*. This proposed model for *MUM2* activation by the LUH regulatory complex is shown in Fig. 5.16A, and is mediated through SHP1 and SHP2, that likely binds to CArG box and CL motifs studied in **Section 5.2.1**. This model is supported by a large scale yeast three-hybrid (Y3H) study that demonstrated the role of SEP3 as a mediator ('glue') for the formation of MADS-box TF complexes (Immink et al., 2009). Confirmed MADS-box TF complexes arising from this Y3H study include

interactions between SHP1, SHP2 and SEP3 among many other MADS-box TF combinations (Immink et al., 2009).

To confirm interactions between SEU and SHP2 can be mediated by SEP3, a Y3H was performed with SEP3 as a ternary factor. A truncated version of SEP3 lacking the auto-activation domain C2 was used instead of the full length SEP3 (Immink et al., 2009). Growth on quadruple dropout media only occurred when SEP3-C2 was included as a ternary factor with the SHP2 in the presence of SEU, with both negative controls (empty vector and no SEP3-C2) showing no growth (Fig. 5.16B). As Y3H screens for intermediate protein-protein interactions which are typically weaker, addition of 3-AT for higher stringency was omitted. This result is consistent with SEP3 mediating the formation of a LUH complex with SHP2 through SEU.

Collectively, the yeast-based assays show that while SHP1 and SHP2 cannot directly interact with the LUH regulatory complex, SEP3 is able to mediate indirect SHP2 binding through co-regulator SEU. While the formation of SEU and SHP1 complexes mediated through SEP3 were not experimentally confirmed due to time constraints, Y3H data from Immink (2009) supports the formation of this complex. Further experimental validation of MADS box TF interactions with the LUH regulatory complex is provided in the following section using BiFC assays.

5.2.4.2 Physical interactions between MADS-box TFs and the LUG/LUH regulatory complex *in planta*

To address the concern of improper protein folding in Y2H experiments, Y2H results were validated using BiFC assays performed in *N. benthamiana* leaves following transient transformation. BiFC, like Y2H, has the advantage of being simple and rapid assay to perform, and has the additional benefit of assessing interactions *in planta*. BiFC involves the fusion of candidate proteins with either the N- or C- terminal half of a reporter gene such as yellow fluorescent protein (YFP) or VENUS (modified

YFP) (Kodama & Hu, 2012). Positive interactions between the candidate proteins results in the reconstitution of a functional fluorescent reporter protein that produces an observable fluorescent signal (Kodama & Hu, 2012). As these interactions occur in the epidermal layer of *N. benthamiana* leaves which are more similar to *Arabidopsis* than *S. cerevisiae* cells, candidate proteins are more likely to undergo appropriate post-translational modifications and to emulate the protein folding that occurs in their native context (Lee, 2014). Despite its many strengths, BiFC may produce false positives, as fluorescent protein fragments have been reported to self-assemble independently of candidate protein-protein interactions (Horstman et al., 2014). Notably, BiFC assays produce false negatives to an even greater extent, which has been attributed to steric hindrance caused by fused fluorescent reporter fragments that disrupt candidate protein interactions (Horstman et al., 2014; Kodama & Hu, 2012). False negatives can be reduced by testing both N- and C-terminal fusion combinations to maximise the probability of the YFP fragments coming together in the appropriate orientation, while false positives can be controlled for by using fluorescent protein variants with mutated binding sites where self-interaction is suspected to occur (Horstman et al., 2014; Kodama & Hu, 2012).

Having established the formation of a LUH complex with SEP3, SHP1 and SHP2 in Y2H assays, the next step was to examine if the formation of this complex occurs *in planta* using the BiFC complementation assay in *N. benthamiana* leaves. BiFC constructs were generated using seamless cloning, and a list of associated primers and plasmids can be found in Appendix 5, Supplementary Table 8 & 9. All sets of BiFC infiltrations were performed with three biological replicates, with each leaf from an independent plant examined. Visualisation of infiltrated leaves was performed at 6-9 DPI using fluorescence microscopy under blue light excitation (515 nm). Positive interactions were only noted if fluorescence was reconstituted across all three biological replicates, with the most representative area imaged.

N-terminal (*VENUS_N*) and C-terminal (*VENUS_C*) VENUS fragments were each fused in frame to the 5' end of the coding sequence of LUH and SEU and both construct

combinations tested (Fig. 5.17). While both combinations produced fluorescent signals at 6 DPI, the VENUS_N:LUH and VENUS_C:SEU combination produced the strongest fluorescence with a higher number of fluorescing nuclei (6 vs 2) (Fig. 5.17; overview panel). Thus, VENUS_N:LUH was subsequently used to test for interactions with SEP3, SHP1 and SHP2 fused to VENUS_C fragments. Consistent with Y2H results, no interactions between LUH, SEP3, SHP1 and SHP2 were observed (Fig. 5.17).

Next, BiFC was used to investigate interactions between SEP3, SHP1 and SHP2 with SEU. The combination of VENUS_C:SEU and VENUS_N:SEP3 served as a positive control based on Y2H results. Again, only VENUS_N:SEP3 but neither VENUS_N:SHP1 nor VENUS_N:SHP2, produced a fluorescence signal with VENUS_C:SEU at 6 DPI (Fig. 5.18). In addition, VENUS_N:LUH was found to interact with VENUS_N:SEP3 to produce a fluorescence signal at 6 DPI only when the 35S_{pro}::SEU construct was co-infiltrated (Fig. 5.18). These data indicate that only SEP3 could interact with LUH through SEU, but not SHP1 and SHP2.

Interactions between SEP3, SHP1 and SHP2 were next examined. Interactions between SEP3 and SHP2 were first tested using both construct combinations, with VENUS_N:SEP3 and VENUS_C:SHP2 producing stronger fluorescence signal at 6 DPI compared to VENUS_C:SEP3 and VENUS_N:SHP2 (Fig. 5.19). Similarly, the combination of VENUS_N:SEP3 and VENUS_C:SHP1 produced fluorescence at 6 DPI, indicating protein interactions between SEP3 and SHP1 (Fig. 5.19). This suggests the possibility of SHP1 and SHP2 interacting with LUH-SEU through SEP3. Finally, BiFC was used to test whether a higher-order complex of LUH with SHP1, SHP2 arose in the presence of SEP3 and SEU (Fig 5.16A). Infiltrated leaves containing a combination of VENUS_N:LUH, VENUS_C:SHP2, 35S_{pro}::SEU and 35S_{pro}::SEP3 were examined from 3-9 DPI, with a fluorescence signal detectable from 9 DPI (Fig. 5.20). This interaction was validated by the absence of fluorescence from the corresponding negative control lacking both SEU and SEP3 at 9 DPI (Fig. 5.20). Likewise, the same combination of VENUS_N:LUH, 35S_{pro}::SEU and 35S_{pro}::SEP3 with VENUS_C:SHP1

produced a fluorescence signal at 9 DPI, but not its corresponding negative control (Fig. 5.20). The late onset of VENUS expression in samples containing four constructs (Fig. 5.20; 9 DPI) compared to samples containing less than four constructs (Fig. 5.17-19; 6 DPI) may be reflective of the increased difficulty in assembling higher-order protein complexes using BiFC in *N. benthamiana*.

Overall, these BiFC results are consistent with results from yeast assays, supporting the formation of a higher-order LUH regulatory complex with SHP1 and SHP2 through SEU and SEP3 (Fig. 5.16A). As SHP2 activates *MUM2*, established through the analysis of SRDX fusions and transactivation assays (**Section 5.2.3**), a likely function of this higher-order complex is to activate *MUM2* in the seed coat of *Arabidopsis*.

5.3 Discussion

5.3.1 Mutations of functional CArG box and CL motifs support activation of *MUM2* by MADS-box TFs

As MADS-box TFs act as tetramers in *Arabidopsis* and therefore have a high degree of redundancy (Immink et al., 2009; Theissen et al., 2016), the inactivation of functional CArG box and CL motifs provided a simple method of experimentally determining that MADS-box TFs can bind to *MUM2*. While both deletion and mutation strategies of selected CArG box and CL motifs were used, the deletion strategy proved problematic due to an overlapping MYB binding site (6 bp), MYB II, between the CArG III and CL II motifs (Fig. 5.1). Nevertheless, previous work performed on identifying the regulatory regions of *MUM2* utilised the deletion strategy (Jayawardana, 2016) and therefore necessitated the extension of deletions to additional CArG box and CL motifs identified by analysis of published ChIP-seq data to facilitate direct comparisons with the original work.

A key difference in the functions of identified CArG box and CL motifs arose between the deletion and mutation strategy, with CL II identified as a seed coat-specific *MUM2* repressor element via deletions and as a *MUM2* activator element in both seed coat and embryo via mutations (Table 5.2). As MADS-box TFs require binding to at least two CArG box motifs to facilitate looping of target DNA into an appropriate conformation for either activation or repression (Mendes et al., 2013; Theissen et al., 2016), the deletion strategy may have disrupted the spacing and/or looping of regulatory elements resulting in aberrant VENUSn reporter results. This is the most reasonable explanation for the disparity in results seen between the two strategies, as mutation of CArG box and CL motifs avoids changing the regulatory sequence spacing within the *MUM2* *I1* region A and therefore is likely to loop correctly (Mendes et al., 2013). Additionally, the deletion of CL II results in the removal of MYB II, a *MUM2* activator element based on double MYB deletions (Table 5.1) and may confound the interpretation of CL II function as a *MUM2* repressor. This scenario, however, seems unlikely as inadvertent deletion of a *MUM2*

activator element cannot explain how deletion of CL II in the *MUM2_{pro}+I1^{ΔCArGIII,CLI,CLII,MYBII}* reporter restores VENUSn expression in the *luh* MSC (Table 5.1). Therefore, besides retaining proper spacing of *MUM2*, the mutation strategy is superior to the deletion strategy as it allows for targeted inactivation of TF binding sites; for example, CL II can be mutated while preserving the overlapping MYB II to discern its function, whereas the deletion of CL II removes both CL II and MYB II sites, complicating the interpretation of data arising from this experiment (Fig. 5.1).

Interestingly, previous work showed that the deletion of *MUM2 I1* region A (~200 bp) was sufficient to rescue VENUSn expression in the MSCs of *luh* mutants, indicating a presence of a *MUM2* repressor element in this regulatory region (Jayawardana, 2016). This supported the presence of a *MUM2* inhibitor and therefore increased the credibility of the de-repression model (Fig. 1.9C), which was examined in **Chapter 3**. Initially, the rescue of VENUSn to the MSCs in *luh* seen in *MUM2_{pro}+I1^{ΔCArGIII,CLII,MYBII}* reporter lines supported this model (Fig. 5.3), with CL II identified as the likely *MUM2* repressor element (Table 5.2). However, the subsequent single mutation of CL II was instead shown to function as a *MUM2* activator element, producing decreased VENUSn expression in both seed coat and embryo compared to the full length *MUM2_{pro}+I1* reporter in *luh/+* and *luh* backgrounds (Table 5.2). Indeed, mutations of CArG I, II, III and IV in *MUM2_{pro}+I1* collectively indicated that these binding sites were critical for *MUM2* activation, as a dramatic decrease was observed in VENUSn expression. Ultimately, no repressor elements were identified using the mutation approach (Table 5.2).

A few conclusions can be drawn from these data: First, if it is assumed that the mutational analysis has correctly assigned functions to the CArG box/CL elements, it would seem that there are no genuine *MUM2* repressor elements in the *MUM2* regulatory sequence, leading to a rejection of the de-repression model for *MUM2* regulation (Fig. 1.9C). Second, increase of VENUSn fluorescence in the MSCs of *luh* seeds with the deletion of region A is inconsistent with results of the mutational

approach and are therefore likely to be the consequence of aberrant looping of *MUM2* *I1* that is facilitated by MADS-box TFs as proposed in Fig. 5.16A. Given that region A is 200 bp in length, its deletion is likely to affect proposed DNA looping capabilities. Additionally, on re-examination of region A deletion reporters in Col-0 and *luh*, VENUSn expression was diffuse, with no clear rescue of VENUSn expression observed in the MSC of *luh* (data not shown). The lack of distinct VENUSn localisation in the nucleus suggests the loss of a functional nuclear localisation signal and the instability of the region A deletion reporter, which are potential outcomes stemming from the mis-looping of the reporter construct caused by the deletion.

Overall, these data led us to conclude that mutation approach is more technically reliable than the deletion approach, and clearly showed that CArG box/CL motifs function as *MUM2* activator elements. These data therefore support a model of regulation in which the MADS-box TFs promote expression of *MUM2* in MSCs.

5.3.2 MADS-box TFs are *MUM2* activators

Multiple lines of evidence described in this chapter, as well as data reported in previous chapters supports a model in which MADS-box TFs function as *MUM2* activators. Most notably, the high seed coat expression of *SHP1*, *SHP2* and *SEP3* VENUSn reporters in Col-0 (**Chapter 3, Section 3.2.3**), the mucilage extrusion defects of *sep3*, *shp1 shp2* and *shp1 shp2 sep3* mutants (**Chapter 3, Section 3.2.4**) and the decrease of *MUM2* expression in *shp1 shp2* mutants (**Chapter 4, Section 4.2.3**) all point to these TFs as *MUM2* activators. In addition, mutagenesis of the CArG I, II, III and CL II abolishes *MUM2* expression in wildtype and *luh* mutant MSCs (Table 5.2). While *SHP1*, *SHP2* and *SEP3* are regulated by *LUH* (**Chapter 3**), a limitation of the CArG box motif mutations fused to VENUSn reporters is the difficulty to discern if these motifs are indeed *LUH*-responsive. Specifically, the removal of *LUH*-responsive elements activating *MUM2* is expected to produce an

equal decrease of *MUM2* activity in both Col-0 and *luh* mutants, which is difficult to quantify given that *MUM2* itself is LUH-responsive and therefore produces different *MUM2* baseline activity across both backgrounds (Fig. 5.4). Therefore, these motifs may correspond to LUH-responsive elements that SEP3, SHP1 and SHP2 bind to but further experimentation is required to establish this definitively.

Next, by fusing SHP2 with the SRDX repressor motif, direct evidence that SHP2 is able to function as an activator in *Arabidopsis* was provided, as a range of dominant negative mucilage extrusion defects were produced in Col-0 lines containing *35S_{pro}::SHP2::SRDX* (Fig. 5.7). More importantly, these mucilage defects were associated with a decrease in *MUM2*, thereby directly linking the loss of SHP2 to a reduction of *MUM2* activity in *Arabidopsis* seeds (Fig. 5.9). To verify the loss of *MUM2* protein/enzyme, western blots should be performed on samples from the same lines used for qRT-PCR to provide evidence that the transcriptional loss of *MUM2* results in an actual loss of *MUM2* protein/enzyme. Additionally, mucilage could be extracted from these mutants to determine if the mucilage extrusion defects are arising from a loss of functional *MUM2* (i.e increased *t*-Gal residues in the carbohydrate linkage data). However, such results may be cryptic given the subtle carbohydrate profile obtained from *shp1 shp2* mutants in **Chapter 4, Section 4.2.5**.

The dominant negative effects arising from the SRDX fusion also extended to disruptions to floral development, with poor stamen formation and deformed carpels observed in moderate to strong lines, respectively (Fig. 5.8). This is unsurprising as MADS-box TFs that may interact with SHP2 are involved in floral development, such as AP3, a B class TF needed for stamen development (Wuest et al., 2012). Thus, conversion of SHP2 and therefore interacting AP3 into a repressor, disrupts stamen formation, resulting in poor to deformed stamens (Fig. 5.8). As carpel deformation is also observed in strong SRDX lines, it is likely that *SHP2::SRDX* interferes with AG activity, especially since SHP2 is known to interact with AG and STK through SEPs for carpel identity (Favaro et al., 2003). It is pertinent to note that these disruptions in upstream floral development occurred in numerous Col-0 lines transformed with

35S_{pro}::SHP2::SRDX, leading to either poor or no seed set in these mutants (Fig. 5.8). Thus, the sampling of the 36 independent plants to assay their mucilage defects (Fig. 5.7) are from less severe dominant negative mutants that were able to set seed and thus may not represent the strongest effect in the seed coat. To circumvent this issue, a seed coat-specific promoter that is not responsive to MADS-box TFs should be utilised to drive *SHP2::SRDX* instead of the constitutive *35S_{pro}*, such as the *GILT_{pro}* identified in **Chapter 3, Section 3.2.5**.

Other data supporting MADS-box TFs as *MUM2* activators were obtained from luciferase-based transactivation assays, which provided a high degree of confidence in generating a quantifiable measure of promoter activation. This is particularly notable as the transactivation system was successfully validated using *MYB75_{pro}::LUC* and FIL, a known activator of *MYB75* (Boter et al., 2015) (Fig. 5.11). Besides demonstrating SHP2 as a *MUM2* activator, a series of transactivation assays in *N. benthamiana* also indicated that SHP1 and SEP3 function as *MUM2* activators *in planta* (Fig. 5.12). This is evidenced in the increase of the FLUC/RLUC ratio of *MUM2_{pro}+I1::LUC* when either SHP1, SHP2 or SEP3 was added to the transactivation assay system (Fig. 5.12). Furthermore, these MADS-box TFs appear to be binding to the CArG box and CL motifs, as use of *MUM2_{pro}+I1^{-CArGs}::LUC* in combination with SHP2 produced FLUC/RLUC outputs similar to the baseline control (Fig. 5.12). Collectively, these data strongly support the binding of SEP3, SHP1 and SHP2 to the CArG box and CL motifs present in *MUM2 I1*. Additionally, the power of the transactivation assay is its ability to test multiple combinations of TFs, such as SHP1 and SHP2 with SEP3 respectively (SHP1+SEP3 and SHP2+SEP3), which produced elevated *MUM2_{pro}+I1::LUC* activation compared to their single TF counterparts (Fig. 5.12). This result raises the possibility that this system could be used to reconstitute the LUH regulatory complex of *MUM2* in *N. benthamiana* and observing if an increase in FLUC/RLUC is produced. Unfortunately, this system was unable to test the effect of SHP1, SHP2, SEP3, SEU and LUH collectively on *MUM2_{pro}+I1::LUC*, as co-infiltrations beyond four constructs produced uninterpretable results with high error bars (data not shown). For

example, co-infiltrations with five constructs (*MUM2_{pro}+I1::LUC*, SHP2, SEP3, SEU and LUH) produced FLUC/RLUC outputs lower than baseline control (*MUM2_{pro}+I1::LUC* with empty vector) (data not shown). One explanation is that perhaps beyond four co-infiltrations, insufficient amounts of each protein accumulate in *N. benthamiana* leaves preventing the full LUH complex from being re-constituted. Therefore, despite the ability of the transactivation system to test the effect of multiple TFs on *MUM2_{pro}+I1*, this system was unable to be used to test the direct activation model (Fig. 1.9A).

Collectively, both SRDX fusions and transactivation assays have clearly demonstrated that MADS-box TFs function as *MUM2* activators *in planta* and therefore are not candidate *MUM2* repressors as proposed by Jayawardana (2016).

5.3.3 LUH directly activates *MUM2* via MADS-box TFs

LUH is known to be a positive regulator of *MUM2* although how it mediates this activation was unknown (Huang et al., 2011; Jayawardana, 2016; Walker et al., 2011). The previously proposed de-repression model (Fig. 1.9C) can be excluded on the basis that MADS-box TFs SEP3, SHP1 and SHP2 are clearly *MUM2* activators and not repressors *in planta* (**Section 5.3.2**). While it is unclear if the activation of *MUM2* by MADS-box TFs is either independent or dependent on LUH, there are two models accounting for activation of *MUM2* by LUH that may involve MADS-box TFs (Fig. 1.9A & B). The simplest model is direct activation, where LUH-SEU binds directly to MADS-box TFs and activates *MUM2* together as a complex (Fig. 1.9A). The second model is indirect activation, where LUH-SEU may activate an intermediate *MUM2* activator such as MADS-box TFs, leading to indirect activation of *MUM2* (Fig. 1.9B). Finding that LUH-SEU can physically complex with SHP1/SHP2/SEP3 via yeast assays (Fig. 5.15 & 5.16) and BiFC (Fig. 5.17-5.20) supports the direct activation model of *MUM2* through LUH (Fig. 1.9A). Both Y2H and BiFC indicate that LUH associates with MADS-box TFs via SEU, which is able to bind to SEP3, but not

SHP1 and SHP2 directly (Fig. 5.16A). Based on this model, it would be reasonable to conclude that loss of either SEP3 or SEU results in loss of the LUH activation complex. This premise is somewhat intriguing as *sep3* mutants have subtle mucilage extrusion defects compared to *shp1 shp2* but may be explained by the functional redundancy with SEP1, SEP2 and SEP4. Consistent with this possibility is the high expression of SEP1 and SEP3 in the seed coat (**Chapter 3, Table 3.3**) (Winter et al., 2007). Accordingly, *sep1 sep3* double mutants have not only a severe mucilage defect, but abnormal seed shape, implicating that SEP1 and SEP3 are functionally redundant in regulating processes in both early seed development and mucilage extrusion (**Chapter 3, Section 3.2.4.1**). According to the direct activation model it might be expected that *seu* mutants have mucilage defects. However, this is not observed and may be due to functional redundancy with SLK1, SLK2 and SLK3 (Bui et al., 2011). However, it is unclear which of these SLKs would be function redundantly as neither the SLK1 or SLK2 reporters gave clear signals in the seed coat (Fig. 5.13). This may in part be due to issues with the YFP reporter used for examining the expression pattern of *SLK2_{pro}* and should be repeated with the VENUSn reporter for clearer results. Additionally, these promoter reporters should be extended to include *SLK3_{pro}*, which was not performed due to time constraints. Nonetheless, semi-quantitative RT-PCR indicated that all three SLKs are expressed in seed coat tissue (Walker et al., 2011)

Given that the loss of *LUH* results in a lack of mucilage extrusion, it is reasonable to expect that LUH is needed for the function of the LUH activation complex in the seed coat (Fig. 5.16A). Indeed, LUH and not LUG is expressed in the seed coat (Walker et al., 2011). In the same manner, the combinatorial loss of SEU, SHP1, SHP2 and SEP3 should result in the loss of the LUH activation complex (Fig. 5.16A) and therefore phenocopy *luh* mutants. This theory is supported by the enhanced mucilage extrusion of the triple mutant *shp1 shp2 sep3* that is slightly less severe than *luh* (Fig. 3.9). Thus, generation of mutants such as *shp1 shp2 sep3 seu* should be performed to determine if this can phenocopy *luh* mutants and therefore support the physical model predicted in Fig. 5.16A. Such results would also indicate that LUH primarily activates

MUM2 through a complex consisting of MADS-box TFs, and not through other TFs such as MYBs. That is, MADS-box TFs act to regulate *MUM2* in a LUH-dependent fashion.

Overall, while there is strong evidence supporting the direct activation model of *MUM2* expression (Fig. 1.9A), there is no direct evidence that the LUH-SEU-MADS box complex is responsible for activating *MUM2* in the seed coat. Future work to address this includes ChIP-PCR on *Arabidopsis* seeds to provide direct evidence that MADS-box TFs, SEU and LUH are capable of binding to the 5' region of *MUM2 I1*. While ChIP-PCR on seeds cannot provide cell-specific resolution such as in MSCs, the context of SHP1, SHP2 and SEP3 activation of *MUM2* can be inferred from their respective promoter-VENUSn activity, which appear primarily active or exclusive to the seed coat (**Chapter 3, Section 3.2.3**). Unfortunately, ChIP-PCR is technically challenging, making it difficult to obtain useful data within reasonable timeframes if parameters such as crosslinking and DNA shearing are not sufficiently optimised (Yamaguchi et al., 2014). Alternatively, a simpler technique is the *in vitro* electrophoretic mobility shift assay (EMSA), which involves incubating radioactive labelled DNA probes (ranging from short oligonucleotides to several thousand bp in size) with protein(s) of interest and subsequently running the mixture on polyacrylamide/agarose gels (Fried & Daugherty, 1998; Rasimas et al., 2007). A positive DNA-protein interaction is denoted by an upward shift in size compared to the lone DNA control (Fried & Daugherty, 1998; Garner & Revzin, 1981; Hellman & Fried, 2007). Therefore, EMSA could be performed on region A of *MUM2 I1* and/or full length *MUM2 I1* as DNA probes and incubated with MADS-box TFs SHP1, SHP2 and SEP3 to determine if these TFs are able to bind to *MUM2*. Overall, EMSA is simple to perform, rapid, sensitive, and has the ability to test multiple TFs on a single probe, allowing for identification of complexes that can bind to the target DNA (Fried & Daugherty, 1998; Garner & Revzin, 1981; Hellman & Fried, 2007). However, of the two techniques, ChIP-PCR remains the gold-standard as it provides a means to measure *in vivo* activity and should therefore be attempted. Lastly, given that MADS-box TFs form tetramers in *Arabidopsis*, strong evidence such as the

enhanced mucilage defect in triple *shp1 shp2 sep3* mutants (Fig. 3.7) and the increased activation of *MUM2_{pro}+I1* in transactivation assays by combinations of these MADS-box TFs (Fig. 5.12) indicate that these MADS-box TFs work together in the same tetramer. As *shp2* produces the most severe mucilage defects compared to *shp1* and *sep3* single mutants, it is likely that this tetramer comprises of two SHP2, one SHP1 and one SEP3 protein. However, the exact composition of this tetramer requires further study to resolve.

5.3.4 Potential role of MYBs in *MUM2* regulation

The MYB TF family is the largest in TF family in *Arabidopsis*, with over 190 members (Stracke et al., 2001; Yanhui et al., 2006) and are known to participate in diverse roles such as embryogenesis (Wang et al., 2009), floral development (Mandaokar et al., 2006) and lateral organ formation (Lee et al., 2009). The deletion of MYB I and MYB II motifs in *MUM2 I1* suggest that MYBs may have roles in mucilage polysaccharides regulation, specifically as *MUM2* activators since their removal results in an overall decrease in VENUSn expression in both Col-0 and *luh* seeds (Table 5.2, Fig. 5.2). A reasonable first experimental approach to addressing their role includes extending the mutation approach to MYB binding sites to determine if these sites are actual *MUM2* activator elements, or merely a side effect of altering the spacing of elements within *MUM2 I1*. Disruptions to *MUM2* spacing arising from MYB deletions are predicted to be less likely, given that MYB binding sites are only 5-6 bp in size (Fig. 5.1) (Prouse & Campbell, 2013). Additionally, the Belmonte microarray dataset can be used to narrow down MYB proteins that bind to these motifs by mining for MYBs that are expressed in the seed coat. Such MYBs include MYB61, MYB5, MYB23 and MYB52: *myb61* and *myb5* mutants have reduced mucilage polysaccharides production and abnormal columella formation (Li et al., 2009; Penfield et al., 2001) which are enhanced in *myb5 myb23* double mutants (Li et al., 2009), while *myb52* mutants have mucilage extrusion defects arising from a loss of *PECTIN METHYLESTERASE INHIBITOR6 (PMEI6)* and *SUBTILISIN-LIKE*

SERINE PROTEASE 1.7 (SBT1.7) activity (Shi et al., 2018). Although *myb23* mutants do not appear to have mucilage extrusion defects, *MYB23* fused to a repressor motif reportedly had reduced mucilage extrusion (Matsui et al., 2005). Therefore, these mutants and *MYB23* repressor lines should be examined to determine if *MUM2* levels are affected. One caveat is that *myb5* and *myb61* mutants have reduced columella formation and therefore likely arise from the reduced mucilage polysaccharides biosynthesis. Thus, even if *MYB5* and *MYB61* are found to activate *MUM2*, it is unlikely that the loss of *MUM2* alone contributes to their observed mucilage defects.

5.3.5 Conclusion

LUH appears to be directly activating *MUM2* through a complex with MADS-box TFs SHP1, SHP2 and SEP3 mediated by SEU. While SHP2 has been convincingly shown to act as a *MUM2* activator, and not a *MUM2* repressor, through both SRDX motif fusions and transactivation assays, there is no direct evidence that SHP2 binds to *MUM2* in the *Arabidopsis* seed coat. Thus, future work includes performing ChIP-PCR on SHP1/SHP2/SEP3 to show that these TFs can bind to *MUM2* in the seed coat. Alternatively, direct binding of *MUM2* could be demonstrated *in vitro* by EMSA.

Additionally, the exact combination of MADS-box TFs in the tetramer binding to LUH-SEU is not known. Based on the increased severity of *shp2* mutants compared to *shp1* and *sep3* in terms of mucilage extrusion, this tetramer is proposed to include two SHP2, one SHP1 and one SEP3 protein. However, it is possible that other MADS-box TFs are involved and form part of the complex, and therefore this proposition requires further investigation. Besides MADS-box TFs, deletion constructs of MYB binding sites in *MUM2 I1* indicate that MYBs may also have a role in activating *MUM2*. However, given that *myb5* and *myb61* mutants produce less seed mucilage, probably due to reduced mucilage polysaccharides production, it is unlikely that *MUM2* has any major role in these mutants. Instead, the generation of quadruple mutant *shp1 shp2 sep3 seu* may resolve if *MUM2* is primarily activated by

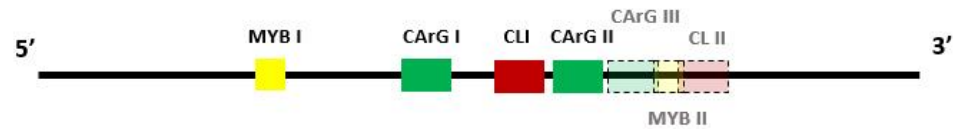
the LUH regulatory complex with SHP1, SHP2 and SEP3, should this mutant phenocopy *luh/mum2*.

Table 5.1. Analysis of *MUM2* promoter/intron-reporter constructs with various deletions.

CAR_G box, CL motifs and MYB binding sites identified in region A of *MUM2* I1 were deleted in various combinations as indicated below. VENUSn fluorescence activity was assessed and standardised across all lines. Asterisk (*) indicates data provided by Dr. Nadeeka Jayawardana (Jayawardana, 2016).

Promoter-reporter construct		Background	Strength of fluorescence observed in seed coats				# of lines
			Strong	Intermediate	Weak	Very weak	
<i>Region A</i>		Col*	24	7	3	-	34
		<i>luh</i> *	-	9	25	3	37
<i>ΔCAR_G III, MYB II (double deletion)</i>		Col	-	-	3	13	16
		<i>luh</i>	-	-	3	13	16

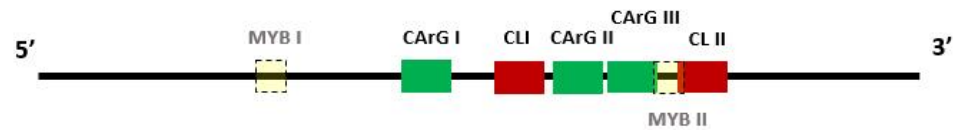
Chapter 5



Δ CArG III, CL II, MYB II
(triple deletion)



Δ CArG III, CL I, CL II, MYB II
(quadruple deletion)



Δ MYB I, MYB II
(MYB deletion)

Col	5	3	6	2	16

<i>luh</i>	1	8	5	2	16

Col	5	5	6	-	16

<i>luh</i>	1	7	8	-	16

Col	4	4	8	-	16

<i>luh</i>	-	-	-	16	16

Table 5.2. Summary of CArG box, CL motifs and MYB binding sites deletions and mutations.

Proposed activity of the CArG box, CL motifs and MYB binding sites within region A based on deletion and mutation analysis (see Table 5.1 and 5.3). In purple shading are motifs selected for further analysis by a base mutation approach (see text for details). Asterisk (*) indicates unpublished findings from Dr. Nadeeka Jayawardana. Bolded statements are final conclusions drawn from mutation analysis when deletion and mutation analysis are not in agreement. SC: seed coat.

TF binding site	Deletions/ mutations	VENUSn expression compared to full length region A		Conclusion
		SC	Embryo	
CArG I	*Deletion	*Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	*No data	<i>MUM2</i> activator (active in SC and embryo)
	Mutation	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	
CArG II	*Deletion	*Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	*No data	<i>MUM2</i> activator (active in SC and embryo)
	Mutation	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	
CArG III	Deletion	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	<i>MUM2</i> activator (active in SC and embryo)
	Mutation	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (<i>luh</i> /+ & <i>luh</i>)	
CL I	Deletion	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>), but no obvious differences	<i>MUM2</i> activator (active in SC only)

between quad and triple deletions				
CL II	Deletion	Restores some <i>VENUSn</i> expression in SC of <i>luh</i>	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>), but no obvious differences between triple and double deletions	<i>MUM2</i> repressor (active in SC only)
	Mutation	Decrease in <i>VENUSn</i> expression (<i>luh/+</i> & <i>luh</i>) No <i>VENUSn</i> restoration in SC of <i>luh</i>	Decrease in <i>VENUSn</i> expression (<i>luh/+</i> & <i>luh</i>)	<i>MUM2</i> activator (active in SC and embryo)
MYB I	Deletion	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	<i>MUM2</i> activator (active in SC and embryo)
MYB II	Deletion	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	Decrease in <i>VENUSn</i> expression (Col-0 & <i>luh</i>)	<i>MUM2</i> activator (active in SC and embryo)

Table 5.3. Summary of CArG box motifs mutations.

CArG box and CL motifs identified in region A were mutated in various combinations as indicated with dotted and faded boxes. MYB binding sites not shown in the schematic. Fluorescence strength was assessed and standardised across all lines and is indicated below. Asterisk (*) indicates data provided by Dr. Nadeeka Jayawardana (Jayawardana, 2016).

Promoter-reporter construct		Strength of fluorescence observed in seed coats					# of Lines
		Background	Strong	Intermediate	Weak	Very weak	
Region A		Col*	24	7	3	-	34
		<i>luh</i> *	-	9	25	3	37
<i>mut</i>CL II (single mutation)		<i>luh</i> +	6	8	2	-	16
		<i>luh</i>	2	10	4	-	16



*mut*CArG I, CArG II, CArG III
(triple mutation)

<i>luh</i> /+	-	3	10	3	16
<i>luh</i>	-	2	7	7	16



*mut*CArG I, CArG II, CArG III, CL II
(quadruple mutation)

<i>luh</i> /+	-	-	3	13	16
<i>luh</i>	-	-	3	13	16

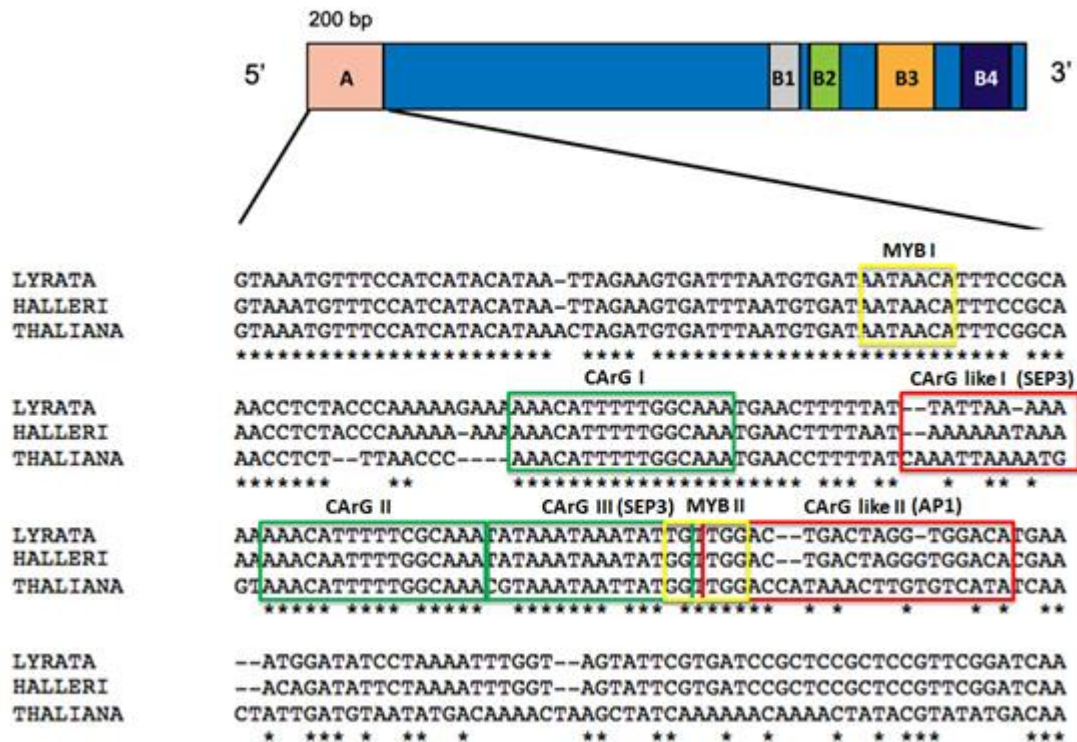


Figure 5.1. Potential TF binding sites in region A of the *MUM2* intron 1.

Mining of the 5' 200 bp region revealed several TF binding sites such as CArG box motifs (in green), CArG-like motifs (in red) and MYB binding sites (in yellow). Asterisks (*) indicate nucleotides conserved across mucilage-producing species *A. lyrata*, *A. halleri* and *A. thaliana*. Adapted and modified from Jayawardana (2016).

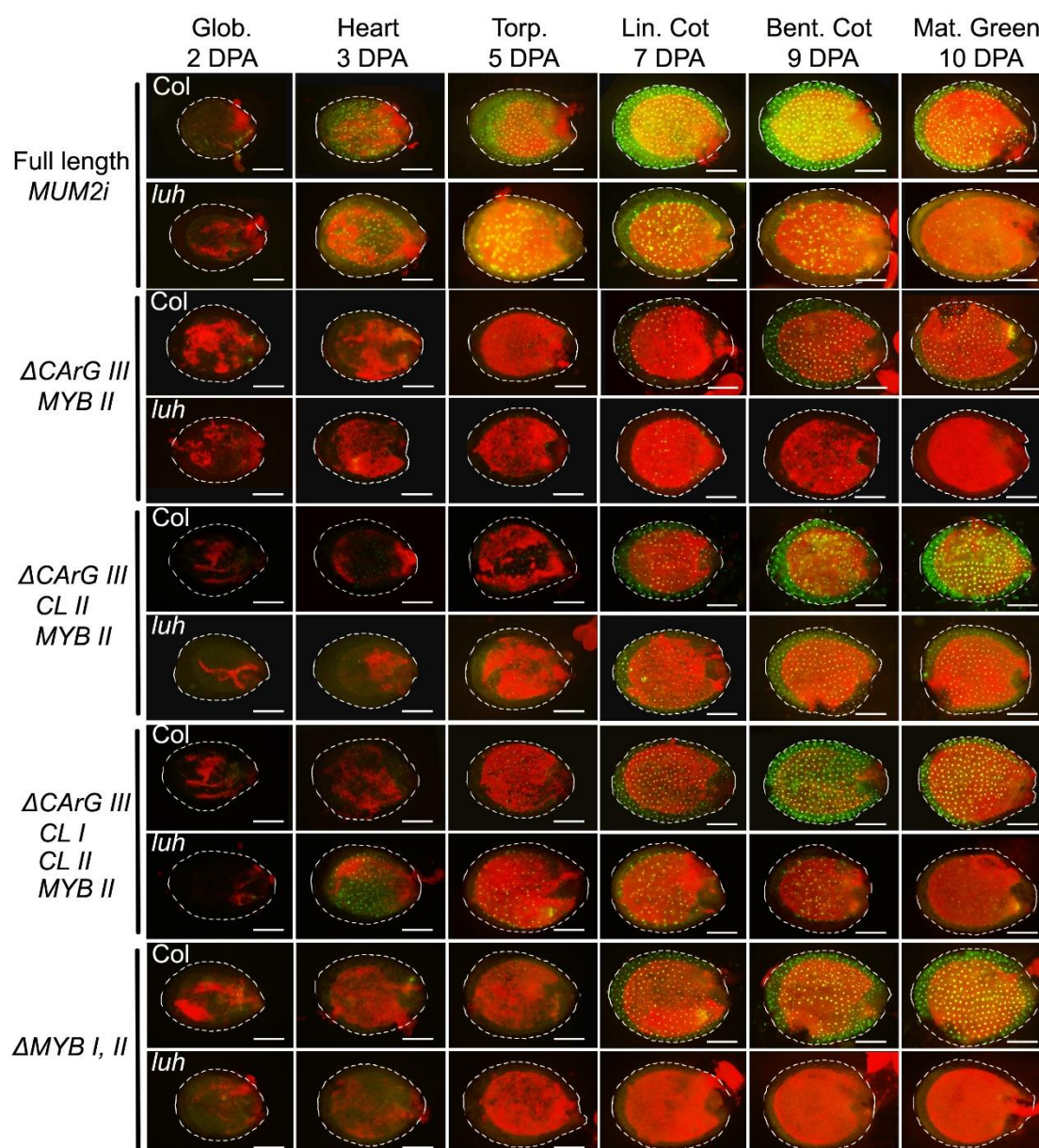


Figure 5.2. Activity of *MUM2* promoter/intron-reporter deletions constructs within the developing seed coat of T₁ transgenic lines.

CAR_G box, CL motifs and MYB binding sites (indicated in Table 5.1) were deleted within the promoter and first intron (region A) of *MUM2* and subsequently fused to the *VENUSn* reporter. These *MUM2*_{pro+I1} deletion constructs were then introduced into Col-0 and *luh* backgrounds, and a representative line from each background was imaged. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear

Cotyledon (Lin. Cot.), Bending Cotyledon (Bent. Cot) and Mature Green (Mat. Green). Dotted lines around seed indicate the boundary of the outer seed coat layer.

Scale bar = 200 μm .

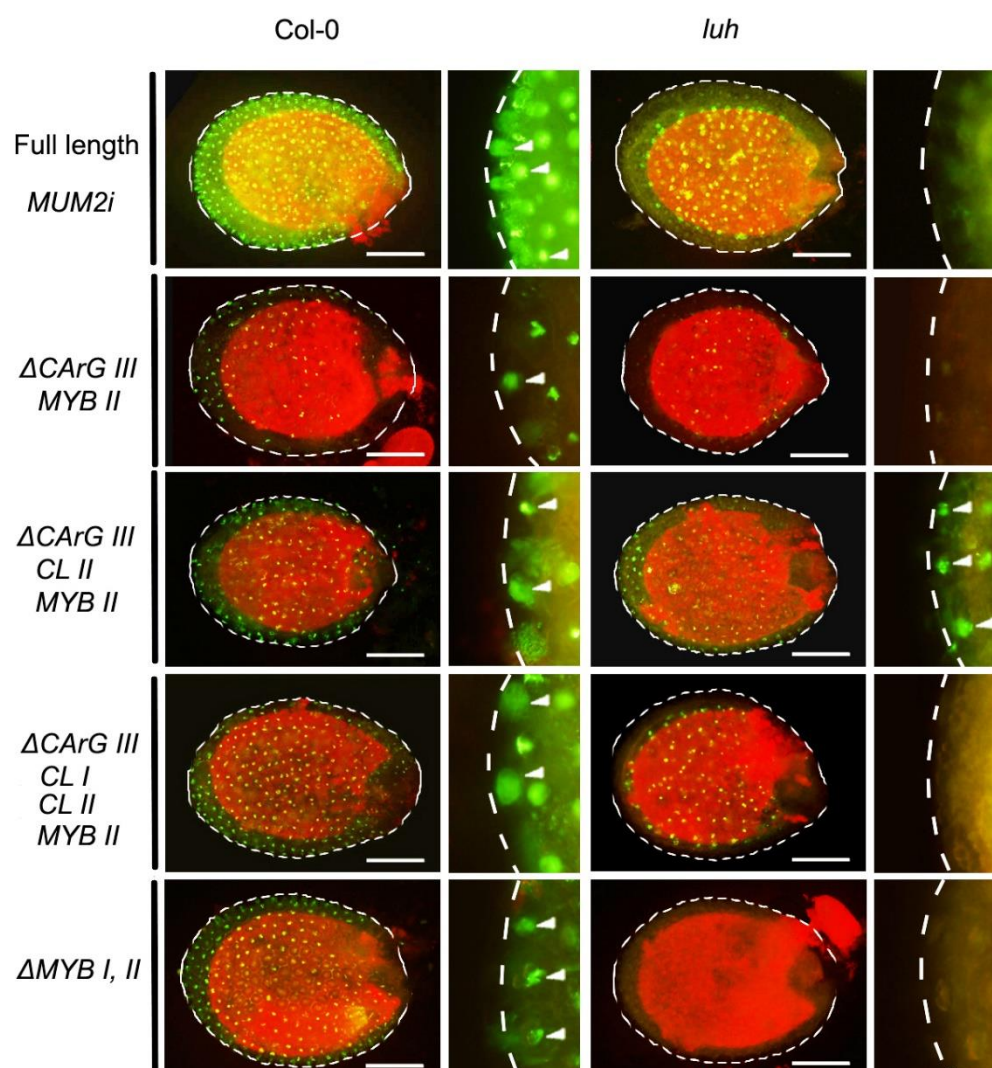


Figure 5.3. Enlarged view of *MUM2* promoter/intron-reporter deletions constructs within the epidermal layer of the seed coat.

See Fig. 5.2 for details of the generation of transgenic lines. An image of whole seeds at the linear cotyledon stage is shown on the left and a magnified view of the corresponding seed coat is shown on the right. Dotted lines around seed indicate the extent of the outer seed coat, with arrowheads indicating nuclear localised VENUSn fluorescence in the epidermal layer.

Scale bar = 200 μm .

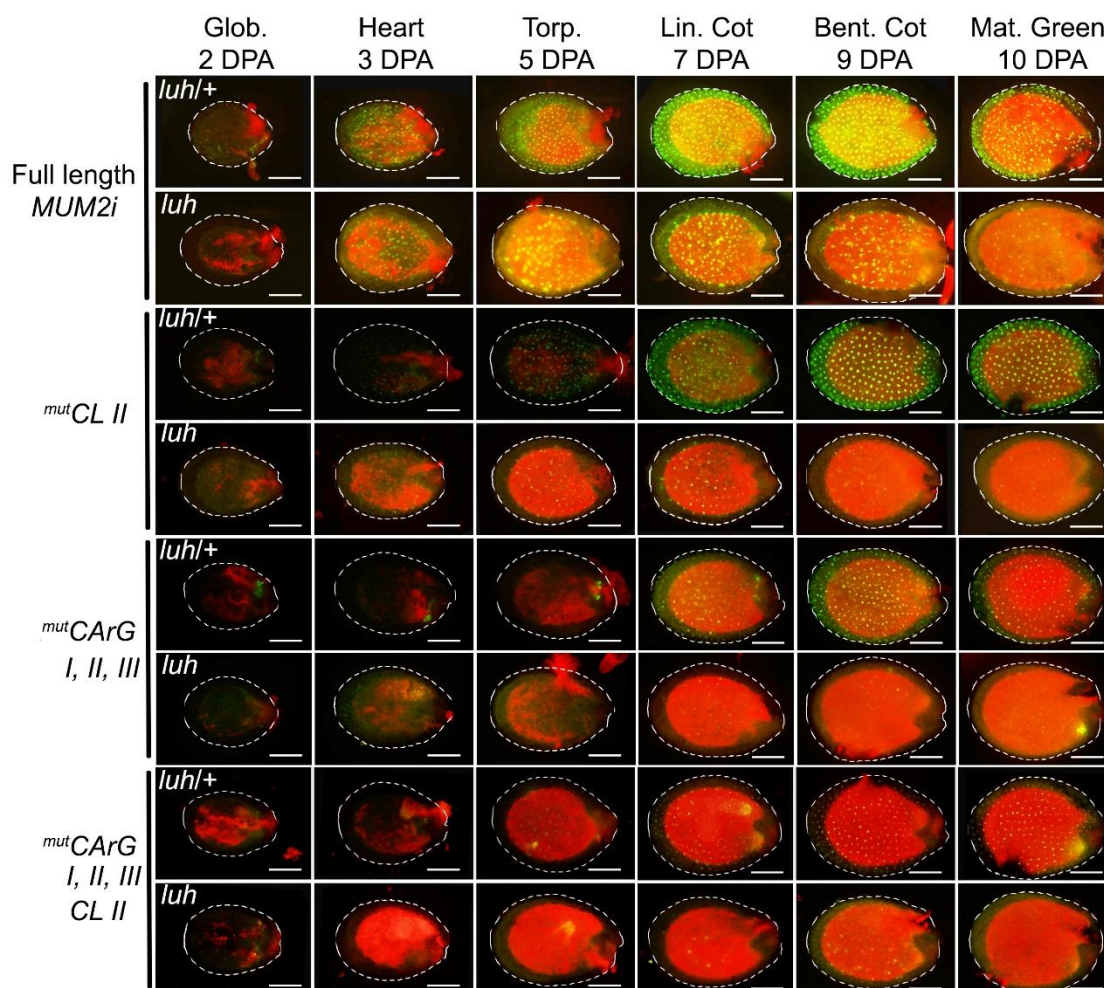


Figure 5.4. Activity of *MUM2* promoter/intron-reporter mutation constructs within the developing seed coat of T₁ transgenic lines.

CAR_G box and CL motifs (indicated in Table 5.1) present in region A were mutated prior to fusion to the *VENUSn* reporter. The *MUM2*_{pro}+*I1* mutations were then introduced into a *luh* background, and representative lines were backcrossed to Col-0 and *luh* mutants to generate lines in two different backgrounds (*luh*/+ vs *luh*/*luh*) where activity of the promoter-reporter is directly comparable. All images were captured using the same exposure (ISO 200) across all developmental stages with the exception of *MUM2*_{pro}+*I1*^{-CARGI,CAR_GII,CAR_GIII,CLII}, which was taken at 4x exposure (ISO 800). All images were adjusted for brightness in Adobe Photoshop (ver. 12.0) to reflect equal exposure. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent. Cot) and Mature

Green (Mat. Green). Dotted lines around seed indicate the boundary of the outer seed coat layer (MSCs).

Scale bar = 200 μm .

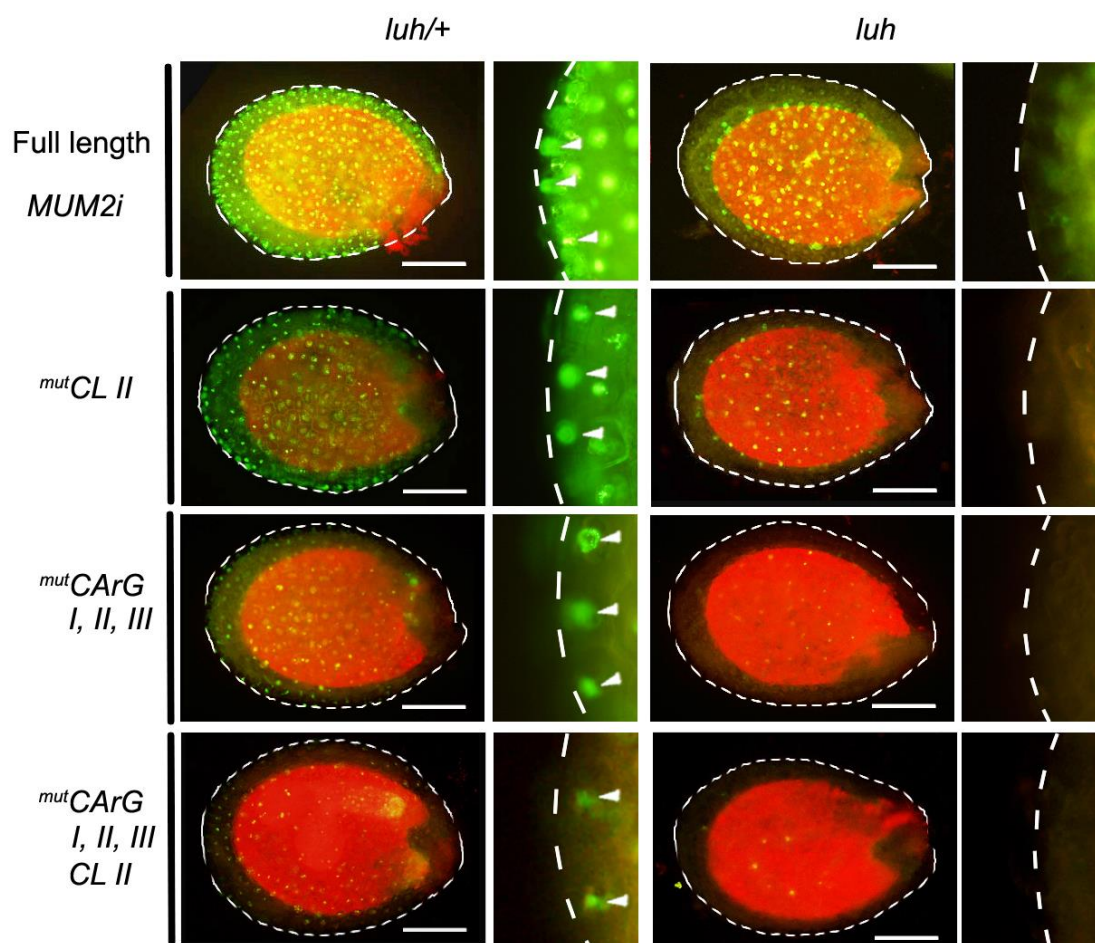


Figure 5.5. Enlarged view of *MUM2* promoter/intron-reporter mutations constructs within the epidermal layer of the seed coat.

Constructs bearing CArG box and CL motifs mutations in the *MUM2* promoter/intron-reporter have been described in Fig. 5.4. An image of whole seeds at the linear cotyledon stage is shown on the left and a magnification of their corresponding seed coat is shown on the right. Dotted lines around seed indicate the outer seed coat boundary, with arrowheads indicating the *VENUSn* expression in the epidermal layer (MSCs).

Scale bar = 200 μ m.

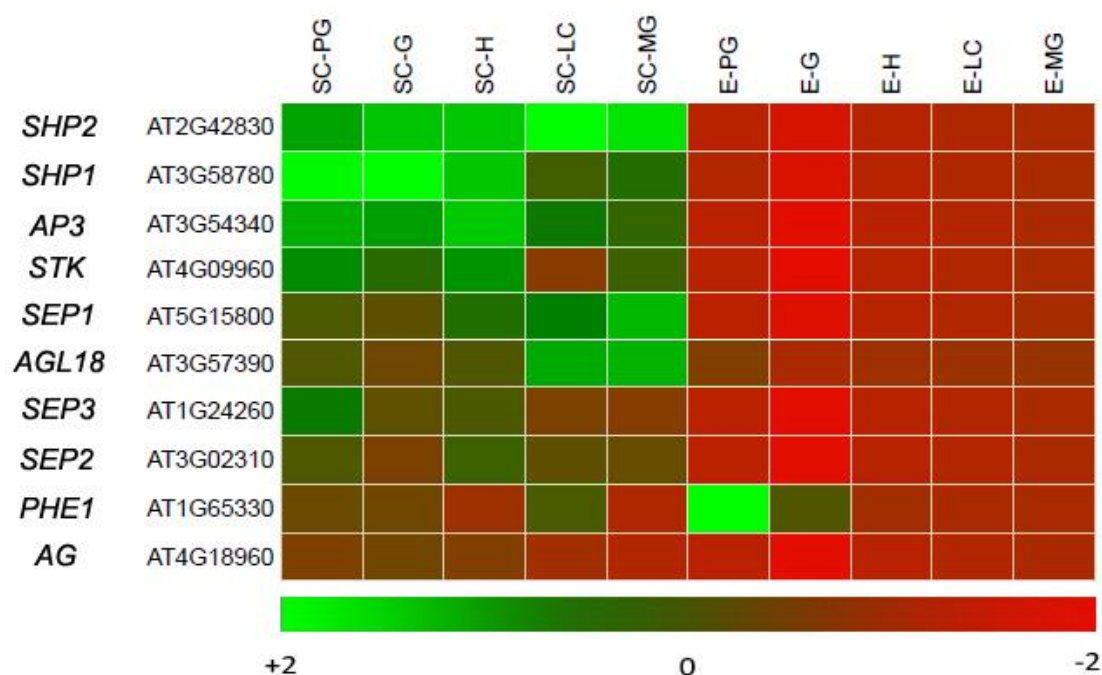


Figure 5.6. Expression pattern of MADS-box TFs in the seed coat and embryos at different developmental stages.

The heat map was generated using graphical tool Protovis, with data derived from publicly available microarray datasets of seed expressed genes (Belmonte et al., 2013; Bostock & Heer, 2009). Only MADS-box TFs with seed coat expression are shown here, arranged from highest to lowest overall expression in the seed coat. Colour scale reflects deviation from the overall mean across genes, with red representing low expression values (<-2 S.D) and green representing high expression values ($>+2$ S.D). To accurately assign colours to expression values while maintaining expression patterns within and across genes, a modifier score of 50 was added to each readout to prevent saturation. Developmental stages are denoted as following: SC-PG: Seed Coat Pre-Globular; SC-G: Seed Coat Globular; SC-H: Seed Coat Heart; SC-LC: Seed Coat Linear Cotyledon; SC-MG: Seed Coat Mature Green; E-PG: Embryo Pre-Globular; E-G: Embryo Globular; E-H: Embryo Heart; E-LC: Embryo Linear Cotyledon; E-MG: Embryo Mature Green.



Figure 5.7. Mucilage extrusion phenotypes of $35S_{pro}::SHP2::SRDX$ lines.

Four categories of mucilage extrusion phenotypes were observed from the 36 T₁ lines assayed with Ruthenium Red staining; strong, moderate, weak and no mucilage release. Strong mucilage release is comparable to Col-0, while weak mucilage release is comparable to *shp2* mutants, and no mucilage release comparable to *luh/mum2* mutants.

Scale bar = 500 μ m.



Figure 5.8. Floral defects observed in some $35S_{pro}::SHP2::SRDX$ lines.

Three classes of flowers were apparent in T_1 transgenic lines and classified as wildtype, mild and strong floral defects. The wildtype class produce flowers that have no obvious defects and are capable of good seed set. The mild class have flowers in which stamens are shortened relative to wildtype and have decreased pollen production. The strong class have flowers that lack petals, have fewer stamens which fail to produce pollen and often produce split carpels. Both mild and strong floral phenotypes were sterile.

Scale bar = 1 mm.

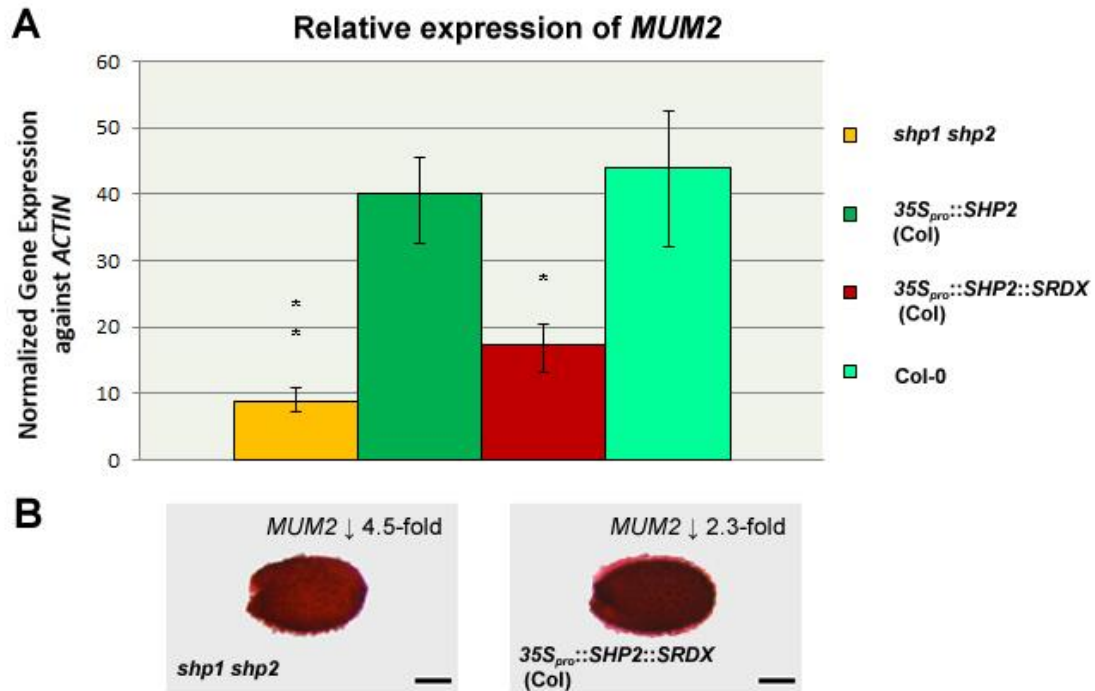


Figure 5.9. Expression of *MUM2* in *35S_{pro}::SHP2::SRDX* lines.

A: *MUM2* expression in whole seeds of *shp1 shp2*, *35S_{pro}::SHP2*, *35S_{pro}::SHP2::SRDX* and *Col-0* at the linear cotyledon stage of development was determined by qRT-PCR. Error bars represent +/- standard error (S.E) of mean for three biological replicates and statistical significance was calculated using Student's unpaired *t*-test (* $p < 0.05$ & ** $p < 0.001$).

B: A representative Ruthenium Red stained seed derived from *shp1 shp2* and *35S_{pro}::SHP2::SRDX* lines with fold change of *MUM2* expression relative to *35S_{pro}::SHP2* indicated.

Scale bar = 200 μ m

Effectors (TFs):

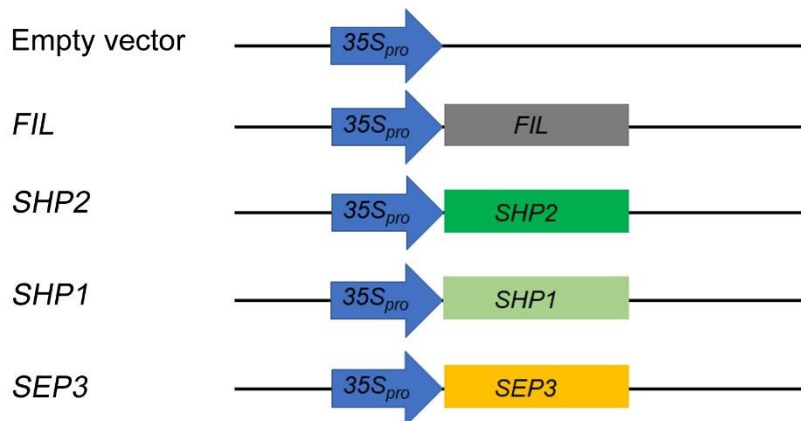
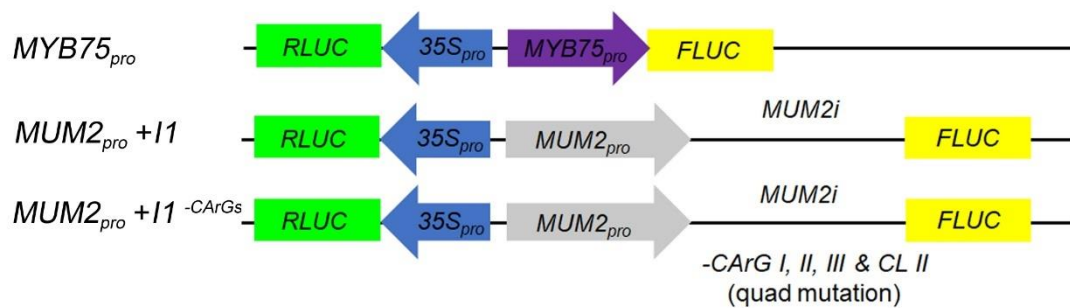
**Reporters:**

Figure 5.10. Constructs generated for transactivation assays.

Constructs used for the transactivation assays are shown, with arrows indicating direction of promotor activity. Figures not drawn to scale.

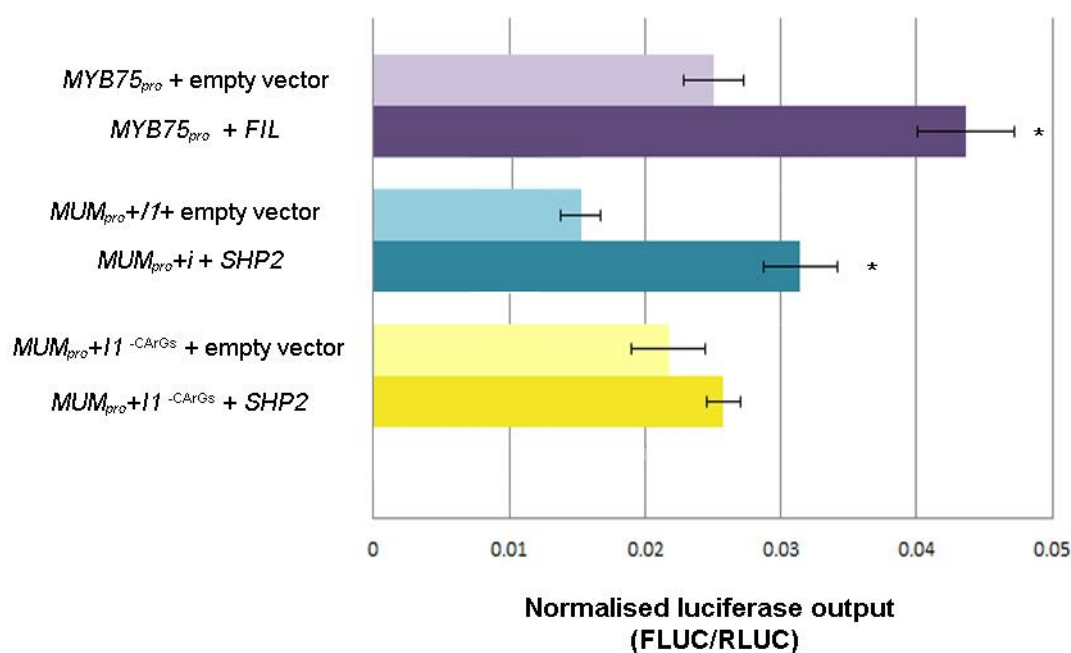


Figure 5.11. Transactivation assays on the *MUM2* promoter/intron reporter.

MYB75 promoter (*MYB75_{pro}*), intact *MUM2* promoter/intron (*MUM2_{pro}+I1*) or the *MUM2* promoter/intron with CArG box/CL motif mutation in *I1* (see text for details; *MUM2_{pro}+I1^{-CArGs}*) driving a FLUC reporter were co-infiltrated with empty pGREENII vector to determine baseline activity for each promoter. Shown in purple are the positive controls for the assay, in which *MYB75_{pro}*-FLUC is co-infiltrated with and without FILAMENTOUS FLOWER (*FIL*), a known TF activator of *MYB75_{pro}* (Boter et al., 2015). Shown in blue are *MUM2_{pro}+I1*-FLUC assays with and without *SHP2*, while in yellow are *MUM2_{pro}+I1^{-CArGs}*-FLUC assays with and without *SHP2*. Statistical significance between baseline and co-infiltration with either *FIL* or *SHP2* was calculated from three biological replicates using Student's *t*-test (* $p < 0.05$).

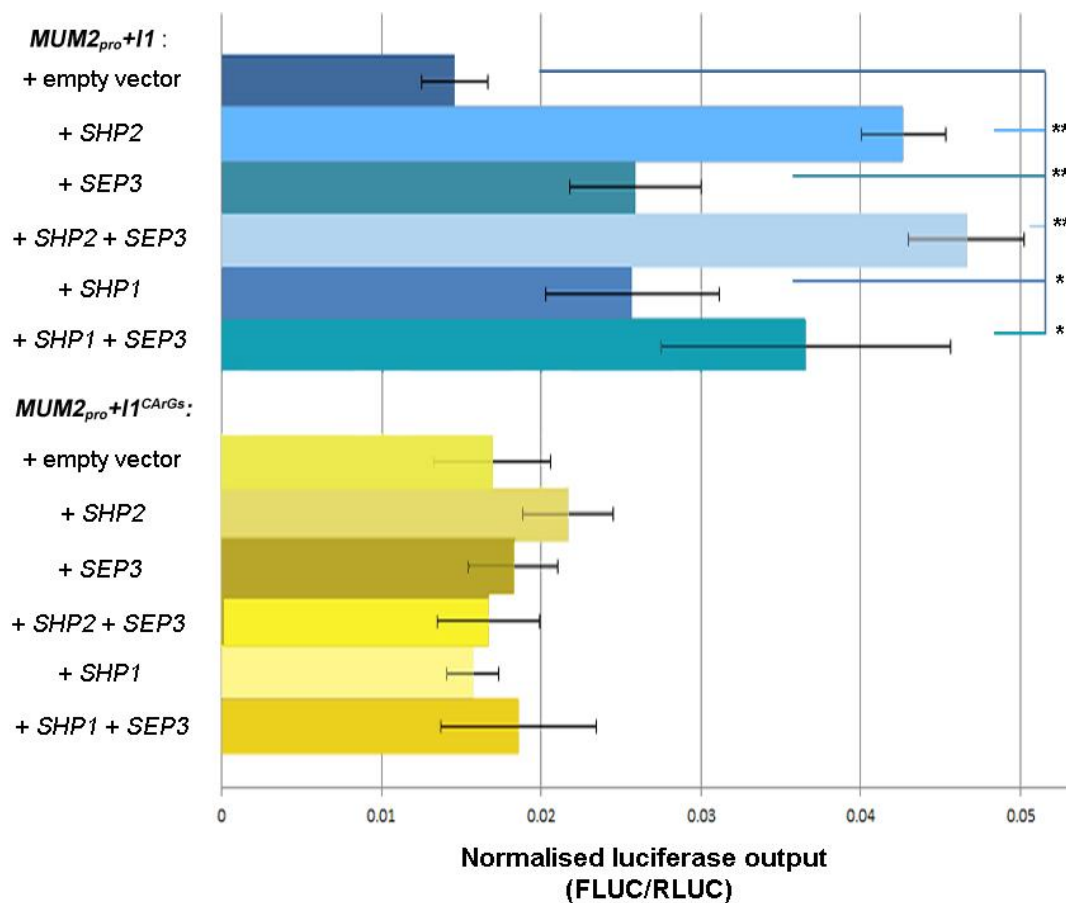


Figure 5.12. Transactivation assay performed on the *MUM2* promoter/intron using different combinations of MADS-box TFs.

The intact *MUM2* promoter/intron-FLUC construct and *MUM2* promoter/intron with CARG box/CL mutation-FLUC construct are described in Fig. 5.11. Shown in blue are *MUM2_{pro}+I1*-FLUC constructs co-infiltrated with various MADS-box TFs, while shown in yellow are *MUM2_{pro}+I1^{CARGs}* constructs. Statistical significance between respective baselines of *MUM2_{pro}+I1*, *MUM2_{pro}+I1^{CARGs}* and co-infiltrations with MADS-box TFs were calculated from three biological replicates using Student's *t*-test (* $p < 0.05$ & ** $p < 0.001$).

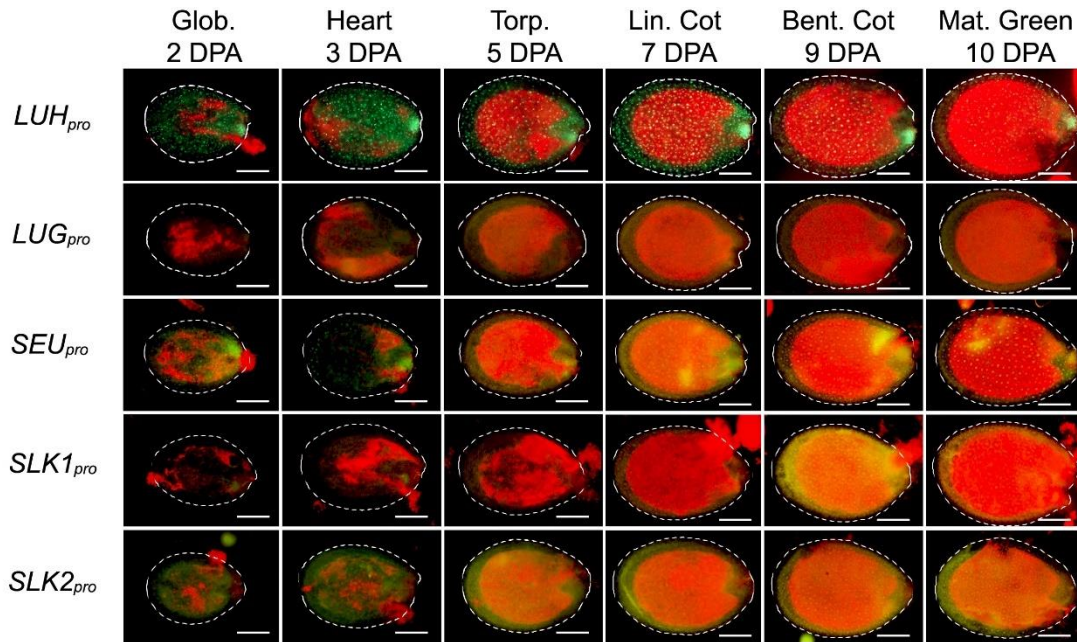


Figure 5.13. Promoter-reporter activity of LUG/LUH regulatory complex components in the developing seed coat.

LUH (*LUH_{pro}*), *LUG* (*LUG_{pro}*), *SEU* (*SEU_{pro}*), *SLK1* (*SLK1_{pro}*) and *SLK2* (*SLK2_{pro}*) promoters were fused to a fluorescent reporter (*VENUSn* or *YFP*, see text for details) and introduced into a Col-0 background. Reporter activity was assessed at early to late seed developmental stages in representative T₁ lines. Six stages were captured: Globular (Glob.), Heart, Torpedo (Torp.), Linear Cotyledon (Lin. Cot.), Bending Cotyledon (Bent. Cot) and Mature Green (Mat. Green). Dotted lines around seed indicate the extent of the outer seed coat layer (MSCs).

Scale bar = 200 μ m.

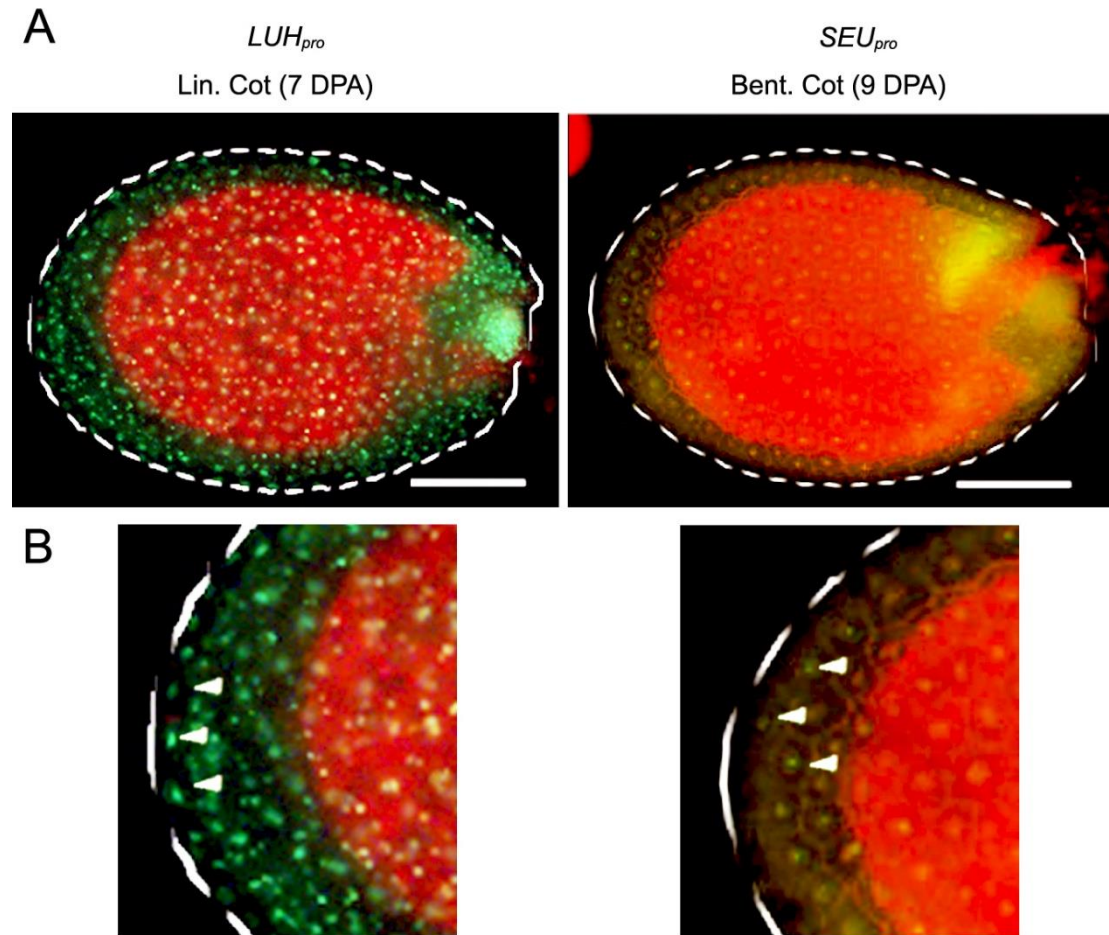


Figure 5.14. Visualising maximal reporter activity in seeds of lines transformed with the LUH and SEU promoter-reporters.

A: Maximal promoter activity was observed at the linear cotyledon (Lin. Cot; 7 DPA) for *LUH* and bending cotyledon (Bent. Cot; 9 DPA) for *SEU*.

B: 2x magnification of seeds shown in **A**. Arrowheads indicating the fluorescence expression in the epidermal layer (MSCs).

Dotted lines around seed indicate the extent of the outer seed coat layer, with.

Scale bar = 200 μm .

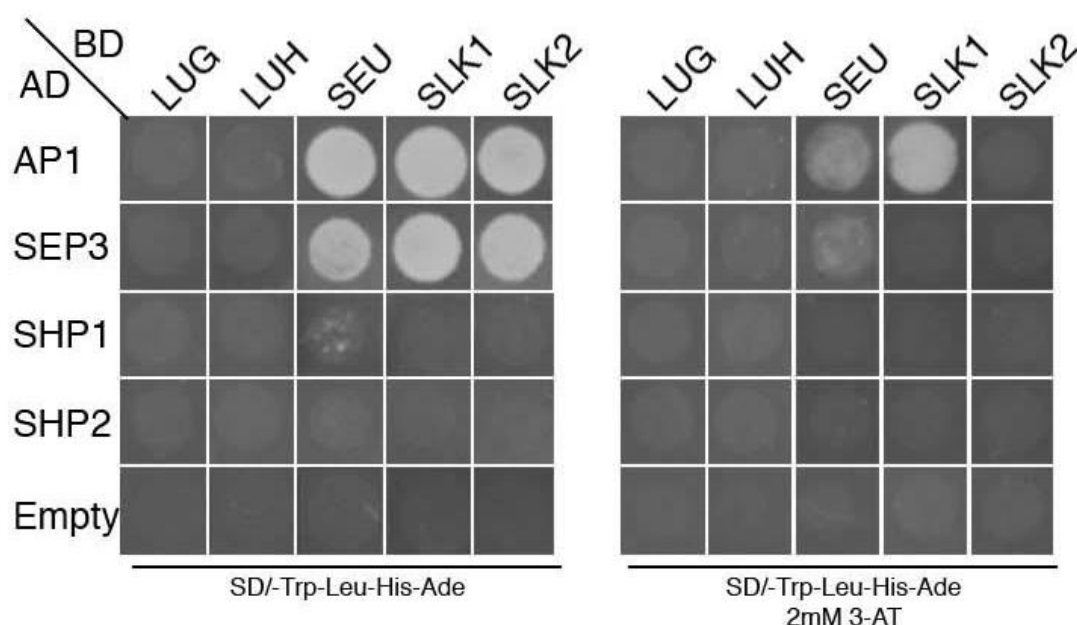


Figure 5.15. Y2H assays between MADS-box TFs and components of the LUG/LUH regulatory complex.

MADS-box TFs SHP1 and SHP2 fused to the GAL4 activation domain (AD) were tested for interactions with components of the LUG/LUH regulatory complex fused to the GAL4 DNA binding domain (BD) by assaying yeast growth on quadruple dropout media (SD/-Trp-Leu-His-Ade). MADS-box TFs AP1 and SEP3 were used as positive controls for the yeast assays, as previous work had shown that they interact with the co-adaptors SEU, SLK1 and SLK2 (Sridhar et al., 2006), while empty prey vector *pGADT7* was used as a negative control. Interactions were also tested on quadruple dropout media containing 2 mM 3-AT to determine the strength of interactions. Successful interactions between bait and prey proteins are indicated by growth of yeast colonies on the plates.

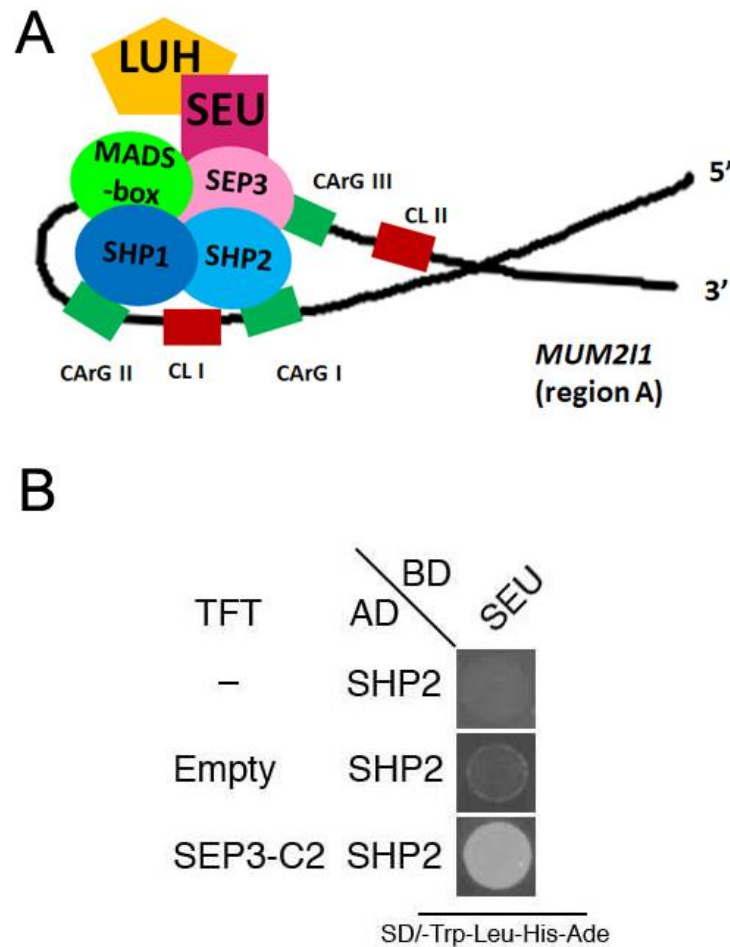


Figure 5.16. Y3H assays to assess the capacity of SEP3 to function as a ternary factor.

A: Proposed structure of the LUH regulatory complex in the MSCs. LUH is recruited to CARG box motifs located in region A of *MUM211* by physically interacting with a SEU-SEP3 complex, mediating *MUM211* looping for regulation. Previous work has shown that SEP3 can interact with both SHP1 and SHP2 in developing flowers (Immink et al., 2009), hence it is inferred that SEP3 is also complexes with SHP1 and SHP2 within the MSCs. Figure not drawn to scale and is only indicative.

B: Y3H assay in which SHP2 fused to the GAL4 AD is tested for interactions to SEU fused to DB, either alone or in the presence of SEP3-C2 in yeast vector pTFT (TFT). SEP3-C2 is a truncated version of SEP3 lacking a domain involved in autoactivation (Immink et al., 2009). Growth of yeast transformed with the combinations listed in the figure was assessed on quadruple dropout media.

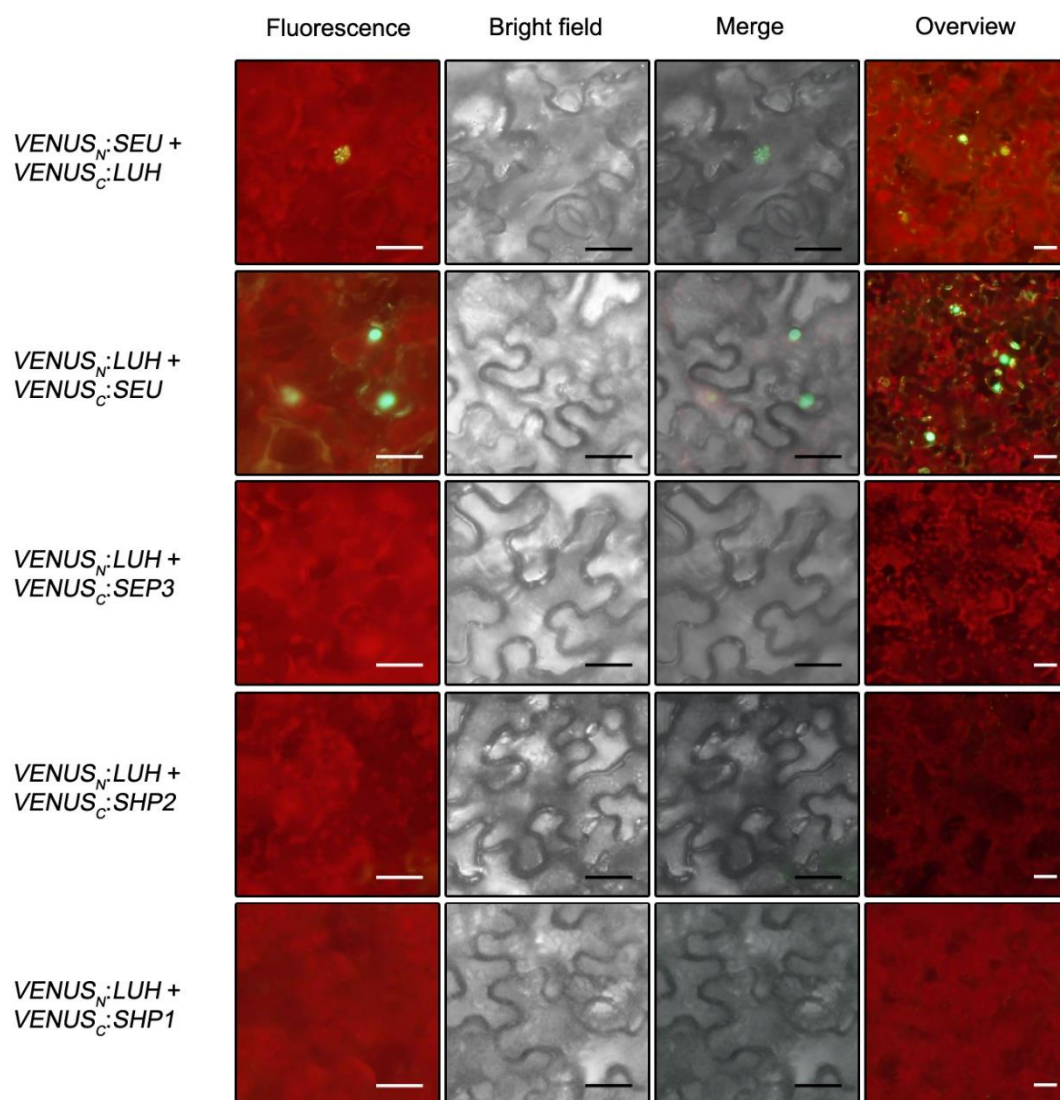


Figure 5.17. BiFC interactions between MADS-box TFs and LUH.

LUH and the MADS-box TFs SEP3, SHP1 and SHP2 were each fused to one half of the fluorescent protein VENUS (either $VENUS_N$ or $VENUS_C$) and their capacity to interact assessed following their introduction into *N. benthamiana* leaves. Interactions resulting in BiFC produce nuclear fluorescence, which is detected by fluorescence microscopy. LUH/SEU combinations were included as positive controls for the BiFC assay (top two rows). Brightfield microscopy were also performed and this image was merged with the output of fluorescence microscopy to provide cellular context for the BiFC fluorescence. Overview shows number of nuclei observed to produce fluorescence as a measure overall fluorescence strength. No BiFC signals were produced between LUH and the MADS-box TF proteins.

Scale bar = 20 μ m.

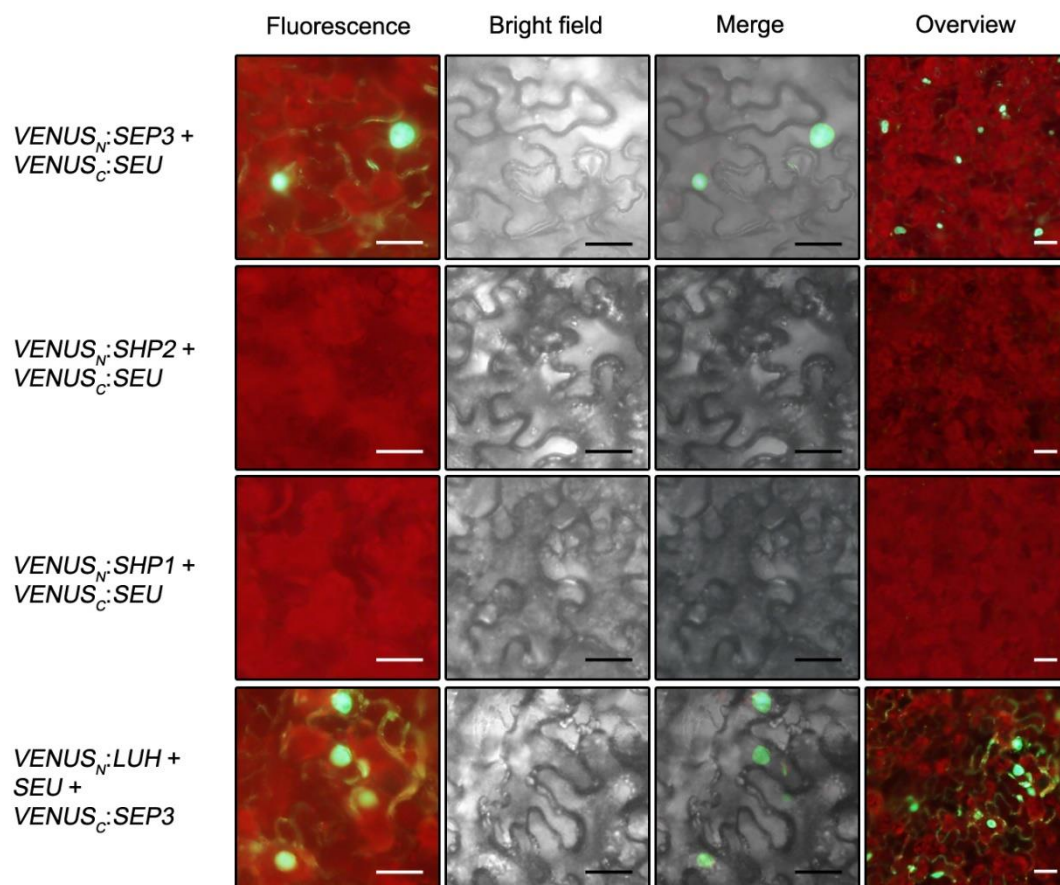


Figure 5.18. BiFC interactions between MADS-box TFs and SEU.

Fusions were generated as described in Fig. 5.17. Presence of nuclear fluorescence when between SEU and SEP3 are co-expressed confirms that these proteins are capable of interacting (top row). No BiFC signal was produced with either SHP/SEU combination (2nd and 3rd rows). Robust BiFC signals were generated when the LUH/SEP3-VENUS fusions were combined with SEU. In this case, it is assumed that SEU functions as a ternary factor linking LUH to SEP3. Microscopy was performed as described in Fig. 5.17.

Scale bar = 20 μ m.

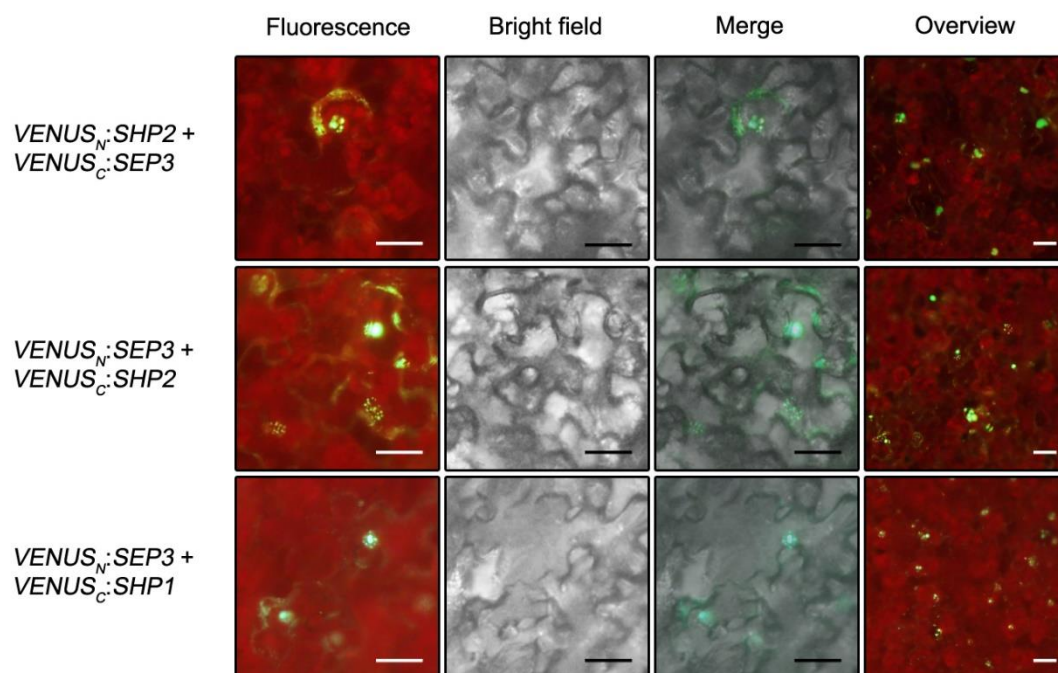


Figure 5.19. BiFC interactions between SEP3, SHP1 and SHP2.

Interaction between SEP3 and SHP2 were first examined in both orientations. Subsequent examination of interactions between SEP3 and SHP1 utilised *VENUS_N:SEP3* as this produced the strongest signal with more fluorescing nuclei when infiltrated with *VENUS_C:SHP2*. Both fluorescence and brightfield microscopy were performed as described in Fig. 5.17.

Scale bar = 20 μ m.

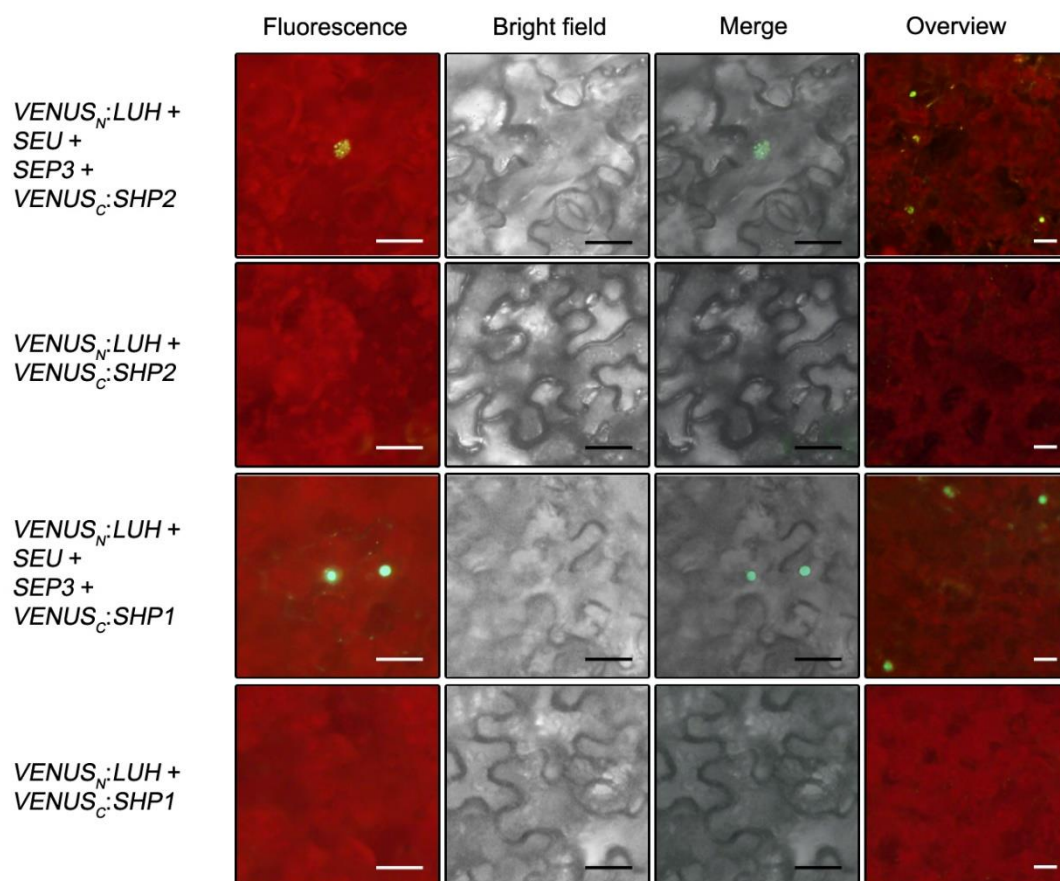


Figure 5.20. BiFC interactions between SEP3, SHP1, SHP2 and components of the LUH regulatory complex.

Interactions between *VENUS_N:LUH*, *VENUS_C:SHP2* and *VENUS_C:SHP1* were assessed in combination with SEP3 and SEU as ternary factors. Infiltrations with only BiFC reporters that did not produce a positive interaction (*VENUS_N:LUH*, *VENUS_C:SHP2* and *VENUS_N:LUH*, *VENUS_C:SHP1*) were included as negative controls and examined from 3-9 DPI. Both fluorescence and brightfield microscopy were performed as described in Fig. 5.17.

Scale bar = 20 μ m.

Chapter 6

Final discussion and conclusions

Final discussion and conclusions

6.1 The roles of MADS-box TFs in *Arabidopsis* seeds

The overall aim of this thesis was to understand how LUH regulates *MUM2* and to determine the molecular mechanism underlying this regulation, which involves MADS-box TFs (Fig. 1.9A). Prior to this study, LUH was well-characterised as activating *MUM2* for proper mucilage extrusion but how this was achieved was unknown. Consequently, three models were proposed (Fig. 1.9), with the de-repression model (Fig. 1.9C) initially pursued due to the similarity of LUH with LUG, a well-characterised transcriptional corepressor. While this led to the identification of MADS-box TFs as candidate *MUM2* repressors (**Chapter 3**) described in Fig. 1.9C, subsequent analysis instead strongly supported MADS-box TFs as recruiting LUH to *MUM2* for activation (Fig. 1.9A). Additionally, this thesis also identified the novel roles of MADS-box TFs in seed development ranging from the control of seed shape (SEP1, SEP3) through crossing *sep1* and *sep3* mutants and the control of seed size (AG) through amiRNA knockdown of AG within the seed coat mediated by seed coat-specific *GILT_{pro}* to overcome sterility issues associated with *ag* mutants.

Although the initial premise of this study was the activation of *MUM2* by LUH through a de-repression model (Fig. 1.9C), the identification of several MADS-box TFs as candidate *MUM2* repressors elevated in *luh* mutants was considered likely to be inconsistent with the mucilage extrusion defects observed in *sep3*, *shp2* and *shp1 shp2* mutants. Instead, given that these mutants possessed normal MSCs, mucilage extrusion defects in *sep3*, *shp2* and *shp1 shp2* indicated that SEP3, SHP1 and SHP2 have roles in regulating downstream effector enzymes for proper mucilage extrusion, and may instead function as *MUM2* activators. While the downstream effector enzyme(s) responsible for the mucilage defects seen in *sep3* and *shp1 shp2* mutants remain unidentified due to inconclusive carbohydrate analysis results, several mis-regulated MSC tool box genes were identified as potential causative factors in a *shp1 shp2* background, most notably *MUM2*, *ASD1* and *ASD2*. Given that *MUM2* seems

only partially responsible for mucilage defects in both *sep3* and *shp1 shp2* as inferred from experiments increasing MUM2 activity using heterologous promoters, it is likely that the other mis-regulated MSC tool box genes play a role in these mutants. Additionally, changes in the expression of some cell-wall components such as *FLA7*, *FLA11* and *FLA15* may lead to changes in mucilage extrusion seen in *shp1 shp2* mutants due to potential changes in cell wall architecture, especially since changes in *SOS5/FLA4* have been shown to lead to mucilage defects (Basu et al., 2016; Griffiths et al., 2014; Harpaz-Saad et al., 2011).

Decreased *MUM2* levels in the MSCs of *sep3* and *shp1 shp2* are consistent with SEP3, SHP1 and SHP2 being involved in the activation of *MUM2*. Unfortunately, this did not correspond to an increase in *t*-Gal residues in mucilage extracted from these mutants, which arises from the loss of MUM2 activity (Dean et al., 2007; Macquet et al., 2007b; Walker et al., 2011). It is possible that the reduction of *t*-Gal in these mutants was simply too subtle and below the detection limit of carbohydrate analysis, but strong enough to result in mucilage extrusion defects (see **Section 6.2** below). Nevertheless, other lines of evidence obtained strongly supported SEP3, SHP1 and SHP2 as *MUM2* activators *in planta*. Firstly, CArG box/CL motif mutations in the regulatory region of the *MUM2* intron resulted in decreased *MUM2* promoter activity in the MSC of both Col-0 and *luh* seeds. Secondly, *SHP2::SRDX* fusion motifs in Col-0 provided compelling evidence that SHP2 acts as a *MUM2* activator *in vivo* by producing dominant negative mutants with both mucilage extrusion defects and decreased *MUM2* expression. Lastly, SEP3, SHP1 and SHP2 all have the capacity to act as *MUM2* activators as demonstrated by transactivation assays, where these TFs promoted *MUM2* promoter activity *in planta*.

With SEP3, SHP1 and SHP2 firmly established as *MUM2* activators, there were two models left to pursue that described how these MADS-box TFs may activate *MUM2* in the MSCs. The first was a LUH-independent model and the second was a model dependent on LUH, either collectively as a complex (Fig. 1.9A) or as *MUM2* activators that are positively regulated by LUH (Fig. 1.9B). The observed increase of

SEP3, SHP1 and SHP2 in *luh* mutants seems to exclude the model in which these TFs are activated by LUH (Fig. 1.9B). Furthermore, finding physical interactions between LUH and SEP3 mediated by SEU in yeast assays provided evidence that MADS-box TFs may work in a complex with LUH to activate *MUM2* (Fig. 1.9A). This was further supported by BiFC assays, which showed that the formation of a LUH-SEU complex with SHP1 and SHP2 mediated by SEP3 is possible. Unfortunately, due to time constraints it was not possible to test whether these MADS-box TFs recruit LUH to *MUM2 I1*, although such experiments will need to be performed in the future. Additionally, the exact composition of MADS-box TFs in the proposed tetramer recruiting LUH to *MUM2* remains unresolved, but presumably includes a combination of SEP3, SHP1 and SHP2, as well as other MADS-box TFs that are expressed in the seed coat such as SEP1, AG and AP3.

Overall, this study has furthered our understanding of how LUH activates *MUM2* in the seed coat – via recruitment to *MUM2* by MADS-box TFs – and therefore firmly establishes MADS-box TFs SHP1, SHP2 and SEP3 as tier 1 TFs in the hierarchy of mucilage polysaccharides regulation based on the classification criteria specified by Golz (2018) (Fig. 1.6).

Other limitations of this study include a lack of direct evidence that MADS-box TFs bind to the identified CArG box/CL motifs and the unresolved composition of the MADS-box TF tetramer involved in *MUM2* activation. These limitations represent scope for further investigation as discussed in the next section. However, ChIP-seq data on floral tissue demonstrates that SEP3 can bind to *MUM2 I1* (Kaufmann et al., 2009; Pajoro et al., 2014), indicating that binding is possible.

Beyond the roles of MADS-box TFs in mucilage polysaccharides regulation, this study also identified novel roles for MADS-box TFs in the seed, such as the control of seed shape by SEP1 and SEP3 and the control of seed size by AG. Therefore, extending the study of MADS-box TFs in *Arabidopsis* seeds is important to advance understanding of seed developmental processes and may lead to the identification of targets of agronomic value, such as the potential application of AG in seed size. Last

but not least, the process of studying the role of AG in the *Arabidopsis* seed resulted in the identification and evaluation of seed coat specific tool *GILT_{pro}* and represents an important tool in seed biology.

6.2 Future directions

Although MADS-box TFs SEP3, SHP1 and SHP2 have been shown to be *MUM2* activators, direct evidence that these TFs can bind to the identified CArG box/CL motifs (CArG I, II, II and CL I; Fig. 5.1) in *MUM2 I1* has not been demonstrated due to a lack of time. This may be addressed by performing ChIP-PCR on wildtype seeds and showing that the resultant binding sites detected map onto the identified CArG box/CL motifs. As loss of *SHP2* gives rise to the most severe mucilage defects in single MADS-box TF mutants, SHP2 would be the ideal target for ChIP-PCR, particularly as a tagged version (*SHP2_{pro}:SHP2-VENUSn*) has been shown to induce complementation in *shp1 shp2* mutants (**Chapter 4, Section 4.2.1**). Alternatively, EMSA may be performed to show that MADS-box TFs can bind to the regulatory region of *MUM2* containing CArG box/CL motifs. Mutation of these motifs should prevent binding and therefore prove that these CArG box/CL motifs are necessary for the binding of MADS-box TFs to *MUM2*. Additionally, the exact combination of MADS-box TFs in the tetramer facilitating the recruitment of LUH to *MUM2* remains undefined. While this tetramer likely consists of SEP3, SHP1 and SHP2, it is unclear if there is another MADS-box TF member present, or if one of the three TFs are represented twice. Hence further study to discern the identity this tetramer is required which may involve crossing of *sep3 shp1 shp2* triple mutants to other MADS-box TF mutants such as *sep1* to look for enhanced mucilage extrusion defects. However, a major hurdle to this approach is the potential disruptions to flower and ovule development that may prevent seed formation and therefore prevent examination of mucilage defects in this mutant. Additionally, the loss of *SEP1* and *SEP3* leads to patterning defects resulting in abnormal seed shape (**Chapter 3, Section 3.2.4.1**), preventing the study of downstream mucilage extrusion. To circumvent these issues,

GILT_{pro} can be used to specifically knock-down *SEP1* expression in a *sep3 shp1 shp2* mutant background to evaluate the loss of *SEP1*, *SEP3*, *SHP1* and *SHP2* on mucilage extrusion in seeds.

In addition, the loss of *MUM2* in mutants lacking these TFs is only partially responsible for defects in mucilage extrusion (**Chapter 4, Section 4.3.3**). Unfortunately, due to the inconclusive results from the carbohydrate analysis, efforts to identify other downstream enzyme effectors contributing to mucilage defects in *shp1 shp2* and *sep3* mutants in this study failed. This can be addressed through a combination of repeating carbohydrate analyses on these mutants using the same protocol as described in Walker et al. (2011) to increase confidence in the reliability of results and by further examination of mucilage defects by immuno-TEM and glycan microarrays (see **Chapter 4, Section 4.3.3**). Assuming that the loss of *t*-Gal residues in *shp1 shp2* and *sep3* mutants are below the detectable threshold of carbohydrate analysis, a stronger mutant such as *sep3 shp1 shp2* may be used in such analysis to improve detection rates (**Chapter 4, Section 4.3.3**). Additionally, several candidate enzymatic factors impacting on *shp1 shp2* identified by RNA-seq such as *ASD1* and *ASD2* could be evaluated for the rescue of mucilage extrusion when heterologously expressed in *shp1 shp2* mutants, with rescue of mucilage extrusion indicating that these downstream enzyme effectors are indeed contributory to observed mucilage defects. While carbohydrate analysis has proven useful in analysing cell-wall related mutants, it may not reveal other subtleties such as a change in cell wall architecture. Hence, the disruption to *FLA7*, *FLA11* and *FLA15* in *shp1 shp2* mutants may be better studied with immuno-TEM.

Lastly, beyond the scope of this thesis is the role of MADS-box TFs in upstream seed development such as the control of seed shape and seed size. This would entail the characterisation of *sep1 sep3* mutants and *AG* knock-down mutants to further understand how *SEP1* and *SEP3* control seed shape and how *AG* controls seed size respectively. For example, histological sections on *sep1 sep3* double mutants can provide insight as to where disruptions to normal seed development occurs and if

SEP1 and SEP3 are involved in regulating the funiculus, inner and outer integument development in ovules like similarly shaped *stk* and *ats* mutants (**Chapter 3, Section 3.3.3**). Similarly, histological sections on AG knock-down mutants can show if integument cell sizes are enlarged like phenotypically similar mutants *da1*, *da2*, *arf2* and *ap2*, or are due to disruptions to other developmental processes. Next, RNA-seq on *sep1 sep3* and AG knock-down mutants can generate a list of robust DEGs as candidate genes regulated by SEP1/SEP3 and AG for the control of seed shape and size respectively. Mutants of these candidate genes can then be crossed with *sep1 sep3*, along with mutants known to produce abnormal seed shape such as *stk*, *ats*, *gl2* and *ap2* to evaluate for genetic interactions with SEP1 and SEP3, with additive phenotypes indicating that these genes work in separate pathways. Likewise, this can be applied to AG knock down mutants with *da1*, *da2*, *arf2* and *ap2* mutants. Finally, AG knock-down mutants should also be assayed for any increase in oil production in seeds, as this represents a potential economic interest if applicable to crops such as canola.

Chapter 7

References

References

- Adam, C. L., Williams, P. A., Garden, K. E., Thomson, L. M., & Ross, A. W. (2015). Dose-dependent effects of a soluble dietary fibre (pectin) on food intake, adiposity, gut hypertrophy and gut satiety hormone secretion in rats. *PLoS One*, 10(1), e0115438. doi:10.1371/journal.pone.0115438
- Alvarez-Buylla, E. R., Pelaz, S., Liljegren, S. J., Gold, S. E., Burgeff, C., Ditta, G. S., Ribas de Pouplana, L., Martinez-Castilla, L., & Yanofsky, M. F. (2000). An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc Natl Acad Sci U S A*, 97(10), 5328-5333.
- Alvarez, J. P., Pekker, I., Goldshmidt, A., Blum, E., Amsellem, Z., & Eshed, Y. (2006). Endogenous and synthetic microRNAs stimulate simultaneous, efficient, and localized regulation of multiple targets in diverse species. *Plant Cell*, 18(5), 1134-1151. doi:10.1105/tpc.105.040725
- Arsovski, A. A., Haughn, G. W., & Western, T. L. (2010). Seed coat mucilage cells of *Arabidopsis thaliana* as a model for plant cell wall research. *Plant Signal Behav*, 5(7), 796-801.
- Arsovski, A. A., Popma, T. M., Haughn, G. W., Carpita, N. C., McCann, M. C., & Western, T. L. (2009a). AtBXL1 Encodes a Bifunctional beta-D-Xylosidase/alpha-L-Arabinofuranosidase Required for Pectic Arabinan Modification in *Arabidopsis* Mucilage Secretory Cells. *Plant Physiol*, 150(3), 1219-1234. doi:DOI 10.1104/pp.109.138388
- Arsovski, A. A., Villota, M. M., Rowland, O., Subramaniam, R., & Western, T. L. (2009b). MUM ENHANCERS are important for seed coat mucilage production and mucilage secretory cell differentiation in *Arabidopsis thaliana*. *J Exp Bot*, 60(9), 2601-2612. doi:Doi 10.1093/Jxb/Erp102
- Balanza, V., Roig-Villanova, I., Di Marzo, M., Masiero, S., & Colombo, L. (2016). Seed abscission and fruit dehiscence required for seed dispersal rely on similar genetic networks. *Development*, 143(18), 3372-3381. doi:10.1242/dev.135202

- Bao, F., Azhakanandam, S., & Franks, R. G. (2010). SEUSS and SEUSS-LIKE transcriptional adaptors regulate floral and embryonic development in Arabidopsis. *Plant Physiol*, 152(2), 821-836. doi:10.1104/pp.109.146183
- Basu, D., Tian, L., Debrosse, T., Poirier, E., Emch, K., Herock, H., Travers, A., & Showalter, A. M. (2016). Glycosylation of a Fasciclin-Like Arabinogalactan-Protein (SOS5) Mediates Root Growth and Seed Mucilage Adherence via a Cell Wall Receptor-Like Kinase (FEI1/FEI2) Pathway in Arabidopsis. *PLoS One*, 11(1), e0145092. doi:10.1371/journal.pone.0145092
- Baudry, A., Caboche, M., & Lepiniec, L. (2006). TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in Arabidopsis thaliana. *Plant Journal*, 46(5), 768-779. doi:10.1111/j.1365-313X.2006.02733.x
- Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., & Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *Plant Journal*, 39(3), 366-380. doi:10.1111/j.1365-313X.2004.02138.x
- Becker, A., & Theissen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol*, 29(3), 464-489.
- Beeckman, T., De Rycke, R., Viane, R., & Inze, D. (2000). Histological Study of Seed Coat Development in Arabidopsis thaliana. *J. Plant Res.*, 113, 139-148.
- Belmonte, M. F., Kirkbride, R. C., Stone, S. L., Pelletier, J. M., Bui, A. Q., Yeung, E. C., Hashimoto, M., Fei, J., Harada, C. M., Munoz, M. D., Le, B. H., Drews, G. N., Brady, S. M., Goldberg, R. B., & Harada, J. J. (2013). Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proc Natl Acad Sci U S A*, 110(5), E435-444. doi:10.1073/pnas.1222061110
- Ben-Tov, D., Abraham, Y., Stav, S., Thompson, K., Loraine, A., Elbaum, R., de Souza, A., Pauly, M., Kieber, J. J., & Harpaz-Saad, S. (2015). COBRA-LIKE2, a member of the glycosylphosphatidylinositol-anchored COBRA-

- LIKE family, plays a role in cellulose deposition in arabidopsis seed coat mucilage secretory cells. *Plant Physiol*, 167(3), 711-724. doi:10.1104/pp.114.240671
- Bhargava, A., Ahad, A., Wang, S., Mansfield, S. D., Haughn, G. W., Douglas, C. J., & Ellis, B. E. (2013). The interacting MYB75 and KNAT7 transcription factors modulate secondary cell wall deposition both in stems and seed coat in Arabidopsis. *Planta*, 237(5), 1199-1211. doi:10.1007/s00425-012-1821-9
- Blackwood, E. M., & Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science*, 281(5373), 60-63.
- Bostock, M., & Heer, J. (2009). Protovis: a graphical toolkit for visualization. *IEEE Trans Vis Comput Graph*, 15(6), 1121-1128. doi:10.1109/TVCG.2009.174
- Boter, M., Golz, J. F., Gimenez-Ibanez, S., Fernandez-Barbero, G., Franco-Zorrilla, J. M., & Solano, R. (2015). FILAMENTOUS FLOWER Is a Direct Target of JAZ3 and Modulates Responses to Jasmonate. *Plant Cell*, 27(11), 3160-3174. doi:10.1105/tpc.15.00220
- Bowman, J. L., Baum, S. F., Eshed, Y., Putterill, J., & Alvarez, J. (1999). Molecular genetics of gynoecium development in Arabidopsis. *Curr Top Dev Biol*, 45, 155-205.
- Bozolla, J. J. (2007). Conventional Specimen Preparation Techniques for Scanning Electron Microscopy of Biological Specimens. In J. Kuo (Ed.), *Electron Microscopy: Methods and Protocols* (2nd ed.). Crawley, Australia: Humana Press.
- Brouns, F., Theuwissen, E., Adam, A., Bell, M., Berger, A., & Mensink, R. P. (2012). Cholesterol-lowering properties of different pectin types in mildly hypercholesterolemic men and women. *Eur J Clin Nutr*, 66(5), 591-599. doi:10.1038/ejcn.2011.208
- Bui, M., Lim, N., Sijacic, P., & Liu, Z. (2011). LEUNIG_HOMOLOG and LEUNIG regulate seed mucilage extrusion in Arabidopsis. *Journal of Integrative Plant Biology*, 53(5), 399-408. doi:10.1111/j.1744-7909.2011.01036.x

- Caffall, K. H., & Mohnen, D. (2009). The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res*, 344(14), 1879-1900. doi:10.1016/j.carres.2009.05.021
- Camacho Londono, J., & Philipp, S. E. (2016). A reliable method for quantification of splice variants using RT-qPCR. *BMC Mol Biol*, 17, 8. doi:10.1186/s12867-016-0060-1
- Cazzonelli, C. I., & Velten, J. (2006). An in vivo, luciferase-based, Agrobacterium-infiltration assay system: implications for post-transcriptional gene silencing. *Planta*, 224(3), 582-597. doi:10.1007/s00425-006-0250-z
- Chavez Montes, R. A., Ranocha, P., Martinez, Y., Minic, Z., Jouanin, L., Marquis, M., Saulnier, L., Fulton, L. M., Cobbett, C. S., Bitton, F., Renou, J. P., Jauneau, A., & Goffner, D. (2008). Cell wall modifications in Arabidopsis plants with altered alpha-L-arabinofuranosidase activity. *Plant Physiol*, 147(1), 63-77. doi:10.1104/pp.107.110023
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1), 117-143.
- Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal*, 16(6), 735-743.
- Cnops, G., Jover-Gil, S., Peters, J. L., Neyt, P., De Block, S., Robles, P., Ponce, M. R., Gerats, T., Micol, J. L., & Van Lijsebettens, M. (2004). The rotunda2 mutants identify a role for the LEUNIG gene in vegetative leaf morphogenesis. *J Exp Bot*, 55(402), 1529-1539. doi:10.1093/jxb/erh165
- Coen, E. S., & Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6339), 31-37. doi:10.1038/353031a0
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol*, 17, 13. doi:10.1186/s13059-016-0881-8

- Conner, J., & Liu, Z. (2000). LEUNIG, a putative transcriptional corepressor that regulates AGAMOUS expression during flower development. *Proc Natl Acad Sci U S A*, 97(23), 12902-12907. doi:10.1073/pnas.230352397
- de Folter, S., Immink, R. G. H., Kieffer, M., Parenicova, L., Henz, S. R., Weigel, D., Busscher, M., Kooiker, M., Colombo, L., Kater, M. M., Davies, B., & Angenent, G. C. (2005). Comprehensive interaction map of the Arabidopsis MADS box transcription factors. *Plant Cell*, 17(5), 1424-1433. doi:10.1105/tpc.105.031831
- de Folter, S., Urbanus, S. L., van Zuijlen, L. G., Kaufmann, K., & Angenent, G. C. (2007). Tagging of MADS domain proteins for chromatin immunoprecipitation. *Bmc Plant Biology*, 7, 47. doi:10.1186/1471-2229-7-47
- Dean, G., Cao, Y., Xiang, D., Provart, N. J., Ramsay, L., Ahad, A., White, R., Selvaraj, G., Datla, R., & Haughn, G. (2011). Analysis of gene expression patterns during seed coat development in Arabidopsis. *Mol Plant*, 4(6), 1074-1091. doi:10.1093/mp/ssr040
- Dean, G., Zheng, H., Tewari, J., Huang, J., Young, D. S., Hwang, Y. T., Western, T. L., Carpita, N. C., McCann, M. C., Mansfield, S. D., & Haughn, G. W. (2007). The Arabidopsis MUM2 gene encodes a beta-galactosidase required for the production of seed coat mucilage with correct hydration properties. *Plant Cell*, 19(12), 4007-4021. doi:10.1105/tpc.107.050609
- Dey, N., Sarkar, S., Acharya, S., & Maiti, I. B. (2015). Synthetic promoters in planta. *Planta*, 242(5), 1077-1094. doi:10.1007/s00425-015-2377-2
- Doblin, M. S., Johnson, K. L., Humphries, J., Newbigin, E. J., & Bacic, A. (2014). Are designer plant cell walls a realistic aspiration or will the plasticity of the plant's metabolism win out? *Curr Opin Biotechnol*, 26, 108-114. doi:10.1016/j.copbio.2013.11.012
- Doblin, M. S., Pettolino, F., & Bacic, A. (2010). Plant cell walls: the skeleton of the plant world. *Functional Plant Biology*, 37(5), 357-381. doi:10.1071/Fp09279

- 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(6), 1349.
- Ehlers, K., Bhide, A. S., Tekleyohans, D. G., Wittkop, B., Snowdon, R. J., & Becker, A. (2016). The MADS Box Genes ABS, SHP1, and SHP2 Are Essential for the Coordination of Cell Divisions in Ovule and Seed Coat Development and for Endosperm Formation in *Arabidopsis thaliana*. *PLoS One*, 11(10), e0165075. doi:10.1371/journal.pone.0165075
- Esfandiari, E., Jin, Z. Q., Abdeen, A., Griffiths, J. S., Western, T. L., & Haughn, G. W. (2013). Identification and analysis of an outer-seed-coat-specific promoter from *Arabidopsis thaliana*. *Plant Molecular Biology*, 81(1-2), 93-104. doi:DOI 10.1007/s11103-012-9984-0
- Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048. doi:10.1093/bioinformatics/btw354
- Ezquer, I., Mizzotti, C., Nguema-Ona, E., Gotte, M., Beauzamy, L., Viana, V. E., Dubrulle, N., Costa de Oliveira, A., Caporali, E., Koroney, A. S., Boudaoud, A., Driouich, A., & Colombo, L. (2016). The Developmental Regulator SEEDSTICK Controls Structural and Mechanical Properties of the *Arabidopsis* Seed Coat. *Plant Cell*, 28(10), 2478-2492. doi:10.1105/tpc.16.00454
- Fang, Z., & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Brief Bioinform*, 12(3), 280-287. doi:10.1093/bib/bbr004
- Faure, J. E., & Dumas, C. (2001). Fertilization in flowering plants. New approaches for an old story. *Plant Physiol*, 125(1), 102-104.
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., & Colombo, L. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell*, 15(11), 2603-2611. doi:10.1105/tpc.015123

- Ferrandiz, C., Pelaz, S., & Yanofsky, M. F. (1999). Control of carpel and fruit development in *Arabidopsis*. *Annu Rev Biochem*, 68, 321-354. doi:10.1146/annurev.biochem.68.1.321
- Fields, S., & Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230), 245-246. doi:10.1038/340245a0
- Figueiredo, D. D., Batista, R. A., Roszak, P. J., Hennig, L., & Kohler, C. (2016). Auxin production in the endosperm drives seed coat development in *Arabidopsis*. *Elife*, 5. doi:10.7554/eLife.20542
- Francoz, E., Ranocha, P., Burlat, V., & Dunand, C. (2015). *Arabidopsis* seed mucilage secretory cells: regulation and dynamics. *Trends Plant Sci*, 20(8), 515-524. doi:10.1016/j.tplants.2015.04.008
- Freeman, W. M., Walker, S. J., & Vrana, K. E. (1999). Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26(1), 112-122, 124-115.
- Fried, M. G., & Daugherty, M. A. (1998). Electrophoretic analysis of multiple protein-DNA interactions. *Electrophoresis*, 19(8-9), 1247-1253. doi:10.1002/elps.1150190804
- Garcia, D., Fitz Gerald, J. N., & Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell*, 17(1), 52-60. doi:10.1105/tpc.104.027136
- Garner, M. M., & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res*, 9(13), 3047-3060.
- Gelvin, S. B. (2003). *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev*, 67(1), 16-37, table of contents.
- Gerlitz, G., Darhin, E., Giorgio, G., Franco, B., & Reiner, O. (2005). Novel functional features of the Lis-H domain: role in protein dimerization, half-life and cellular localization. *Cell Cycle*, 4(11), 1632-1640. doi:10.4161/cc.4.11.2151

- Gleave, A. P. (1992). A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology*, 20(6), 1203-1207.
- Golz, J. F., Allen, P. J., Li, S. F., Parish, R. W., Jayawardana, N. U., Bacic, A., & Doblin, M. S. (2018). Layers of regulation - Insights into the role of transcription factors controlling mucilage production in the Arabidopsis seed coat. *Plant Science*, 272, 179-192. doi:10.1016/j.plantsci.2018.04.021
- Gomez, L. D., Steele-King, C. G., & McQueen-Mason, S. J. (2008). Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist*, 178(3), 473-485. doi:10.1111/j.1469-8137.2008.02422.x
- Gonzalez, A., Brown, M., Hatlestad, G., Akhavan, N., Smith, T., Hembd, A., Moore, J., Montes, D., Mosley, T., Resendez, J., Nguyen, H., Wilson, L., Campbell, A., Sudarshan, D., & Lloyd, A. (2016). TTG2 controls the developmental regulation of seed coat tannins in Arabidopsis by regulating vacuolar transport steps in the proanthocyanidin pathway. *Dev Biol*, 419(1), 54-63. doi:10.1016/j.ydbio.2016.03.031
- Gonzalez, A., Mendenhall, J., Huo, Y., & Lloyd, A. (2009). TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Dev Biol*, 325(2), 412-421. doi:10.1016/j.ydbio.2008.10.005
- Gonzalez , D., Bowen, A. J., Carroll, T. S., & Conlan, R. S. (2007). The transcription corepressor LEUNIG interacts with the histone deacetylase HDA19 and mediator components MED14 (SWP) and CDK8 (HEN3) to repress transcription. *Mol Cell Biol*, 27(15), 5306-5315. doi:10.1128/MCB.01912-06
- Gramzow, L., Ritz, M. S., & Theissen, G. (2010). On the origin of MADS-domain transcription factors. *Trends Genet*, 26(4), 149-153. doi:10.1016/j.tig.2010.01.004
- Gramzow, L., & Theissen, G. (2010). A hitchhiker's guide to the MADS world of plants. *Genome Biol*, 11(6), 214. doi:10.1186/gb-2010-11-6-214
- Gregis, V., Sessa, A., Dorca-Fornell, C., & Kater, M. M. (2009). The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B

- and C floral homeotic genes. *Plant Journal*, 60(4), 626-637. doi:10.1111/j.1365-313X.2009.03985.x
- Griffiths, J. S., Crepeau, M. J., Ralet, M. C., Seifert, G. J., & North, H. M. (2016). Dissecting Seed Mucilage Adherence Mediated by FEI2 and SOS5. *Front Plant Sci*, 7, 1073. doi:10.3389/fpls.2016.01073
- Griffiths, J. S., & North, H. M. (2017). Sticking to cellulose: exploiting Arabidopsis seed coat mucilage to understand cellulose biosynthesis and cell wall polysaccharide interactions. *New Phytologist*, 214(3), 959-966. doi:10.1111/nph.14468
- Griffiths, J. S., Sola, K., Kushwaha, R., Lam, P., Tateno, M., Young, R., Voiniciuc, C., Dean, G., Mansfield, S. D., DeBolt, S., & Haughn, G. W. (2015). Unidirectional movement of cellulose synthase complexes in Arabidopsis seed coat epidermal cells deposit cellulose involved in mucilage extrusion, adherence, and ray formation. *Plant Physiol*, 168(2), 502-520. doi:10.1104/pp.15.00478
- Griffiths, J. S., Tsai, A. Y., Xue, H., Voiniciuc, C., Sola, K., Seifert, G. J., Mansfield, S. D., & Haughn, G. W. (2014). SALT-OVERLY SENSITIVE5 Mediates Arabidopsis Seed Coat Mucilage Adherence and Organization through Pectins. *Plant Physiol*, 165(3), 991-1004. doi:10.1104/pp.114.239400
- Grigorova, B., Mara, C., Hollender, C., Sijacic, P., Chen, X. M., & Liu, Z. C. (2011). LEUNIG and SEUSS co-repressors regulate miR172 expression in Arabidopsis flowers. *Development*, 138(12), 2451-2456. doi:10.1242/Dev.058362
- Harholt, J., Suttangkakul, A., & Vibe Scheller, H. (2010). Biosynthesis of pectin. *Plant Physiol*, 153(2), 384-395. doi:10.1104/pp.110.156588
- Harpaz-Saad, S., McFarlane, H. E., Xu, S. L., Divi, U. K., Forward, B., Western, T. L., & Kieber, J. J. (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 Journal*, 68(6), 941-953. doi:10.1111/j.1365-313X.2011.04760.x

- Hatem, A., Bozdag, D., Toland, A. E., & Catalyurek, U. V. (2013). Benchmarking short sequence mapping tools. *BMC Bioinformatics*, 14, 184. doi:10.1186/1471-2105-14-184
- Haughn, G., & Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. *Trends Plant Sci*, 10(10), 472-477. doi:10.1016/j.tplants.2005.08.005
- Haughn, G. W., & Western, T. L. (2012). Arabidopsis Seed Coat Mucilage is a Specialized Cell Wall that Can be Used as a Model for Genetic Analysis of Plant Cell Wall Structure and Function. *Front Plant Sci*, 3, 64. doi:10.3389/fpls.2012.00064
- He, J. B., Zhao, X. H., Du, P. Z., Zeng, W., Beahan, C. T., Wang, Y. Q., Li, H. L., Bacic, A., & Wu, A. M. (2018). KNAT7 positively regulates xylan biosynthesis by directly activating IRX9 expression in Arabidopsis. *Journal of Integrative Plant Biology*, 60(6), 514-528. doi:10.1111/jipb.12638
- Heisler, M. G., Atkinson, A., Bylstra, Y. H., Walsh, R., & Smyth, D. R. (2001). SPATULA, a gene that controls development of carpel margin tissues in Arabidopsis, encodes a bHLH protein. *Development*, 128(7), 1089-1098.
- Heisler, M. G., Ohno, C., Das, P., Sieber, P., Reddy, G. V., Long, J. A., & Meyerowitz, E. M. (2005). Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the Arabidopsis inflorescence meristem. *Current Biology*, 15(21), 1899-1911. doi:10.1016/j.cub.2005.09.052
- Hellens, R. P., Allan, A. C., Friel, E. N., Bolitho, K., Grafton, K., Templeton, M. D., Karunairetnam, S., Gleave, A. P., & Laing, W. A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods*, 1, 13. doi:10.1186/1746-4811-1-13
- Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S., & Mullineaux, P. M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, 42(6), 819-832.

- Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols*, 2(8), 1849-1861. doi:10.1038/nprot.2007.249
- Hengen, P. N. (1997). False positives from the yeast two-hybrid system. *Trends Biochem Sci*, 22(1), 33-34.
- Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Munster, T., & Theissen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. *Mol Biol Evol*, 19(6), 801-814. doi:10.1093/oxfordjournals.molbev.a004137
- Hernandez-Garcia, C. M., Martinelli, A. P., Bouchard, R. A., & Finer, J. J. (2009). A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. *Plant Cell Reports*, 28(5), 837-849. doi:10.1007/s00299-009-0681-7
- Hiratsu, K., Matsui, K., Koyama, T., & Ohme-Takagi, M. (2003). Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*. *Plant Journal*, 34(5), 733-739.
- Honma, T., & Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, 409(6819), 525-529. doi:10.1038/35054083
- Horstman, A., Tonaco, I. A. N., Boutilier, K., & Immink, R. G. H. (2014). A Cautionary Note on the Use of Split-YFP/BiFC in Plant Protein-Protein Interaction Studies. *International Journal of Molecular Sciences*, 15(6), 9628-9643. doi:10.3390/ijms15069628
- Hu, R., Li, J., Wang, X., Zhao, X., Yang, X., Tang, Q., He, G., Zhou, G., & Kong, Y. (2016). Xylan synthesized by Irregular Xylem 14 (IRX14) maintains the structure of seed coat mucilage in *Arabidopsis*. *J Exp Bot*, 67(5), 1243-1257. doi:10.1093/jxb/erv510
- Huang, J., DeBowles, D., Esfandiari, E., Dean, G., Carpita, N. C., & Haughn, G. W. (2011). The *Arabidopsis* transcription factor LUH/MUM1 is required for extrusion of seed coat mucilage. *Plant Physiol*, 156(2), 491-502. doi:10.1104/pp.111.172023

- Iglesias-Fernandez, R., Barrero-Sicilia, C., Carrillo-Barral, N., Onate-Sanchez, L., & Carbonero, P. (2013). *Arabidopsis thaliana* bZIP44: a transcription factor affecting seed germination and expression of the mannanase-encoding gene AtMAN7. *Plant Journal*, 74(5), 767-780. doi:10.1111/tpj.12162
- Immink, R. G., Ferrario, S., Busscher-Lange, J., Kooiker, M., Busscher, M., & Angenent, G. C. (2003). Analysis of the petunia MADS-box transcription factor family. *Molecular Genetics and Genomics*, 268(5), 598-606. doi:10.1007/s00438-002-0781-3
- Immink, R. G., Tonaco, I. A., de Folter, S., Shchennikova, A., van Dijk, A. D., Busscher-Lange, J., Borst, J. W., & Angenent, G. C. (2009). SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome Biol*, 10(2), R24. doi:10.1186/gb-2009-10-2-r24
- Irish, V. F., & Sussex, I. M. (1990). Function of the *Apetala-1* Gene during *Arabidopsis* Floral Development. *Plant Cell*, 2(8), 741-753.
- Jayawardana, N. U. (2016). *The role of LEUNIG_HOMOLOG in regulating mucilage release from the Arabidopsis testa*. (Doctoral Thesis), University of Melbourne, Victoria, Australia.
- Jetha, K., Theissen, G., & Melzer, R. (2014). *Arabidopsis* SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. *Nucleic Acids Res*, 42(17), 10927-10942. doi:10.1093/nar/gku755
- Jofuku, K. D., den Boer, B. G., Van Montagu, M., & Okamuro, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell*, 6(9), 1211-1225.
- Jofuku, K. D., Omidyar, P. K., Gee, Z., & Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc Natl Acad Sci U S A*, 102(8), 3117-3122. doi:10.1073/pnas.0409893102
- Johnson, C. S., Kolevski, B., & Smyth, D. R. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell*, 14(6), 1359-1375.
- Jordan, B. R. (2006). *The molecular biology and biotechnology of flowering* (2nd ed.). Wallingford, Oxfordshire, UK ; Cambridge, MA: CAB Pub.

- Kaufmann, K., Muino, J. M., Jauregui, R., Airoidi, C. A., Smaczniak, C., Krajewski, P., & Angenent, G. C. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol*, 7(4), e1000090. doi:10.1371/journal.pbio.1000090
- Khan, D., Millar, J. L., Girard, I. J., & Belmonte, M. F. (2014). Transcriptional circuitry underlying seed coat development in Arabidopsis. *Plant Science*, 219-220, 51-60. doi:10.1016/j.plantsci.2014.01.004
- Kim, J. B., & Carpita, N. C. (1992). Changes in Esterification of the Uronic Acid Groups of Cell Wall Polysaccharides during Elongation of Maize Coleoptiles. *Plant Physiol*, 98(2), 646-653.
- Kodama, Y., & Hu, C. D. (2012). Bimolecular fluorescence complementation (BiFC): A 5-year update and future perspectives. *Biotechniques*, 53(5), 285-+. doi:10.2144/0000113943
- Kong, Y. Z., Zhou, G. K., Abdeen, A. A., Schafhauser, J., Richardson, B., Atmodjo, M. A., Jung, J. Y., Wicker, L., Mohnen, D., Western, T., & Hahn, M. G. (2013). GALACTURONOSYLTRANSFERASE-LIKE5 Is Involved in the Production of Arabidopsis Seed Coat Mucilage. *Plant Physiol*, 163(3), 1203-1217. doi:DOI 10.1104/pp.113.227041
- Koornneef, M. (1981). The complex syndrome of ttg mutants. *Arabid Inf Serv*, 18, 45-51.
- Krizek, B. A., & Meyerowitz, E. M. (1996). The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development*, 122(1), 11-22.
- Kulich, I., Cole, R., Drdova, E., Cvrckova, F., Soukup, A., Fowler, J., & Zarsky, V. (2010). Arabidopsis exocyst subunits SEC8 and EXO70A1 and exocyst interactor ROH1 are involved in the localized deposition of seed coat pectin. *New Phytologist*, 188(2), 615-625. doi:DOI 10.1111/j.1469-8137.2010.03372.x
- Kunieda, T., Mitsuda, N., Ohme-Takagi, M., Takeda, S., Aida, M., Tasaka, M., Kondo, M., Nishimura, M., & Hara-Nishimura, I. (2008). NAC family

- proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in Arabidopsis. *Plant Cell*, 20(10), 2631-2642. doi:10.1105/tpc.108.060160
- Kunieda, T., Shimada, T., Kondo, M., Nishimura, M., Nishitani, K., & Hara-Nishimura, I. (2013). Spatiotemporal Secretion of PEROXIDASE36 Is Required for Seed Coat Mucilage Extrusion in Arabidopsis. *Plant Cell*, 25(4), 1355-1367. doi:DOI 10.1105/tpc.113.110072
- Lau, E., & Bacic, A. (1993). Capillary gas chromatography of partially methylated alditol acetates on a high-polarity, cross-linked, fused-silica BPX70 column. *J Chromatogr*, 637, 100-103.
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, 15(2), R29. doi:10.1186/gb-2014-15-2-r29
- Lee, D. K., Geisler, M., & Springer, P. S. (2009). LATERAL ORGAN FUSION1 and LATERAL ORGAN FUSION2 function in lateral organ separation and axillary meristem formation in Arabidopsis. *Development*, 136(14), 2423-2432. doi:10.1242/dev.031971
- Lee, J., & Lee, I. (2010). Regulation and function of SOC1, a flowering pathway integrator. *J Exp Bot*, 61(9), 2247-2254. doi:10.1093/jxb/erq098
- Lee, J. E. (2014). *Regulation of shoot apical meristem formation and cotyledon development by SEUSS and SEUSS-LIKE 2 in Arabidopsis*.
- Lee, J. E., & Golz, J. F. (2012). Diverse roles of Groucho/Tup1 co-repressors in plant growth and development. *Plant Signal Behav*, 7(1), 86-92. doi:10.4161/psb.7.1.18377
- Lee, N., Park, J., Kim, K., & Choi, G. (2015). The Transcriptional Coregulator LEUNIG_HOMOLOG Inhibits Light-Dependent Seed Germination in Arabidopsis. *Plant Cell*, 27(8), 2301-2313. doi:10.1105/tpc.15.00444
- Leon-Kloosterziel, K. M., Keijzer, C. J., & Koornneef, M. (1994). A Seed Shape Mutant of Arabidopsis That Is Affected in Integument Development. *Plant Cell*, 6(3), 385-392. doi:10.1105/tpc.6.3.385

- Li, P., Piao, Y., Shon, H. S., & Ryu, K. H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC Bioinformatics*, 16, 347. doi:10.1186/s12859-015-0778-7
- Li, S., Bashline, L., Lei, L., & Gu, Y. (2014). Cellulose synthesis and its regulation. *Arabidopsis Book*, 12, e0169. doi:10.1199/tab.0169
- Li, S., Milliken, O. N., Pham, H., Seyit, R., Napoli, R., Preston, J., Koltunow, A. M., & Parish, R. W. (2009). The Arabidopsis MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell*, 21(1), 72-89. doi:10.1105/tpc.108.063503
- Li, Y., Bhargava, A., Qiang, W., Friedmann, M. C., Forneris, N., Savidge, R. A., Johnson, L. A., Mansfield, S. D., Ellis, B. E., & Douglas, C. J. (2012). The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in Populus. *New Phytologist*, 194(1), 102-115. doi:10.1111/j.1469-8137.2011.04016.x
- Li, Y., Zheng, L., Corke, F., Smith, C., & Bevan, M. W. (2008). Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. *Genes Dev*, 22(10), 1331-1336. doi:10.1101/gad.463608
- Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*, 41(10), e108. doi:10.1093/nar/gkt214
- Liljegren, S. J., Ditta, G. S., Eshed, H. Y., Savidge, B., Bowman, J. L., & Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in Arabidopsis. *Nature*, 404(6779), 766-770.
- Liu, S., Lin, L., Jiang, P., Wang, D., & Xing, Y. (2011). A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Res*, 39(2), 578-588. doi:10.1093/nar/gkq817
- Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*, 30(3), 301-304. doi:10.1093/bioinformatics/btt688

- Liu, Z., & Karmarkar, V. (2008). Groucho/Tup1 family co-repressors in plant development. *Trends Plant Sci*, 13(3), 137-144. doi:10.1016/j.tplants.2007.12.005
- Liu, Z., & Meyerowitz, E. M. (1995). LEUNIG regulates AGAMOUS expression in Arabidopsis flowers. *Development*, 121(4), 975-991.
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4), 402-408. doi:10.1006/meth.2001.1262
- Lu, S., Du, X., Lu, W., Chong, K., & Meng, Z. (2007). Two AGAMOUS-like MADS-box genes from Taihangia rupestris (Rosaceae) reveal independent trajectories in the evolution of class C and class D floral homeotic functions. *Evol Dev*, 9(1), 92-104. doi:10.1111/j.1525-142X.2006.00140.x
- Lydall, D., Ammerer, G., & Nasmyth, K. (1991). A new role for MCM1 in yeast: cell cycle regulation of SW15 transcription. *Genes Dev*, 5(12B), 2405-2419.
- Ma, K., Chan, J. K., Zhu, G., & Wu, Z. (2005). Myocyte enhancer factor 2 acetylation by p300 enhances its DNA binding activity, transcriptional activity, and myogenic differentiation. *Mol Cell Biol*, 25(9), 3575-3582. doi:10.1128/MCB.25.9.3575-3582.2005
- Macquet, A., Ralet, M. C., Kronenberger, J., Marion-Poll, A., & North, H. M. (2007a). In situ, chemical and macromolecular study of the composition of Arabidopsis thaliana seed coat mucilage. *Plant Cell Physiol*, 48(7), 984-999. doi:10.1093/pcp/pcm068
- Macquet, A., Ralet, M. C., Loudet, O., Kronenberger, J., Mouille, G., Marion-Poll, A., & North, H. M. (2007b). A naturally occurring mutation in an Arabidopsis accession affects a beta-D-galactosidase that increases the hydrophilic potential of rhamnogalacturonan I in seed mucilage. *Plant Cell*, 19(12), 3990-4006. doi:10.1105/tpc.107.050179
- Mandaokar, A., Thines, B., Shin, B., Lange, B. M., Choi, G., Koo, Y. J., Yoo, Y. J., Choi, Y. D., Choi, G., & Browse, J. (2006). Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant Journal*, 46(6), 984-1008. doi:10.1111/j.1365-313X.2006.02756.x

- Masiero, S., Colombo, L., Grini, P. E., Schnittger, A., & Kater, M. M. (2011). The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell*, 23(3), 865-872. doi:10.1105/tpc.110.081737
- Matsui, K., Hiratsu, K., Koyama, T., Tanaka, H., & Ohme-Takagi, M. (2005). A chimeric AtMYB23 repressor induces hairy roots, elongation of leaves and stems, and inhibition of the deposition of mucilage on seed coats in Arabidopsis. *Plant Cell Physiol*, 46(1), 147-155. doi:10.1093/pcp/pci005
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*, 40(10), 4288-4297. doi:10.1093/nar/gks042
- McCarthy, D. J., & Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 25(6), 765-771. doi:10.1093/bioinformatics/btp053
- McFarlane, H. E., Gendreau, D., & Western, T. L. (2014). Seed coat ruthenium red staining assay. *bio-protocol*, 4(7), 1-4.
- McFarlane, H. E., Watanabe, Y., Gendreau, D., Carruthers, K., Levesque-Tremblay, G., Haughn, G. W., Bhalerao, R. P., & Samuels, L. (2013). Cell wall polysaccharides are mislocalized to the Vacuole in echidna mutants. *Plant Cell Physiol*, 54(11), 1867-1880. doi:10.1093/pcp/pct129
- Mehla, J., Caufield, J. H., Sakhawalkar, N., & Uetz, P. (2017). A Comparison of Two-Hybrid Approaches for Detecting Protein-Protein Interactions. *Methods Enzymol*, 586, 333-358. doi:10.1016/bs.mie.2016.10.020
- Mendes, M. A., Guerra, R. F., Berns, M. C., Manzo, C., Masiero, S., Finzi, L., Kater, M. M., & Colombo, L. (2013). MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in Arabidopsis. *Plant Cell*, 25(7), 2560-2572. doi:10.1105/tpc.112.108688
- Mendu, V., Griffiths, J. S., Persson, S., Stork, J., Downie, A. B., Voiniciuc, C., Haughn, G. W., & DeBolt, S. (2011). Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. *Plant Physiol*, 157(1), 441-453. doi:10.1104/pp.111.179069

- Mizukami, Y., & Fischer, R. L. (2000). Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proc Natl Acad Sci U S A*, 97(2), 942-947.
- Mizzotti, C., Ezquer, I., Paolo, D., Rueda-Romero, P., Guerra, R. F., Battaglia, R., Rogachev, I., Aharoni, A., Kater, M. M., Caporali, E., & Colombo, L. (2014). SEEDSTICK is a master regulator of development and metabolism in the Arabidopsis seed coat. *PLoS Genet*, 10(12), e1004856. doi:10.1371/journal.pgen.1004856
- Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R., & Colombo, L. (2012). The MADS box genes SEEDSTICK and ARABIDOPSIS Bsisiter play a maternal role in fertilization and seed development. *Plant Journal*, 70(3), 409-420. doi:10.1111/j.1365-313X.2011.04878.x
- Mohnen, D. (2008). Pectin structure and biosynthesis. *Current Opinion in Plant Biology*, 11(3), 266-277. doi:10.1016/j.pbi.2008.03.006
- Moore, I., Galweiler, L., Grosskopf, D., Schell, J., & Palme, K. (1998). A transcription activation system for regulated gene expression in transgenic plants. *Proc Natl Acad Sci U S A*, 95(1), 376-381.
- Muino, J. M., Smaczniak, C., Angenent, G. C., Kaufmann, K., & van Dijk, A. D. (2014). Structural determinants of DNA recognition by plant MADS-domain transcription factors. *Nucleic Acids Res*, 42(4), 2138-2146. doi:10.1093/nar/gkt1172
- Nakaune, S., Yamada, K., Kondo, M., Kato, T., Tabata, S., Nishimura, M., & Hara-Nishimura, I. (2005). A vacuolar processing enzyme, deltaVPE, is involved in seed coat formation at the early stage of seed development. *Plant Cell*, 17(3), 876-887. doi:10.1105/tpc.104.026872
- Nam, J., dePamphilis, C. W., Ma, H., & Nei, M. (2003). Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol Biol Evol*, 20(9), 1435-1447. doi:10.1093/molbev/msg152
- Nassuth, A., Siddiqua, M., Xiao, H. G., Moody, M. A., & Carlow, C. E. (2014). Newly developed quantitative transactivation system shows difference in

- activation by Vitis CBF transcription factors on DRE/CRT elements. *Plant Methods*, 10. doi:Artn 32
10.1186/1746-4811-10-32
- Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I., Caboche, M., & Lepiniec, L. (2002). The TRANSPARENT TESTA16 Locus Encodes the ARABIDOPSIS BSISTER MADS Domain Protein and Is Required for Proper Development and Pigmentation of the Seed Coat. *The Plant Cell Online*, 14(10), 2463-2479. doi:10.1105/tpc.004127
- Ng, M., & Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nat Rev Genet*, 2(3), 186-195. doi:10.1038/35056041
- Nickless, A., Bailis, J. M., & You, Z. (2017). Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell Biosci*, 7, 26. doi:10.1186/s13578-017-0153-7
- North, H. M., Berger, A., Saez-Aguayo, S., & Ralet, M. C. (2014). Understanding polysaccharide production and properties using seed coat mutants: future perspectives for the exploitation of natural variants. *Ann Bot*, 114(6), 1251-1263. doi:10.1093/aob/mcu011
- Ohto, M. A., Floyd, S. K., Fischer, R. L., Goldberg, R. B., & Harada, J. J. (2009). Effects of APETALA2 on embryo, endosperm, and seed coat development determine seed size in Arabidopsis. *Sex Plant Reprod*, 22(4), 277-289. doi:10.1007/s00497-009-0116-1
- Oka, T., Nemoto, T., & Jigami, Y. (2007). Functional analysis of Arabidopsis thaliana RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J Biol Chem*, 282(8), 5389-5403. doi:10.1074/jbc.M610196200
- Okushima, Y., Mitina, I., Quach, H. L., & Theologis, A. (2005). AUXIN RESPONSE FACTOR 2 (ARF2): a pleiotropic developmental regulator. *Plant Journal*, 43(1), 29-46. doi:10.1111/j.1365-313X.2005.02426.x
- Onate-Sanchez, L., & Vicente-Carbajosa, J. (2008). DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC Res Notes*, 1, 93. doi:10.1186/1756-0500-1-93

- Pajoro, A., Madrigal, P., Muino, J. M., Matus, J. T., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J. F., Balazadeh, S., Arif, M., O'Maoileidigh, D. S., Wellmer, F., Krajewski, P., Riechmann, J. L., Angenent, G. C., & Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol*, 15(3), R41. doi:10.1186/gb-2014-15-3-r41
- Parenicova, L., de Folter, S., Kieffer, M., Horner, D. S., Favalli, C., Busscher, J., Cook, H. E., Ingram, R. M., Kater, M. M., Davies, B., Angenent, G. C., & Colombo, L. (2003). Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. *Plant Cell*, 15(7), 1538-1551.
- Penfield, S., Meissner, R. C., Shoue, D. A., Carpita, N. C., & Bevan, M. W. (2001). MYB61 is required for mucilage deposition and extrusion in the Arabidopsis seed coat. *Plant Cell*, 13(12), 2777-2791.
- Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nat Rev Genet*, 14(4), 288-295. doi:10.1038/nrg3458
- Pettolino, F. A., Walsh, C., Fincher, G. B., & Bacic, A. (2012). Determining the polysaccharide composition of plant cell walls. *Nature Protocols*, 7(9), 1590-1607. doi:10.1038/nprot.2012.081
- Pinyopich, A., Ditta, G., Savidge, B., Liljegren, S., Baumann, E., Wisman, E., & Yanofsky, M. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 424(6944), 4.
- Pollock, R., & Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev*, 5(12A), 2327-2341.
- Prouse, M. B., & Campbell, M. M. (2013). Interactions between the R2R3-MYB transcription factor, AtMYB61, and target DNA binding sites. *PLoS One*, 8(5), e65132. doi:10.1371/journal.pone.0065132
- Puranik, S., Acajjaoui, S., Conn, S., Costa, L., Conn, V., Vial, A., Marcellin, R., Melzer, R., Brown, E., Hart, D., Theissen, G., Silva, C. S., Parcy, F., Dumas, R., Nanao, M., & Zubieta, C. (2014). Structural basis for the oligomerization

- of the MADS domain transcription factor SEPALLATA3 in Arabidopsis. *Plant Cell*, 26(9), 3603-3615. doi:10.1105/tpc.114.127910
- Radonic, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun*, 313(4), 856-862.
- Rajani, S., & Sundaresan, V. (2001). The Arabidopsis myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Current Biology*, 11(24), 1914-1922.
- Ralet, M. C., Crepeau, M. J., Vigouroux, J., Tran, J., Berger, A., Salle, C., Granier, F., Botran, L., & North, H. M. (2016). Xylans Provide the Structural Driving Force for Mucilage Adhesion to the Arabidopsis Seed Coat. *Plant Physiol*, 171(1), 165-178. doi:10.1104/pp.16.00211
- Ranocha, P., Francoz, E., Burlat, V., & Dunand, C. (2014). Expression of PRX36, PME16 and SBT1.7 is controlled by complex transcription factor regulatory networks for proper seed coat mucilage extrusion. *Plant Signal Behav*, 9(11), e977734. doi:10.4161/15592324.2014.977734
- Rasimas, J. J., Kar, S. R., Pegg, A. E., & Fried, M. G. (2007). Interactions of human O6-alkylguanine-DNA alkyltransferase (AGT) with short single-stranded DNAs. *J Biol Chem*, 282(5), 3357-3366. doi:10.1074/jbc.M608876200
- Rautengarten, C., Usadel, B., Neurnetzler, L., Hartmann, J., Buessis, D., & Altmann, T. (2008). A subtilisin-like serine protease essential for mucilage release from Arabidopsis seed coats. *Plant Journal*, 54(3), 466-480. doi:DOI 10.1111/j.1365-313X.2008.03437.x
- Reynoso, M. A., Juntawong, P., Lancia, M., Blanco, F. A., Bailey-Serres, J., & Zanetti, M. E. (2015). Translating Ribosome Affinity Purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translomes. *Methods Mol Biol*, 1284, 185-207. doi:10.1007/978-1-4939-2444-8_9
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. doi:10.1093/bioinformatics/btp616

- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*, 11(3), R25. doi:10.1186/gb-2010-11-3-r25
- Romano, J. M., Dubos, C., Prouse, M. B., Wilkins, O., Hong, H., Poole, M., Kang, K. Y., Li, E., Douglas, C. J., Western, T. L., Mansfield, S. D., & Campbell, M. M. (2012). AtMYB61, an R2R3-MYB transcription factor, functions as a pleiotropic regulator via a small gene network. *New Phytologist*, 195(4), 774-786. doi:10.1111/j.1469-8137.2012.04201.x
- Ruprecht, C., Bartetzko, M. P., Senf, D., Dallabernadina, P., Boos, I., Andersen, M. C. F., Kotake, T., Knox, J. P., Hahn, M. G., Clausen, M. H., & Pfrengle, F. (2017). A Synthetic Glycan Microarray Enables Epitope Mapping of Plant Cell Wall Glycan-Directed Antibodies. *Plant Physiol*, 175(3), 1094-1104. doi:10.1104/pp.17.00737
- Sablowski, R. (2007). Flowering and determinacy in Arabidopsis. *J Exp Bot*, 58(5), 899-907. doi:10.1093/jxb/erm002
- Saedler, H., Becker, A., Winter, K. U., Kirchner, C., & Theissen, G. (2001). MADS-box genes are involved in floral development and evolution. *Acta Biochim Pol*, 48(2), 351-358.
- Saez-Aguayo, S., Ralet, M. C., Berger, A., Botran, L., Ropartz, D., Marion-Poll, A., & North, H. M. (2013). PECTIN METHYLESTERASE INHIBITOR6 promotes Arabidopsis mucilage release by limiting methylesterification of homogalacturonan in seed coat epidermal cells. *Plant Cell*, 25(1), 308-323. doi:10.1105/tpc.112.106575
- Saez-Aguayo, S., Rondeau-Mouro, C., Macquet, A., Kronholm, I., Ralet, M. C., Berger, A., Salle, C., Poulain, D., Granier, F., Botran, L., Loudet, O., de Meaux, J., Marion-Poll, A., & North, H. M. (2014). Local evolution of seed flotation in Arabidopsis. *PLoS Genet*, 10(3), e1004221. doi:10.1371/journal.pgen.1004221
- Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., & Scott, R. J. (2006). The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin

- signalling, cell division, and the size of seeds and other organs. *Development*, 133(2), 251-261. doi:10.1242/dev.02194
- Shea, E. M., Gibeaut, D. M., & Carpita, N. C. (1989). Structural analysis of the cell walls regenerated by carrot protoplasts. *Planta*, 179(3), 293-308. doi:10.1007/BF00391074
- Shi, D., Ren, A., Tang, X., Qi, G., Xu, Z., Chai, G., Hu, R., Zhou, G., & Kong, Y. (2018). MYB52 Negatively Regulates Pectin Demethylesterification in Seed Coat Mucilage. *Plant Physiol*, 176(4), 2737-2749. doi:10.1104/pp.17.01771
- Shore, P., & Sharrocks, A. D. (1995). The MADS-box family of transcription factors. *Eur J Biochem*, 229(1), 1-13.
- Shrestha, B., Guragain, B., & Sridhar, V. V. (2014). Involvement of co-repressor LUH and the adapter proteins SLK1 and SLK2 in the regulation of abiotic stress response genes in Arabidopsis. *Bmc Plant Biology*, 14. doi:10.1186/1471-2229-14-54
- Sieburth, L. E., & Meyerowitz, E. M. (1997). Molecular Dissection of the AGAMOUS Control Region Shows That cis Elements for Spatial Regulation Are Located Intragenically. *Plant Cell*, 9, 355-365.
- Sitaraman, J. (2004). Functional analysis of three *Arabidopsis* transcriptional corepressors linking floral regulators to ABA hormone signaling. *Department Of Cell Biology and Molecular Genetics, Doctor of Philosophy*.
- Sitaraman, J., Bui, M., & Liu, Z. (2008). LEUNIG_HOMOLOG and LEUNIG perform partially redundant functions during Arabidopsis embryo and floral development. *Plant Physiol*, 147(2), 672-681. doi:10.1104/pp.108.115923
- Smaczniak, C., Immink, R. G., Angenent, G. C., & Kaufmann, K. (2012). Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development*, 139(17), 3081-3098. doi:10.1242/dev.074674
- Sommer, H., Beltran, J. P., Huijser, P., Pape, H., Lonnig, W. E., Saedler, H., & Schwarz-Sommer, Z. (1990). Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J*, 9(3), 605-613.

- Sprunck, S., & Dresselhaus, T. (2015). Three Cell Fusions during Double Fertilization. *Cell*, 161(4), 708-709. doi:10.1016/j.cell.2015.04.032
- Sridhar, V. V., Surendrarao, A., Gonzalez, D., Conlan, R. S., & Liu, Z. (2004). Transcriptional repression of target genes by LEUNIG and SEUSS, two interacting regulatory proteins for Arabidopsis flower development. *Proc Natl Acad Sci U S A*, 101(31), 11494-11499. doi:10.1073/pnas.0403055101
- Sridhar, V. V., Surendrarao, A., & Liu, Z. (2006). APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. *Development*, 133(16), 3159-3166. doi:10.1242/dev.02498
- Stahle, M. I., Kuehlich, J., Staron, L., von Arnim, A. G., & Golz, J. F. (2009). YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis. *Plant Cell*, 21(10), 3105-3118. doi:10.1105/tpc.109.070458
- Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. *Current Opinion in Plant Biology*, 4(5), 447-456.
- Sullivan, S., Ralet, M. C., Berger, A., Diatloff, E., Bischoff, V., Gonneau, M., Marion-Poll, A., & North, H. M. (2011). CESA5 Is Required for the Synthesis of Cellulose with a Role in Structuring the Adherent Mucilage of Arabidopsis Seeds. *Plant Physiol*, 156(4), 1725-1739. doi:DOI 10.1104/pp.111.179077
- Szabados, L., Kovacs, I., Oberschall, A., Abraham, E., Kerekcs, I., Zsigmond, L., Nagy, R., Alvarado, M., Krasovskaja, I., Gal, M., Berente, A., Redei, G. P., Haim, A. B., & Koncz, C. (2002). Distribution of 1000 sequenced T-DNA tags in the Arabidopsis genome. *Plant Journal*, 32(2), 233-242.
- Takenaka, Y., Kato, K., Ogawa-Ohnishi, M., Tsuruhama, K., Kajiura, H., Yagyu, K., Takeda, A., Takeda, Y., Kunieda, T., Hara-Nishimura, I., Kuroha, T., Nishitani, K., Matsubayashi, Y., & Ishimizu, T. (2018). Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. *Nat Plants*, 4(9), 669-676. doi:10.1038/s41477-018-0217-7
- Tang, W., & Perry, S. E. (2003). Binding site selection for the plant MADS domain protein AGL15: an in vitro and in vivo study. *J Biol Chem*, 278(30), 28154-28159. doi:10.1074/jbc.M212976200

- Thakur, B. R., Singh, R. K., & Handa, A. K. (1997). Chemistry and uses of pectin--a review. *Crit Rev Food Sci Nutr*, 37(1), 47-73. doi:10.1080/10408399709527767
- Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology*, 4(1), 75-85.
- Theissen, G., Melzer, R., & Rumpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development*, 143(18), 3259-3271. doi:10.1242/dev.134080
- Theissen, G., & Saedler, H. (2001). Plant biology. Floral quartets. *Nature*, 409(6819), 469-471. doi:10.1038/35054172
- Tilly, J. J., Allen, D. W., & Jack, T. (1998). The CArG boxes in the promoter of the Arabidopsis floral organ identity gene APETALA3 mediate diverse regulatory effects. *Development*, 125(9), 1647-1657.
- Tiwary, C. M., Ward, J. A., & Jackson, B. A. (1997). Effect of pectin on satiety in healthy US Army adults. *J Am Coll Nutr*, 16(5), 423-428.
- Tsai, A. Y., Kunieda, T., Rogalski, J., Foster, L. J., Ellis, B. E., & Haughn, G. W. (2017). Identification and characterization of Arabidopsis seed coat mucilage proteins. *Plant Physiol*. doi:10.1104/pp.16.01600
- Turner, R., & Foster, G. D. (1995). The potential exploitation of plant viral translational enhancers in biotechnology for increased gene expression. *Mol Biotechnol*, 3(3), 225-236. doi:10.1007/BF02789333
- van Dijk, A. D., Morabito, G., Fiers, M., van Ham, R. C., Angenent, G. C., & Immink, R. G. (2010). Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. *PLoS Comput Biol*, 6(11), e1001017. doi:10.1371/journal.pcbi.1001017
- Verelst, W., Saedler, H., & Munster, T. (2007). MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific Arabidopsis promoters. *Plant Physiol*, 143(1), 447-460. doi:10.1104/pp.106.089805
- Voiniciuc, C., Dean, G. H., Griffiths, J. S., Kirchsteiger, K., Hwang, Y. T., Gillett, A., Dow, G., Western, T. L., Estelle, M., & Haughn, G. W. (2013). Flying saucer1 is a transmembrane RING E3 ubiquitin ligase that regulates the degree of

- pectin methylesterification in Arabidopsis seed mucilage. *Plant Cell*, 25(3), 944-959. doi:10.1105/tpc.112.107888
- Voiniciuc, C., Engle, K. A., Gunl, M., Dieluweit, S., Schmidt, M. H., Yang, J. Y., Moremen, K. W., Mohnen, D., & Usadel, B. (2018). Identification of Key Enzymes for Pectin Synthesis in Seed Mucilage. *Plant Physiol.* doi:10.1104/pp.18.00584
- Voiniciuc, C., Gunl, M., Schmidt, M. H., & Usadel, B. (2015a). Highly Branched Xylan Made by IRREGULAR XYLEM14 and MUCILAGE-RELATED21 Links Mucilage to Arabidopsis Seeds. *Plant Physiol*, 169(4), 2481-2495. doi:10.1104/pp.15.01441
- Voiniciuc, C., Schmidt, M. H., Berger, A., Yang, B., Ebert, B., Scheller, H. V., North, H. M., Usadel, B., & Gunl, M. (2015b). MUCILAGE-RELATED10 Produces Galactoglucomannan That Maintains Pectin and Cellulose Architecture in Arabidopsis Seed Mucilage. *Plant Physiol*, 169(1), 403-420. doi:10.1104/pp.15.00851
- Voiniciuc, C., Yang, B., Schmidt, M. H. W., Gunl, M., & Usadel, B. (2015c). Starting to Gel: How Arabidopsis Seed Coat Epidermal Cells Produce Specialized Secondary Cell Walls. *International Journal of Molecular Sciences*, 16(2), 3452-3473. doi:10.3390/ijms16023452
- Voiniciuc, C., Zimmermann, E., Schmidt, M. H., Gunl, M., Fu, L., North, H. M., & Usadel, B. (2016). Extensive Natural Variation in Arabidopsis Seed Mucilage Structure. *Front Plant Sci*, 7, 803. doi:10.3389/fpls.2016.00803
- Voragen, A. G. J., Coenen, G.-J., Verhoef, R. P., & Schols, H. A. (2009). Pectin, a versatile polysaccharide present in plant cell walls. *Structural Chemistry*, 20(2), 263-275. doi:10.1007/s11224-009-9442-z
- Voss, U., Larrieu, A., & Wells, D. M. (2013). From jellyfish to biosensors: the use of fluorescent proteins in plants. *Int J Dev Biol*, 57(6-8), 525-533. doi:10.1387/ijdb.130208dw
- Walker, M., Tehseen, M., Doblin, M. S., Pettolino, F. A., Wilson, S. M., Bacic, A., & Golz, J. F. (2011). The transcriptional regulator LEUNIG_HOMOLOG

- regulates mucilage release from the Arabidopsis testa. *Plant Physiol*, 156(1), 46-60. doi:10.1104/pp.111.172692
- Wang, X., Niu, Q. W., Teng, C., Li, C., Mu, J., Chua, N. H., & Zuo, J. (2009). Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in Arabidopsis. *Cell Res*, 19(2), 224-235. doi:10.1038/cr.2008.276
- West, A. G., Shore, P., & Sharrocks, A. D. (1997). DNA binding by MADS-box transcription factors: a molecular mechanism for differential DNA bending. *Mol Cell Biol*, 17(5), 2876-2887.
- Western, T. L., Burn, J., Tan, W. L., Skinner, D. J., Martin-McCaffrey, L., Moffatt, B. A., & Haughn, G. W. (2001). Isolation and characterization of mutants defective in seed coat mucilage secretory cell development in Arabidopsis. *Plant Physiol*, 127(3), 998-1011.
- Western, T. L., Skinner, D. J., & Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant Physiol*, 122(2), 345-356.
- Western, T. L., Young, D. S., Dean, G. H., Tan, W. L., Samuels, A. L., & Haughn, G. W. (2004). MUCILAGE-MODIFIED4 encodes a putative pectin biosynthetic enzyme developmentally regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the Arabidopsis seed coat. *Plant Physiol*, 134(1), 296-306. doi:10.1104/pp.103.035519
- Williams, C. R., Baccarella, A., Parrish, J. Z., & Kim, C. C. (2016). Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics*, 17, 103. doi:10.1186/s12859-016-0956-2
- Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provar, N. J. (2007). An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLoS One*, 2(8), e718. doi:10.1371/journal.pone.0000718
- Wolf, S., Mouille, G., & Pelloux, J. (2009). Homogalacturonan methyl-esterification and plant development. *Mol Plant*, 2(5), 851-860. doi:10.1093/mp/ssp066

- Wu, L., El-Mezawy, A., & Shah, S. (2011). A seed coat outer integument-specific promoter for *Brassica napus*. *Plant Cell Reports*, 30(1), 75-80. doi:10.1007/s00299-010-0945-2
- Wuest, S. E., O'Maoileidigh, D. S., Rae, L., Kwasniewska, K., Raganelli, A., Hanczaryk, K., Lohan, A. J., Loftus, B., Graciet, E., & Wellmer, F. (2012). Molecular basis for the specification of floral organs by APETALA3 and PISTILLATA. *PNAS*, 109(33), 6.
- Xia, T., Li, N., Dumenil, J., Li, J., Kamenski, A., Bevan, M. W., Gao, F., & Li, Y. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in *Arabidopsis*. *Plant Cell*, 25(9), 3347-3359. doi:10.1105/tpc.113.115063
- Xing, S., Wallmeroth, N., Berendzen, K. W., & Grefen, C. (2016). Techniques for the Analysis of Protein-Protein Interactions in Vivo. *Plant Physiol*, 171(2), 727-758. doi:10.1104/pp.16.00470
- Yamaguchi, N., Winter, C. M., Wu, M. F., Kwon, C. S., William, D. A., & Wagner, D. (2014). PROTOCOLS: Chromatin Immunoprecipitation from *Arabidopsis* Tissues. *Arabidopsis Book*, 12, e0170. doi:10.1199/tab.0170
- Yang, Y., Fanning, L., & Jack, T. (2003). The K domain mediates heterodimerization of the *Arabidopsis* floral organ identity proteins, APETALA3 and PISTILLATA. *Plant Journal*, 33(1), 47-59.
- Yang, Y., & Jack, T. (2004). Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins. *Plant Molecular Biology*, 55(1), 45-59. doi:10.1007/s11103-004-0416-7
- Yanhui, C., Xiaoyuan, Y., Kun, H., Meihua, L., Jigang, L., Zhaofeng, G., Zhiqiang, L., Yunfei, Z., Xiaoxiao, W., Xiaoming, Q., Yunping, S., Li, Z., Xiaohui, D., Jingchu, L., Xing-Wang, D., Zhangliang, C., Hongya, G., & Li-Jia, Q. (2006). The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Molecular Biology*, 60(1), 107-124. doi:10.1007/s11103-005-2910-y
- Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., & Meyerowitz, E. M. (1990). The protein encoded by the *Arabidopsis* homeotic

- gene agamous resembles transcription factors. *Nature*, 346(6279), 35-39. doi:10.1038/346035a0
- Yu, L., Shi, D. C., Li, J. L., Kong, Y. Z., Yu, Y. C., Chai, G. H., Hu, R. B., Wang, J., Hahn, M. G., & Zhou, G. K. (2014). CELLULOSE SYNTHASE-LIKE A2, a Glucomannan Synthase, Is Involved in Maintaining Adherent Mucilage Structure in Arabidopsis Seed(1[C][W]). *Plant Physiol*, 164(4), 1842-1856. doi:DOI 10.1104/pp.114.236596
- Zhang, X., Henriques, R., Lin, S. S., Niu, Q. W., & Chua, N. H. (2006). Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. *Nature Protocols*, 1(2), 641-646. doi:10.1038/nprot.2006.97
- Zhao, X., Qiao, L., & Wu, A. M. (2017). Effective extraction of Arabidopsis adherent seed mucilage by ultrasonic treatment. *Sci Rep*, 7, 40672. doi:10.1038/srep40672
- Zobell, O., Faigl, W., Saedler, H., & Munster, T. (2010). MIKC* MADS-box proteins: conserved regulators of the gametophytic generation of land plants. *Mol Biol Evol*, 27(5), 1201-1211. doi:10.1093/molbev/msq005

Appendices

Appendix 3

Supplementary Table 1. (Excel)	RIN values of Col-0 & <i>luh-5</i> at 3, 7 and 10 DPA.
Supplementary Table 2. (Excel)	RNA-seq: List of DEGs that are differentially expressed in 7 DPA <i>luh-5</i> mutant whole seeds compared to Col-0.
Supplementary Table 3. (Excel)	RNA-seq: List of DEGs that are differentially expressed in 7 DPA <i>luh-4</i> mutant whole seeds compared to Col-0.
Supplementary Table 4. (Excel)	RNA-seq: List of DEGs that are differentially expressed in 7 DPA <i>mum2-1</i> mutant whole seeds compared to Col-0.
Supplementary Table 5. (Excel)	RNA-seq: List of genes that are differentially expressed at 7 DPA in both <i>luh-4</i> and <i>luh-5</i> , but not in <i>mum2-1</i> mutant seeds.
Supplementary Table 6. (Excel)	RNA-seq: List of genes that are differentially expressed at 3 DPA in <i>luh-5</i> mutant seeds compared to Col-0
Supplementary Table 7. (Excel)	RNA-seq: List of genes that are differentially expressed at 10 DPA in <i>luh-5</i> mutant seeds compared to Col-0.
Supplementary Table 8. (Excel)	RNA-seq: List of genes that are differentially expressed at 3, 7 and 10 DPA in <i>luh-5</i> mutant seeds compared to Col-0.
Supplementary Table 9. (Excel)	qRT-PCR: List of genes analysed in Col-0 and <i>luh-4</i> mutant seed coats at 7 DPA.
Supplementary Table 10. (Excel)	qRT-PCR: List of genes analysed in Col-0 and <i>luh-4</i> mutant embryos at 7 DPA.
Supplementary Table 11.	Oligonucleotides used for qRT-PCR described in Chapter 3.

Supplementary Table 12.	Oligonucleotides used for the generation of promoter-VENUSn constructs described in Chapter 3.
Supplementary Table 13.	List of plasmids used for the generation of promoter-VENUSn constructs described in Chapter 3.
Supplementary Table 14.	Oligonucleotides used for the generation of amiRNAs constructs described in Chapter 3.
Supplementary Table 15.	List of plasmids used for the generation of amiRNAs constructs described in Chapter 3.
Supplementary Table 16.	Dry weight measurements T ₂ and T ₃ seeds derived from lines transformed with <i>GILT_{pro}::amiR-AG</i> .
Supplementary Figure 1.	Activity of MADS-box promoter-reporters in wildtype and <i>luh</i> embryos.
Supplementary Figure 2.	Activity of early differentiation TFs promoter-reporters in wildtype and <i>luh</i> embryos
Supplementary Figure 3.	Ruthenium Red staining of single MADS-box TF mutants.
Supplementary Figure 4.	Activity of putative seed coat specific promoter-reporters in wildtype and <i>luh</i> embryos.

Supplementary Table 11. Oligonucleotides used for qRT-PCR described in Chapter 3.

Gene	Primer	Sequence (5'-3')
<i>TUB7</i>	TUB7-q-F	GCAAAGAGGCTGAGAACTGTGA
	TUB7-qR	CCAGTTCCTCCTCCCAAGG
<i>AP3</i>	AP3-F	TTGTGTGATGCTAGGGTTTCG
	AP3-R	CTCCTTCGTTGTGGTGTTAGG
<i>AG</i>	AG-F2	ATCAGCCAAATTGCGTCAAC
	AG-R2	TCTAATCTGCCTTCCAAGTTCC
<i>SEP1</i>	SEP1-F	CGTCAACAGAGAAATCTTCTTGG
	SEP1-R	GTGTCTTGATGGACCGAACTTG
<i>SEP3</i>	SEP3-F1	GCTCTCAGGACACAGTTTATGCT
	SEP3-R1	ATGCGTTCCTTACTCTGAAGATCG
<i>STK</i>	STK-qF	GAATTAGGATGGGAAGAGGAAAGA
	STK-qR	ATCGCAAAGGACCGAAAGC
<i>SHP1</i>	SHP1-qF	CTATCAGCAAGAAGCCTCTAAGC
	SHP1-qR	GAAGTTCAAGGAACCAAGTGATTC
<i>SHP2</i>	SHP2-qF	ACAACAGTGTGAGAGGAACAA
	SHP2-qR	GAGTATTAGCTTCGGTGATGGTC
<i>MYB3</i>	MYB3-F	CGCCGCTGGATTACAAAGATGTG
	MYB3-R	TTGTTACCGAGCAAGCTATGGAGC
<i>MYB5</i>	MYB5-F	GACGAGGAAGATCTCATCCTCCGT
	MYB5-R	GATCAATGACCACCTGTTGCCGA
<i>TT2</i>	TT2-F	ACTCTCCCTAACCAAGCTGGTCTC
	TT2-R	CGGTCTTAGGTAGTTCTTCCACC
<i>NAC080</i>	NAC080-qF	CCGCTAAGAATGAATGGGTGA
	NAC080-qR	TGGTGAAGAATCAGTTAAAGGTGG
<i>NARS2</i>	NARS2-F	TCTCTCAGGCTTGATGATTGG
	NARS2-R	GATGATGATGTCTAGATGCTGTACT
<i>BAN</i>	BAN-qF	CCGAAGATCCCGAGAAAGAC
	BAN-qR	GATCACACGCTTGACTGATTTTC
<i>AP2</i>	AP2-F	CCGAGTCATCAGGGAATCCTAC
	AP2-R	AGCTCAAATCGAGGTTGTGATC
<i>TTG2</i>	TTG2-F	TGCTTCAGGGGAATCGTCAACAAA
	TTG2-R	TTGGCTCACACTCTTCCCATAGGA
<i>GL2</i>	GL2-F	CATGAGTGGGATGCTTTGTCAA

	GL2-R	ATTGCCACTGAGTTGCCTCTG
<i>TT8</i>	TT8-F	CTCCATCTGGGATGCCAGG
	TT8-R	CCACTTAGCCATACGTGCTTCC
<i>GILT</i>	GILT-qF	TTGCCAGCATGGTGAAGA
	GILT-qR	GTGATTTCTGCGTTCCATAACC
<i>AT1G62000</i>	At1g62000-F	TGGATACAGCCTAGAGAAGAGG
	At1g62000-R	CAGAAGACATTGGAGGGAGTTC
<i>DP1</i>	DP1-F	GCAAGCATTTTCGTCCTCTACC
	DP1-R	CCCCTAGTCCCAAAGGTTTC
<i>AT1G04660</i>	EmbS1-F	GTCATGTCCCGTCGTCTTTC
	EmbS1-R	GCTTTGTCAGTCGTTTGTGTC

Supplementary Table 12. Oligonucleotides used for the generation of promoter-VENUSn constructs described in Chapter 3.

Underlined text corresponds to sequences required for seamless cloning.

Construct	Primer	Sequence (5'-3')
<i>AP3_{pro}::VENUSn</i>	pAP3F-new	<u>CGTCTCGAGGAATTCGGTAC</u> GATACACAAGTTCTTTGGATA
	pAP3R-new	<u>CCTCGCCCTTGCTCACC</u> ATGATTCTTCTCTCTTTGTTTAA
<i>SHP1_{pro}::VENUSn</i>	pSHP1::JG1000-F	<u>CGTCTCGAGGAATTCGGTAC</u> ACAATTTAAGATAAGTTTTGAC
	pSHP1::JG1000-R	<u>CCTCGCCCTTGCTCACC</u> ATGTGATCCTATATATGACAT
<i>SHP2_{pro}::VENUSn</i>	pSHP2::JG1000-F	<u>CGTCTCGAGGAATTCGGTAC</u> GGAATCCAAAGTCTTGGG
	pSHP2::JG1000-R	<u>CCTCGCCCTTGCTCACC</u> ATGTTCTATAAGCCCTAG
<i>STK_{pro}+intron1(I1)::VENUSn</i>	pSTK-I1::JG1000-F	<u>CGCGTCTCGAGGAATTCGGTAC</u> CATGATATATTAACTT
	pSTK-I1::JG1000-R	<u>CCTCGCCCTTGCTCACC</u> ATGCTTCATTTTAAACATCA
<i>NARS2_{pro}::VENUSn</i>	NARS2-JG1000-F2	<u>CGTCTCGAGGAATTCGGTAC</u> AAGGAACAGAGGGAGTAA
	NARS2-JG1000-R2	<u>CCTCGCCCTTGCTCACC</u> ATGTAAAATGATGATCGGAGAATA
<i>TT2_{pro}::VENUSn</i>	TT2-JG1000-F	<u>CGCGTCTCGAGGAATTCGGTAC</u> TATCAATTGTTTCTTG
	TT2-JG1000-R2	<u>CCTTGCTCACC</u> ATGTCTCACTTTTCTCTCTCTTG
<i>MYB5_{pro}::VENUSn</i>	pMYB5-JG1000-F	<u>CGTCTCGAGGAATTCGGTAC</u> CTCTTCTCGTTTCTTCTTCC
	pMYB5-JG1000-R	<u>CTCGCCCTTGCTCACC</u> ATGCTCCGCCGTCTTCAACAA
<i>DP1_{pro}::VENUSn</i>	DP1-F	<u>CGTCTCGAGGAATTCGGTAC</u> TTTGCGGTTTCCTAACCAGG
	DP1-R1	<u>CCTCGCCCTTGCTCACC</u> ATGTGTTAGAGTGTTAAGTAGAAATGAG
<i>GILT_{pro}::VENUSn</i>	pGILT::JG1000-F	<u>CGTCTCGAGGAATTCGGTAC</u> GTTAGCAATGTTATTAAACC
	pGILT::JG1000-R	<u>CCTCGCCCTTGCTCACC</u> ATGTGTTACTAAAAGAGTTTAAG

Supplementary Table 13. List of plasmids used for the generation of promoter-VENUSn constructs described in Chapter 3.

Plasmids used to generate promoter-VENUSn (3X VENUS-N7) constructs are listed in the table below, along with their stock numbers. Plants were transformed with *Agrobacterium tumefaciens* C58 containing binary constructs (pMLBART/pGREENII) and selected by spraying with Basta.

Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
<i>AP3_{pro}::VENUSn</i>	JG1339	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1342	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1350	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>SHP1_{pro}::VENUSn</i>	JG1548	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1550	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1554	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>SHP2_{pro}::VENUSn</i>	JG1549	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1551	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1555	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>STK_{pro_i}::VENUSn</i>	JG1558	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1563	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1566	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>NARS2_{pro}::VENUSn</i>	JG1492	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1496	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1501	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>TT2_{pro}::VENUSn</i>	JG1494	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1498	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1503	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MYB5_{pro}::VENUSn</i>	JG1625	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1626	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1627	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif

Appendices

$VDD_{pro}::VENUSn$	JG1571	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1573	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1577	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
$DP1_{pro}::VENUSn$	JG1336	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1340	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1349	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
$GILT_{pro}::VENUSn$	JG1535	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1538	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1541	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif

Supplementary Table 14. Oligonucleotides used for the generation of amiRNAs constructs described in Chapter 3.

Underlined text corresponds to sequences required for cloning.

Construct	Primer	Sequence (5'-3')
<i>DP1_{pro}</i> in pMIGRO	DP1-F	<u>CGTCTCGAGGAATTCGGTAC</u> TTTGC GGTTTCCTAACCAGG
	DP1-R2	<u>CTCTAGAACTCCATGGAGTGGCGC</u> TTATGGTACCTGTTAGAGTGTTAAGTAG AAATGAG
<i>DP1_{pro}</i> in pGREENII	pGILT/pGREENII-F	<u>CGTCTCGAGGAATTCGGTAC</u> TTTGC GGTTTCCTAACCAGG
	pGILT/pGREENII-R	<u>CATGGAGTGGATCCGGTACC</u> TGTTAGAGTGTTAAGTAGAAATGAG
<i>DP1_{pro}::amiR-LUH</i>	amiRNA-F2	<u>CTTAACACTCTAACAGGTAC</u> CAAACACACGCTCGGAC
	amiRNA-R2	<u>CTAGAACTCCATGGAGTGGCG</u> ATCCATGGCGATGCCT
<i>DP1_{pro}::amiR-MUM2</i>	pDP1::amiR-MUM2-II-F	<u>CTTAACACTCTAACAGGTAC</u> CTTAACACTCTAACAGG
	pDP1::amiR-MUM2-II-R	<u>CTAGAACTCCATGGAGTGGC</u> CTAGAACTCCATGGAGTG
<i>GILT_{pro}</i> in pGREENII	pGILT/pGREENII-F	<u>CGTCTCGAGGAATTCGGTAC</u> GTTAGCAATGTTATTAAACC
	pGILT/pGREENII-R	<u>CATGGAGTGGATCCGGTACC</u> TGTTACTAAAAGAGTTTAAG
<i>GILT_{pro}::amiR-LUH</i>	pGILT::amiR-LUH-I-F	<u>CTTAAACTCTTTTAGTAACAGGTAC</u> CAAACACACGCTCGG
	pGILT::amiR-LUH-I-R	<u>CTCTAGAACTCCATGGAGTG</u> CATGGCGATGCCTTAAATAA
<i>GILT_{pro}::amiR-MUM2</i>	pGILT::amiR-MUM2-II-F	<u>CTTAAACTCTTTTAGTAACAGGTAC</u> CTTAACACTCTAACAGG
	pGILT::amiR-MUM2-II-R	<u>TCTAGAACTCCATGGAGTG</u> CTAGAACTCCATGGAGTG
<i>35S_{pro}:: amiR-AG</i>	35S_amiR164-F	<u>ACGCTCGAGGAATTCGGTAC</u> GAGAGAATGATGAAGGTGTGTG
	migro-amiR164-R	<u>ATTAAAGCAGGACTCTAGAG</u> CGCACGTGTGTGTTTATCG
<i>GILT_{pro}:: amiR-AG</i>	pGILT::amiR-AG-F	<u>CTTAAACTCTTTTAGTAACAGGTAC</u> GAGAGAATGATGAAGG
	pGILT::amiR-AG-R	<u>CTCTAGAACTCCATGGAGTG</u> CGCACGTGTGTGTTTATC

Supplementary Table 15. List of plasmids used for the generation of amiRNAs constructs described in Chapter 3.

Plasmids used to generate artificial miRNA (amiR) constructs are listed in the table below, along with their stock numbers. Plants were transformed with *Agrobacterium tumefaciens* C58 containing binary constructs (pMLBART/pGREENII) and selected by spraying with Basta.

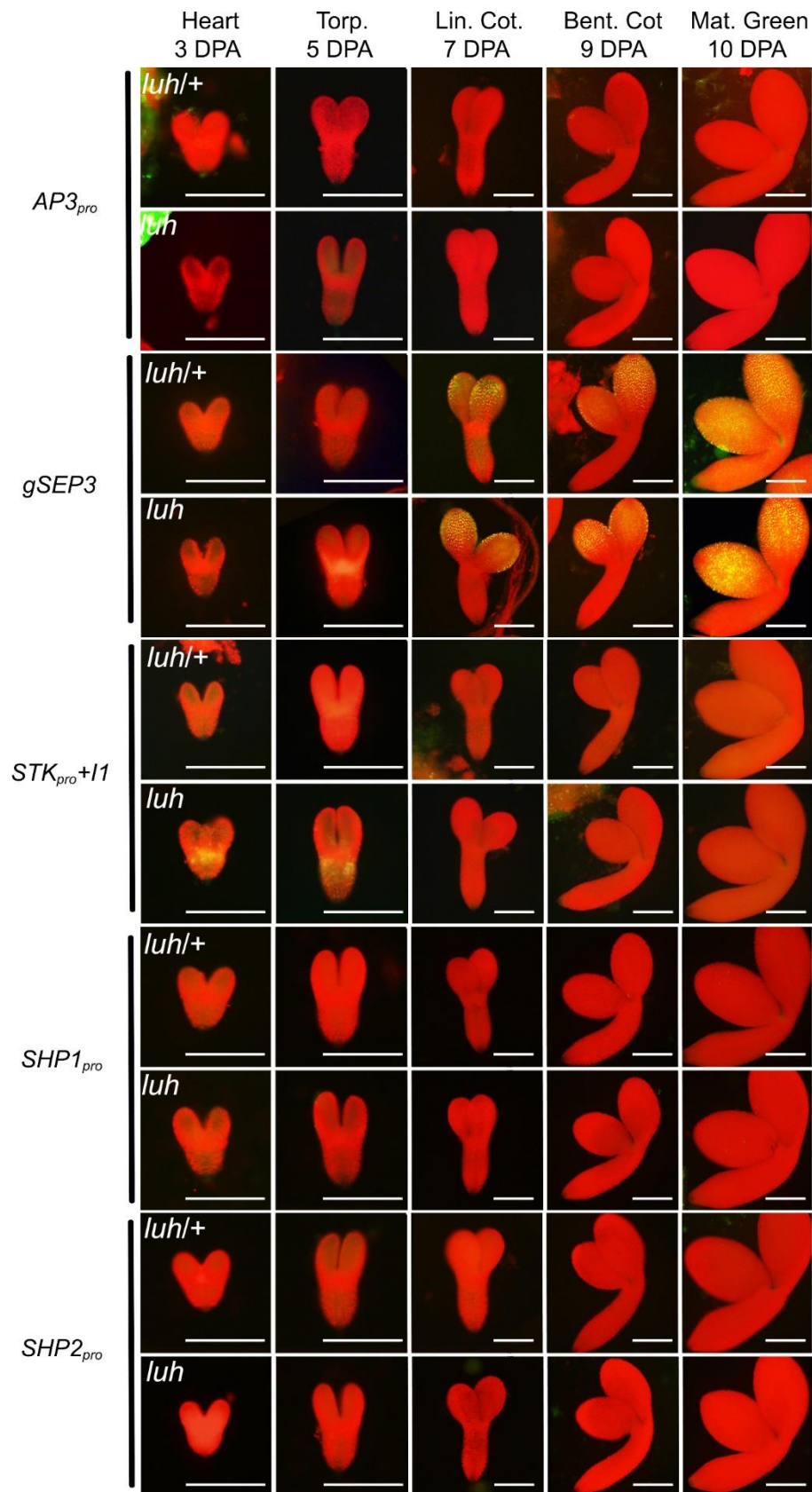
Construct name	Stock number	Vector	Cell line	Antibiotic selection marker
<i>DP1_{pro}</i> in pMIGRO	JG1330	pMIGRO	<i>E.coli</i> Top 10	Amp
<i>DP1_{pro}</i> in pGREENII	JG1455	pGREENII	<i>E.coli</i> Top 10	Kan
<i>DP1_{pro}::amiR-LUH</i>	JG1363	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1368	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1389	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>DP1_{pro}::amiR-MUM2</i>	JG1337	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1341	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1351	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>GILT_{pro}</i> in pGREENII	JG1595	pGREENII	<i>E.coli</i> Top 10	Kan
<i>GILT_{pro}::amiR-LUH</i>	JG1601	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1607	pGREENII	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>GILT_{pro}::amiR-MUM2</i>	JG1603	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1609	pGREENII	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>35S_{pro}:: amiR-AG</i>	JG1409	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1423	pGREENII	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>GILT_{pro}:: amiR-AG</i>	JG1599	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1605	pGREENII	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet

Supplementary Table 16. Dry weight measurements T₂ and T₃ seeds derived from lines transformed with *GILT_{pro}::amiR-AG*.

The dry weight of 300 counted T₂ seeds from 12 T₁ plants transformed with *GILT_{pro}::amiR-AG* was measured in 3*100 lots (Rep 1-3). Lines producing seeds that are at least 1.5x heavier than wildtype were carried through to the next generation (T₂) to determine if the larger seed phenotype was observable in T₃ seeds. Segregation analysis was also performed on these lines to determine transgene copy number.

Rep: replicate; S.D: standard deviation. N/A: not applicable.

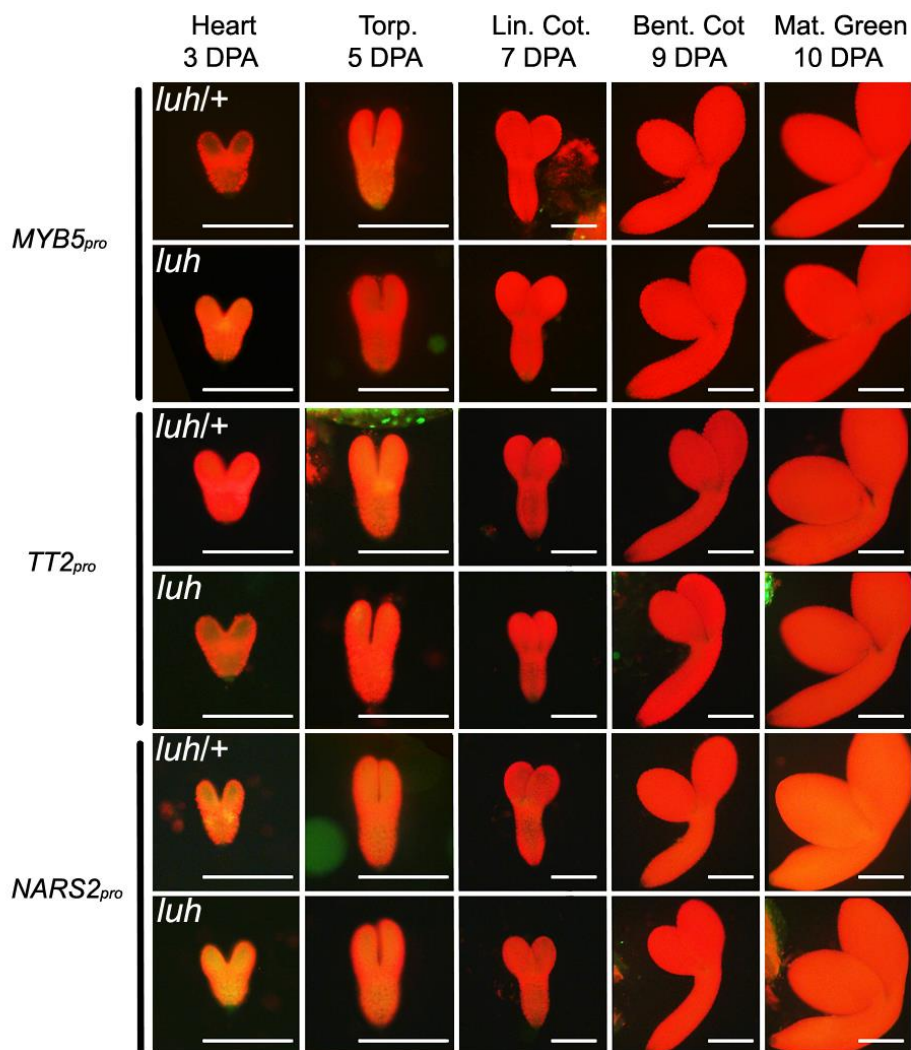
Samples	Rep 1	Rep 2	Rep 3	Average weight (g)	S.D	Fold change (compared to Col-0)	Segregation analysis
T ₁ plants							
Col-0 (wildtype)	0.0021	0.0022	0.0020	0.0021	8.2E-05	N/A	N/A
#5	0.0035	0.0037	0.0034	0.0035	0.00012	1.68	
#8	0.0034	0.0036	0.0035	0.0035	9.4E-05	1.65	
#9	0.0027	0.0030	0.0027	0.0028	0.00014	1.33	
#10	0.0029	0.0030	0.0026	0.0028	0.00017	1.35	
#20	0.0036	0.0038	0.0034	0.0036	0.00016	1.71	
#23	0.0038	0.0040	0.0037	0.0038	0.00012	1.83	
#28	0.0033	0.0034	0.0033	0.0033	4.7E-05	1.59	
#29	0.0025	0.0029	0.0027	0.0028	8.2E-05	1.33	
#31	0.0031	0.0030	0.0032	0.0031	8.2E-05	1.48	
#33	0.0039	0.0041	0.0039	0.0040	9.4E-05	1.89	
#38	0.0027	0.0028	0.0028	0.0028	4.7E-05	1.32	
#40	0.0034	0.0034	0.0035	0.0034	4.7E-05	1.63	
T ₂ plants							
Col-0 (wildtype)	0.0024	0.0021	0.0022	0.0022	0.0001	N/A	N/A
#5	0.0030	0.0029	0.0030	0.0030	5E-05	1.33	1 insert (1:3)
#8	0.0030	0.0032	0.0033	0.0032	0.0001	1.45	1 insert (1:3)
#20	0.0036	0.0036	0.0038	0.0037	9E-05	1.64	1 insert (1:3)
#23	0.0031	0.0034	0.0033	0.0033	0.0001	1.46	1 insert (1:4)
#28	0.0033	0.0031	0.0032	0.0032	8E-05	1.43	1 insert (1:5)
#33	0.0042	0.0040	0.0039	0.0040	0.0001	1.81	1 insert (1:3)



Supplementary Figure 1. Activity of MADS-box promoter-reporters in wildtype and *luh* embryos.

Promoter::VENUSn constructs were introduced into a *luh-5* background. A representative T₁ plants was subsequently crossed to Col-0 and *luh-5* to ensure that the same transgene was introduced into two different backgrounds (*luh-5/+* vs *luh-5/luh-5*) to enable direct comparison. Globular stage embryos are not shown due to difficulties in capturing this stage. Images were captured using the same exposure at all developmental stages.

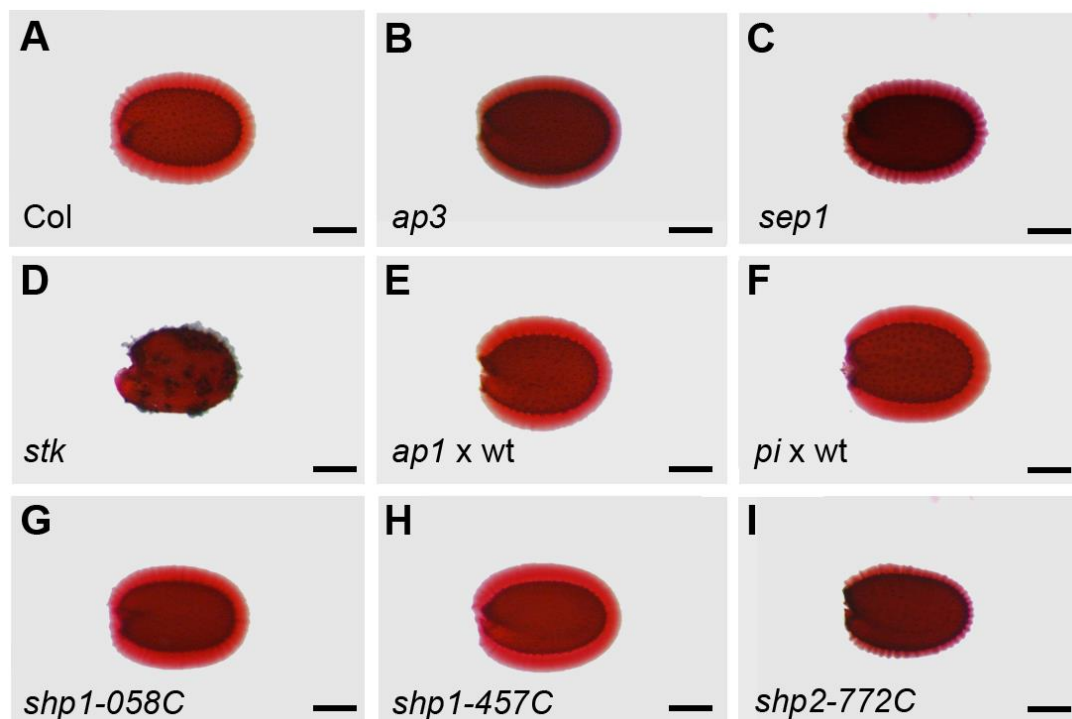
Scale bar = 200 μ m.



Supplementary Figure 2. Activity of early differentiation TFs promoter-reporters in wildtype and *luh* embryos.

Promoter-VENUSn constructs were introduced to a *luh-5* background. A representative T₁ plant was subsequently backcrossed to Col-0 and *luh-5* to ensure that the same transgene was introduced into two different backgrounds (*luh-5/+* vs *luh-5/luh-5*) to enable direct comparison. Globular stage embryos are not shown due to difficulties in capturing this stage. Images were captured using the same exposure at all developmental stages.

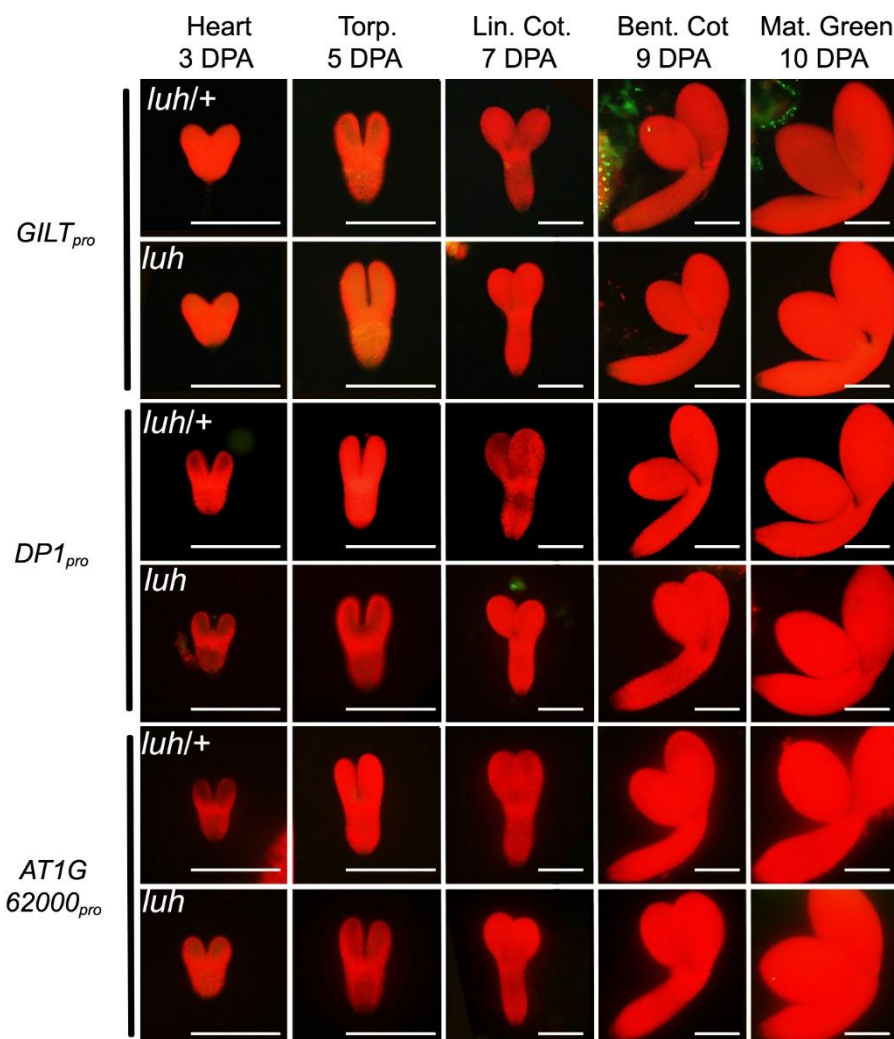
Scale bar = 200 μ m.



Supplementary Figure 3. Ruthenium Red staining of single MADS-box TF mutants.

Col-0 was included as wildtype control. Pictured here are MADS-box TF mutants that have normal mucilage release (A-C, E-H), *stk* (D) mutants which have altered mucilage release, and mutant alleles of *SHP1* (G-H) and *SHP2* (I) not shown in chapter 3. To obtain *ap1* (E) and *pi* (F) seeds, carpels from each mutant line were crossed to wildtype pollen.

Scale bar = 200 μ m.



Supplementary Figure 4. Activity of putative seed coat specific promoter-reporters in wildtype and *luh* embryos.

GILT, *DP1* and *ATIG62000* promoter sequence were fused to *VENUSn* and introduced into *luh-5* mutants. A representative T_1 plant was subsequently backcrossed to wildtype and *luh-5* to generate directly comparable lines with the same transgene in two different genetic backgrounds (*luh-5/+* vs *luh-5/luh-5*) to enable direct comparison. Globular stage embryos are not shown due to difficulties in capturing this stage. Images were captured using the same exposure at all developmental stages.

Scale bar = 200 μ m.

Appendix 4

Supplementary Table 1.	Oligonucleotides used for the generation of constructs described in Chapter 4.
Supplementary Table 2.	List of plasmids used for the generation of constructs described in Chapter 4.
Supplementary Table 3.	Oligonucleotides used for qRT-PCR described in Chapter 4.
Supplementary Table 4. (Excel)	RNA-seq: List of DEGs expressed in 7 DPA <i>shp1 shp2</i> mutant whole seeds compared to Col-0.
Supplementary Table 5. (Excel)	RNA-seq: List of DEGs encoding enzymes expressed in 7 DPA <i>shp1 shp2</i> mutant whole seeds compared to Col-0
Supplementary Table 6. (Excel)	qRT-PCR: List of genes analysed in Col-0, <i>luh-5</i> , <i>shp1 shp2</i> and <i>sep3</i> mutant whole seeds at 7 DPA.

Supplementary Table 1. Oligonucleotides used for the generation of constructs described in Chapter 4.

Underlined text corresponds to sequences required for seamless cloning.

Construct	Primer	Sequence (5'-3')
35S _{pro} ::SHP2	35S:SHP2-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGAGGGTGGTGCAGTAATG
	35S:SHP2-R	ATTAAAGCAGGACTCTAGAGTCAAACAAGTTGCAGAGGTGGTTG
35S _{pro} ::SEP3	35S:SEP3-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGGAAGAGGGAGAGTAGA
	35S:SEP3-R	<u>ATTAAAGCAGGACTCTAGAGTCAA</u> ATAGAGTTGGTGTCTATAAGG

Supplementary Table 2. List of plasmids used for the generation of constructs described in Chapter 4.

Plasmids used to generate constructs are listed in the table below, along with their stock numbers. Plants were transformed with *A. tumefaciens* C58 containing binary constructs (pMLBART/pGREENII) and selected by spraying with Basta. Asterisk (*) indicates pre-existing stocks made by previous members of the Golz lab.

Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
<i>35S_{pro}::SHP2</i>	JG1570	pGREEN II	<i>E.coli</i> Top 10	Kan
	JG1576	pGREEN II	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>35S_{pro}::SEP3</i>	JG1461	pGREEN II	<i>E.coli</i> Top 10	Kan
	JG1481	pGREEN II	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>SHP2_{pro}::SHP2-VENUS*</i>	JG1651	pGREEN II	<i>E.coli</i> Top 10	Kan
	JG1679	pGREEN II	<i>Agrobacterium</i> C58	Kan/Gent/Rif/Tet
<i>MUM2i_{pro}::VENUSn*</i>	JG1234	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1241	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>35S_{pro}::MUM2*</i>	JG726	pMLBART	<i>E.coli</i> Top 10	Spec
	JG751	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>35S_{pro}::LUH*</i>	JG392	pMLBART	<i>E.coli</i> Top 10	Spec
	JG393	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif

Supplementary Table 3. Oligonucleotides used for qRT-PCR described in Chapter 4.

Gene	Primer	Sequence (5'-3')
<i>ACT2</i>	Actin#1 RT	TCTTCCGCTCTTTCTTTCCA
	Actin#2 RT	TCCTTCTGGTTCATCCCAAC
<i>MUM2</i>	BGAL6- qF	GCATTGGTCGTTATTGGGTCTCG
	BGAL6- qR	CCACCTTCCTCTTCAAATACCACC
<i>LUH</i>	MUM1-qF	GCTGACAAGATGCTTGATGTTT
	MUM1-qR	GATCAGGAGAAACCTTCCCTTC

Appendix 5

Supplementary Table 1.	List of plasmids used for the generation of constructs for <i>Arabidopsis</i> transformation described in Chapter 5.
Supplementary Table 2.	Oligonucleotides used for the generation of constructs for <i>Arabidopsis</i> transformation described in Chapter 5.
Supplementary Table 3.	Oligonucleotides used for qRT-PCR described in Chapter 5.
Supplementary Table 4.	Oligonucleotides used for the generation of constructs for transactivation assays described in Chapter 5.
Supplementary Table 5.	List of plasmids used for the generation of constructs for transactivation assays described in Chapter 5.
Supplementary Table 6.	Oligonucleotides used for the generation of constructs for Y2H assays described in Chapter 5.
Supplementary Table 7.	List of plasmids used for the generation of constructs for Y2H assays described in Chapter 5.
Supplementary Table 8.	Oligonucleotides used for the generation of constructs for bimolecular fluorescence complementation (BiFC) assays described in Chapter 5
Supplementary Table 9.	List of plasmids used for the generation of constructs for BiFC assays described in Chapter 5.
Supplementary Figure 1.	Activity of CArG box/CL deletions in <i>MUM2_{pro}+I1</i> reporters in Col-0 and <i>luh</i> embryos.
Supplementary Figure 2.	Activity of CArG box/CL mutations in <i>MUM2_{pro}+I1</i> reporters in Col-0 and <i>luh</i> embryos.

Supplementary Table 1. List of plasmids used for the generation of constructs for *Arabidopsis* transformation described in Chapter 5.

Plasmids used to generate constructs are listed in the table below, along with their stock numbers. Plants were transformed with *A. tumefaciens* C58 containing binary constructs (pMLBART/pGREENII) and selected by spraying with Basta. Asterisks (*) indicates stocks made by previous or other Golz lab members.

Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
<i>MUM2_{pro}::E1(modified)+I1 (minus CArG-III/MYB-II)::VENUSn *</i>	JG1462	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1467	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1471	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (minus CArG-III/IV/V/MYB-II)::VENUSn *</i>	JG1463	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1468	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1472	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (minus CArG-III/IV/MYB-II)::VENUSn *</i>	JG1464	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1469	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1473	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (minus MYB-I/MYB-II)::VENUSn *</i>	JG1465	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1470	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1474	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (CArG-I/II/III mutation)::VENUSn *</i>	JG1559	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1567	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1579	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (CArG-IV mutation)::VENUSn *</i>	JG1560	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1568	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1580	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
<i>MUM2_{pro}::E1(modified)+I1 (CArG-I/II/III/IV mutation)::VENUSn *</i>	JG1743	pMIGRO	<i>E.coli</i> Top 10	Amp
	JG1750	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1758	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif

35S _{pro} ::SHP2	JG1631	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1634	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
35S _{pro} ::SHP2-SRDX	JG1632	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1635	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
LUH _{pro} ::VENUSn *	JG1685	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
LUG _{pro} ::VENUSn *	JG1645	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
SEU _{pro} ::VENUSn *	JG1371	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
SLK1 _{pro} ::VENUSn *	JG1357	pMLBART	<i>Agrobacterium</i> C58	Spec/Gent/Rif
SLK2 _{pro} ::YFP::SLK2 *	JG450	pMLBART	<i>Agrobacterium</i> GV310	Spec/Gent/Rif

Supplementary Table 2. Oligonucleotides used for the generation of constructs for *Arabidopsis* transformation described in Chapter 5.

Construct	Primer	Sequence (5'-3')
35S _{pro} ::SHP2	Topo-SHP2-F	ATGGAGGAAGGTGGGAGTAGTC
	SHP2-STOP-R	AACAAGTTGCAGAGGTGGTTG
35S _{pro} ::SHP2-SRDX	Topo-SHP2-F	ATGGAGGAAGGTGGGAGTAGTC
	SRDX::SHP2-R	AGCAGCTCCACCTCCACCTCCAACAAGTTGCAGAGGTGGTTG

Supplementary Table 3. Oligonucleotides used for qRT-PCR described in Chapter 5.

Gene	Primer	Sequence (5'-3')
<i>ACT2</i>	Actin#1 RT	TCTTCCGCTCTTTCTTTCCA
	Actin#2 RT	TCCTTCTGGTTCATCCCAAC
<i>MUM2</i>	BGAL6- qF	GCATTGGTCGTTATTGGGTCTCG
	BGAL6- qR	CCACCTTCCTCTTCAAATACCACC

Supplementary Table 4. Oligonucleotides used for the generation of constructs for transactivation assays described in Chapter 5.

Underlined text corresponds to sequences required for seamless cloning.

Construct	Primer	Sequence (5'-3')
<i>35S_{pro}::SHP1</i>	35S:SHP1-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGAGGAAGGTGGGAGTAGTC
	35S:SHP1-R	<u>ATTAAAGCAGGACTCTAGAGT</u> TACACAAGTTGAAGAGGAGGT
<i>35S_{pro}::SHP2</i>	35S:SHP2-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGAGGGTGGTGCGAGTAATG
	35S:SHP2-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAACAAGTTGCAGAGGTGGTTG
<i>35S_{pro}::SEP3</i>	35S_SEP3-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGGAAGAGGGAGAGTAGA
	35S_SEP3-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAATAGAGTTGGTGTCTATAAGG

Supplementary Table 5. List of plasmids used for the generation of constructs for transactivation assays described in Chapter 5.

Plasmids used to generate constructs are listed in the table below, along with their stock numbers. *N. benthamiana* leaves were infiltrated with *A.tumefaciens* AGL1 containing binary constructs (pMLBART/pGREENII). Asterisk (*) indicates stocks made by previous or other Golz lab members.

Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
MYB75 _{pro} (2.2kb)::LUC *	JG1727	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1737	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
MYB75 _{pro} (0.8kb)::LUC *	JG1667	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1677	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
MUM2 _{pro} ::E1(modified)+I1::LUC *	JG1723	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1735	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
MUM2 _{pro} ::E1(modified)+I1 (CArG-I/II/III/IV mutation)::LUC *	JG1746	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1756	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::empty *	JG367	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1674	pMLBART	<i>Agrobacterium</i> AGL1	Spec/Amp
35S _{pro} ::FIL *	JG99	pMLBART	<i>E.coli</i> Top 10	Spec
	JG102	pMLBART	<i>Agrobacterium</i> AGL1	Spec/Amp
35S _{pro} ::SHP1	JG1569	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1617	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SHP2	JG1570	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1618	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEP3	JG1461	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1738	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp

Supplementary Table 6. Oligonucleotides used for the generation of constructs for Y2H assays described in Chapter 5.

Construct	Primer	Sequence (5'-3')
<i>SHP1 in pGAD-T7</i>	Topo-SHP1-F2	CACCATGGAGGAAGGTGGGAGTAGTC
	SHP1-STOP-R	ACACAAGTTGAAGAGGAGGT
<i>SHP2 in pGAD-T7</i>	Topo-SHP2-F2	CACCATGGAGGGTGGTGCGAGTAATG
	SHP2-STOP-R	AACAAGTTGCAGAGGTGGTTG
<i>STK in pGAD-T7</i>	Topo-STK-F	CACCATGCTCTTTCCCCATGAAAG
	STK-R	TTATCCGAGATGAAGAATTTCTTGTCG

Supplementary Table 7. List of plasmids used for the generation of constructs for Y2H assays described in Chapter 5.

Plasmids used to generate constructs are listed in the table below, along with their stock numbers. *S. cerevisiae* Y2HGold was co-transformed with bait and prey constructs (pGBK-T7/pGAD-T7) grown in *E. coli* Top 10 and selected for by SD/-Leu/-Trp media.

Asterisk (*) indicates stocks made by previous or other Golz lab members.

Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
<i>LUH in pGBK-T7</i> *	JG44	pGBK-T7	<i>E. coli</i> Top 10	Kan
<i>SEU in pGBK-T7</i> *	JG378	pGBK-T7	<i>E. coli</i> Top 10	Kan
<i>SLK1 in pGBK-T7</i> *	JG473	pGBK-T7	<i>E. coli</i> Top 10	Kan
<i>SLK2 in pGBK-T7</i> *	JG472	pGBK-T7	<i>E. coli</i> Top 10	Kan
<i>pGBK-T7</i> (empty) *	JG136	pGBK-T7	<i>E. coli</i> Top 10	Kan
<i>AP1 in pGAD-T7</i> *	JG1642	pGAD-T7	<i>E. coli</i> Top 10	Amp
<i>SEP3 in pGAD-T7</i> *	JG1641	pGAD-T7	<i>E. coli</i> Top 10	Amp
<i>SHP1 in pGAD-T7</i>	JG1636	pGAD-T7	<i>E. coli</i> Top 10	Amp
<i>SHP2 in pGAD-T7</i>	JG1637	pGAD-T7	<i>E. coli</i> Top 10	Amp
<i>STK in pGAD-T7</i>	JG1638	pGAD-T7	<i>E. coli</i> Top 10	Amp
<i>pGAD-T7</i> (empty) *	JG137	pGAD-T7	<i>E. coli</i> Top 10	Amp

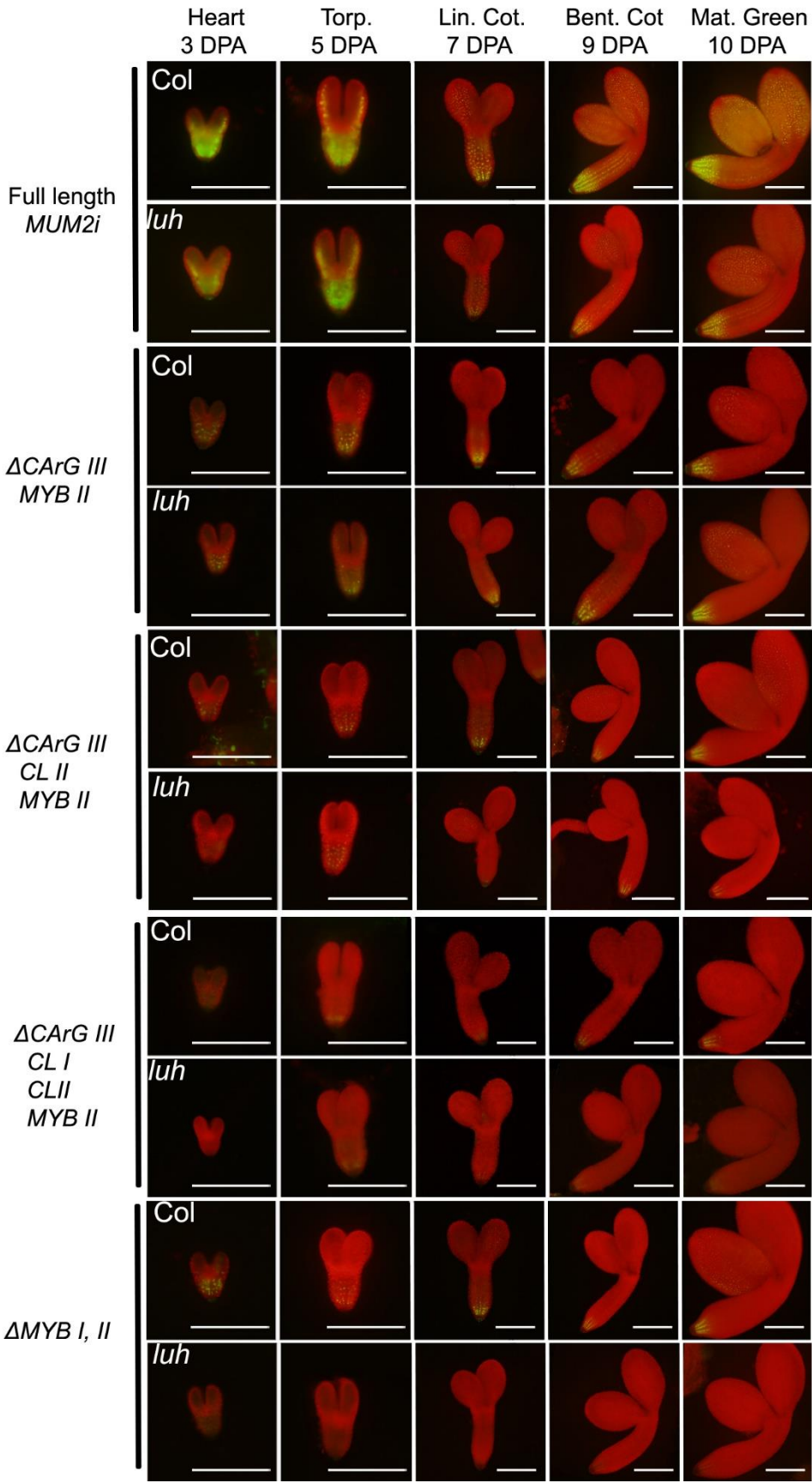
Supplementary Table 8. Oligonucleotides used for the generation of constructs for BiFC assays described in Chapter 5.Underlined text corresponds to sequences required for seamless cloning.

Construct	Primer	Sequence (5'-3')
<i>35S_{pro}::SHP1::VENUS_N</i>	Venus-SHP1-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGAGGAAGGTGGGAGTAGTC
	35S-SHP1-R	<u>ATTAAAGCAGGACTCTAGAGT</u> TACACAAGTTGAAGAGGAGGT
<i>35S_{pro}::SHP1::VENUS_C</i>	VENUSn-SHP1-F	<u>GTGGATCTGGAGGAGGAGGC</u> ATGGAGGAAGGTGGGAGTAGTC
	35S-SHP1-R	<u>ATTAAAGCAGGACTCTAGAGT</u> TACACAAGTTGAAGAGGAGGT
<i>35S_{pro}::SHP2::VENUS_N</i>	Venus-SHP2-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGAGGGTGGTGCGAGTAATG
	35S-SHP2-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAACAAGTTGCAGAGGTGGTTG
<i>35S_{pro}::SHP2::VENUS_C</i>	VENUSn-SHP2-F	<u>GTGGATCTGGAGGAGGAGGC</u> ATGGAGGGTGGTGCGAGTAATG
	35S-SHP2-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAACAAGTTGCAGAGGTGGTTG
<i>35S_{pro}::SEP3::VENUS_N</i>	Venus-SEP3-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGGAAGAGGGAGAGTAGA
	35S-SEP3-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAATAGAGTTGGTGTCATAAGG
<i>35S_{pro}::SEP3::VENUS_C</i>	VENUSn-SEP3-F	<u>GTGGATCTGGAGGAGGAGGC</u> ATGGGAAGAGGGAGAGTAGA
	35S-SEP3-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAATAGAGTTGGTGTCATAAGG
<i>35S_{pro}::SEP3</i>	35S-SEP3-F	<u>ACGCTCGAGGAATTCGGTAC</u> ATGGGAAGAGGGAGAGTAGA
	35S-SEP3-R	<u>ATTAAAGCAGGACTCTAGAGT</u> CAAATAGAGTTGGTGTCATAAGG

Supplementary Table 9. List of plasmids used for the generation of constructs for BiFC assays described in Chapter 5.

Plasmids used to generate constructs are listed in the table below, along with their stock numbers. *N. benthamiana* leaves were infiltrated with *A. tumefaciens* AGL1 containing binary constructs (pMLBART/pGREENII). Asterisk (*) indicates stocks made by previous Golz lab members.

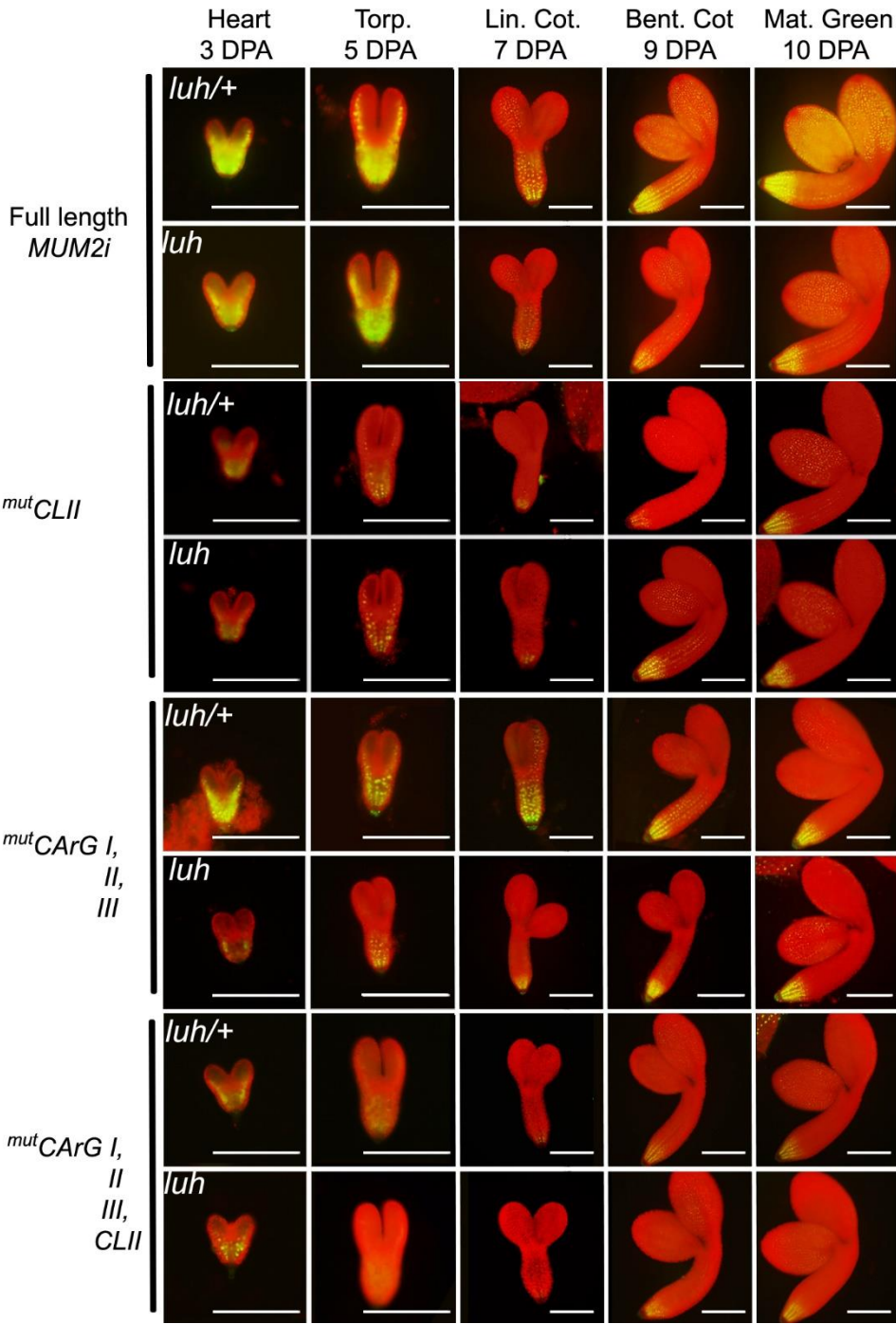
Construct name	Stock numbers	Vector	Cell line	Antibiotic selection marker
35S _{pro} ::LUH::VENUS _N *	JG952	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} :: LUH::VENUS _C *	JG951	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEU::VENUS _N *	JG949	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEU::VENUS _C *	JG950	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SHP1::VENUS _N	JG1716	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1732	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SHP1::VENUS _C	JG1747	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1753	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SHP2::VENUS _N	JG1718	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1734	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SHP2::VENUS _C	JG1748	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1754	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEP3::VENUS _N	JG1717	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1733	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEP3::VENUS _C	JG1749	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1755	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp
35S _{pro} ::SEU *	JG1001	pMLBART	<i>E.coli</i> Top 10	Spec
	JG1614	pMLBART	<i>Agrobacterium</i> AGL1	Spec/Amp
35S _{pro} ::SEP3	JG1461	pGREENII	<i>E.coli</i> Top 10	Kan
	JG1738	pGREENII	<i>Agrobacterium</i> AGL1	Kan/Tet/Amp



Supplementary Figure 1. Activity of CArG box/CL deletions in *MUM2_{pro}+I1* reporters in Col-0 and *luh* embryos.

Promoter::VENUSn constructs were introduced into both Col-0 and *luh-5* background. Sixteen lines were imaged per background, and a representative plant per background was chosen for imaging across early-late stages. Globular stage embryos are not shown due to difficulties in capturing this stage. Images were captured using the same exposure at all developmental stages.

Scale bar = 200 μ m.



Supplementary Figure 2. Activity of CArG box/CL mutations in *MUM2_{pro}+I1* reporters in Col-0 and *luh* embryos.

Promoter::VENUSn constructs were introduced into a *luh-5* background. A representative T₁ plant was subsequently crossed to Col-0 and *luh-5* to ensure that the same transgene was introduced into the two different backgrounds (*luh-5/+* vs *luh-*

5/luh-5) to enable direct comparison. Globular stage embryos are not shown due to difficulties in capturing this stage. Images were captured using the same exposure (ISO 200) at all developmental stages with the exception of *MUM2i^{mutCArG I, II, III, CLII}*, which was taken at 4x exposure (ISO 800). All images were adjusted for brightness in Adobe Photoshop (ver. 12.0) to reflect equal exposure.

Scale bar = 200 μ m.