

AN ABSTRACT OF THE DISSERTATION OF

Estefania Elorriaga for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on March 17, 2020.

Title: Functional Characterization and Classification of Genes Essential to Flower Induction, Flower Development, and Seed Development in *Populus* and *Eucalyptus*

Abstract approved: _____

Steven H. Strauss

This dissertation consists of four studies of *Populus* and *Eucalyptus* biotechnology and genomic science: 1) induction of floral sterility by tapetal expression of the ribonuclease Barnase in *Populus*; 2) CRISPR Cas9-mediated gene editing targeting *LEAFY* (*LFY*) and *AGAMOUS* (*AG*) homologs in *Populus*; 3) induction of floral sterility by CRISPR Cas9 mutagenesis of the *LEAFY* (*LFY*) gene homolog in hybrid *Eucalyptus*; and 4) Gene expression analysis using RNASeq during flowering and seed capsule development in *Eucalyptus grandis* (*E. grandis*).

The *Populus* genus includes species and hybrids that are favored in commercial and research settings because of their fast growth and their relative ease of transformation and regeneration. The *Eucalyptus* genus is the most planted genus of hardwood trees in the world. *Eucalyptus* species and hybrids are grown for wood, pulp, essential oils, and honey. Regardless of their value, the genes and genomes of trees are not as well-studied as those of other plant and crop species, including *Arabidopsis*, *Antirrhinum*, maize, and tobacco. We conducted four studies to address research needs associated with tree floral biology and genetic containment.

In study 1, we performed a four-year field trial of transgenic male poplars that expressed the ribonuclease Barnase in the tapetal layer of their anthers. The purpose was to test the efficiency of this RNase at disrupting pollen development and determine whether growth was affected. During the trial, 17 of 18 transgenic barnase-expressing trees grew on average 40% slower than the WT control trees (i.e., no transformation). The 18 Barnase-expressing trees did not have detectable production of viable pollen

In study 2, we evaluated the mutation efficiency and mutation spectra induced by four CRISPR Cas9 nucleases targeting two different sites in three essential flowering genes, *LFY* and the two *AG* homologs in *Populus*. The average mutation rate was 77.5%; a higher mutation rate than observed before the advent of CRISPR. No undesirable mutations were seen in 310 potential “off-target” loci, and no mutations were seen at any of the target sites in the empty-vector (“Cas9-only” plus markers) control population.

In study 3, we analyzed the mutation efficiency and efficacy of CRISPR Cas9 nucleases at inducing knockout mutations in the *LFY* ortholog in *Eucalyptus*. In 68 transgenic lines, the average mutation rate was 98.5%. After evaluating the floral morphology of 32 lines in the greenhouse, we calculated the average loss-of-function (LOF) rate to be 91%. Lines with LOF mutations failed to produce flowers and viable gametes. Meanwhile, the LOF mutations did not affect growth. The expression of genes upstream and downstream of *LFY* in the floral development pathway suggest that the mutant flowers were not transitioning adequately between inflorescences and flowers. However, occasional sterile and underdeveloped floral organs were seen. Further long-term research is necessary to determine whether mutation of *LFY* is a fully reliable containment technology.

In study 4, we examined the gene expression of 20 samples corresponding to seven different floral and vegetative tissues during late flowering and early seed capsule development in *E. grandis*. Expression libraries were created for flowers and seed capsules at five time points between anthesis and early seed development. Libraries were also constructed for mature pollen and mature leaf (i.e., the vegetative control). We identified differentially expressed genes by comparing the expression of all reproductive tissues to mature flower and also by comparing expression among reproductive tissues.

In total we identified 27,450 unique transcripts, and identified 11,438 differentially expressed transcripts (false discovery rate of 0.05, filtered to genes with at least double the expression if positive change, or half the expression if negative change). We found that genes involved in the biosynthesis of phenylpropanoids were important and differentially expressed during anther and seed capsule maturation. Flower development genes were expressed in tissues homologous to those of other flowers previously characterized, e.g. *Arabidopsis* and *Antirrhinum*. The transcriptome data provides a rich resource to support studies of floral evolution in the Myrtales, and will inform efforts to breed or genetically engineer sexual development in *Eucalyptus*.

©Copyright by Estefania Elorriaga

March 17, 2020

All Rights Reserved

Functional Characterization and Classification of Genes Essential to Flower Induction,
Flower Development, and Seed Development in *Populus* and *Eucalyptus*

by
Estefania Elorriaga

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 17, 2020
Commencement June 2020

Doctor of Philosophy dissertation of Estefania Elorriaga presented on March 17, 2020

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Graduate Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Estefania Elorriaga, Author

ACKNOWLEDGEMENTS

There are so many people I would like to thank. First, I need to thank my advisor for being so passionate. His passion is contagious. I also need to thank him for all the time and dedication he has spent molding me into the scientist and researcher I have become. I am also grateful to the members of my committee, John Fowler, James Myers, Brett Tyler, Glenn Howe, and Amy Klocko. Thank you for patiently guiding me, showing me where I needed to improve, and for giving me encouragement. In particular, I thank Amy for all the time and effort and guidance she provided. She was my main mentor in the lab. I would also like to thank Cathleen Ma for training me on tissue culture, greenhouse care, and transformation. I would like to thank all of my friends for their encouragement and support. In particular, I thank Haiwei Lu and Anna Magnuson. Haiwei's calm temperament and relentless work ethic propelled me forward during the most challenging and exasperating times of my doctorate. I could have not wished for a better labmate/officemate. Anna's support and constant open-door policy allowed me to vent my frustrations and regain my optimism and confidence. I thank my parents for providing me with opportunities to learn and grow and for always rooting for me and loving me. Last, but not least, I thank my husband Gregory for all his love, support, and understanding.

I would also like to thank the members of the Tree Biotechnology and Genetics Research Cooperative including FuturaGene/Suzano, SAPPI, SweTree, Klabin, Arauco, Arborgen and University of Pretoria for their financial support to our laboratory over the years. I would also like to thank FuturaGene for providing us with their SP7 hybrid *Eucalyptus* clone so essential to my work, and SAPPI for collecting and delivering the reproductive tissues from their *Eucalyptus grandis* field trees for my transcriptome studies.

CONTRIBUTIONS OF AUTHORS

Drs. Xinmin An, Amy M. Brunner, Ms. Anna C. Magnuson, Mrs. Cathleen Ma, Richard Meilan, Alexander A. Myburg, Mr. Marc du Plessis, Jeffrey S. Skinner, and Mr. Luke Solomon coauthored one or more of the reported studies. Dr. Amy L. Klocko assisted in the experimental design for this work and co-led several of the grants that funded this work along with Dr. Steve Strauss, who oversaw all of the research performed in the studies.

TABLE OF CONTENTS

	<u>Page</u>
1 Introduction.....	1
1.1 The origin of genetic engineering.....	1
1.2 The generation of transgenic plants.....	2
1.3 Plant tissue culture.....	4
1.4 Regulation of genetically modified plants.....	7
1.5 Commercial application of genetically modified plants.....	9
1.6 Risks of genetically modified plants	11
1.7 Society’s perception of genetically modified plants.....	11
1.8 Genetic use restriction technologies	13
1.9 Genetic containment techniques	13
1.10 CRISPR technology and applications.....	14
1.11 Floral molecular biology	17
2 A tapetal ablation transgene induces stable male-sterility and slows field growth in <i>Populus</i>	22
2.1 Introduction	25
2.2 Materials and Methods	27
2.3 Results	34
2.4 Discussion.....	39
3 Variation in mutation spectra among CRISPR Cas9 mutagenized poplars.....	45
3.1 Introduction	48
3.2 Materials and methods.....	50

TABLE OF CONTENTS (continued)

	<u>Page</u>
3.3 Results	58
3.4 Discussion.....	70
4 CRISPR disruption of <i>LEAFY</i> function in <i>Eucalyptus</i> gives sterile indeterminate inflorescences but normal vegetative development	74
4.1 Introduction	77
4.2 Materials and Methods	80
4.3 Results	87
4.4 Discussion.....	96
5 The fruit and seed pod transcriptome of <i>Eucalyptus grandis</i>	103
5.1 Introduction	104
5.2 Materials and Methods	106
5.3 Results	112
5.4 Discussion.....	130
6 Conclusion	140
Bibliography	143
Appendices.....	179
Appendix A Supplementary material for Chapter 2.....	180
Appendix B Supplementary material for Chapter 3.....	200
Appendix C Supplementary material for Chapter 4.....	216
Appendix D Supplementary material for Chapter 5.....	234

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Fig. 2.1 Field trial during early growth and catkin collection.....	32
Fig. 2.2 Transgenic trees showed reduced growth when compared to non-transgenic control	35
Fig. 2.3 Reporter and non-transgenic trees grew at similar rate.....	36
Fig. 2.4 Absence of visible pollen release from transgenic catkins.....	37
Fig. 2.5 Transgenic catkins lacked visible pollen.....	38
Fig. 2.6 Transgenic catkins lacked visible pollen.....	39
Fig. 3.1 CRISPR Cas9 sgRNA design and mutation detection in <i>LFY</i> and <i>AG</i> paralogs	51
Fig. 3.2 Experimental constructs targeting one or two loci simultaneously.....	54
Fig. 3.3 Transformation event genotyping of <i>LFY</i> and <i>AG</i> paralogs.....	68
Fig. 4.1 Examples of nucleotide sequence alignments of gene-edited <i>ELFY</i> alleles.....	81
Fig. 4.2 Stem volume growth and plant form appear to be unaffected by knock-out of <i>ELFY</i>	89
Fig. 4.3 Wildtype <i>E. grandis</i> flower images.....	91
Fig. 4.4 Flower development stages in control and <i>ELFY</i> knock-outs.....	91
Fig. 4.5 Sterile floral-like buds with underdeveloped ovules belonging to mutant event 30-6	93
Fig. 4.6 Sterile floral-like buds from mutant event 30-10 with many repeated bract-like and pedicel-like organs, and no underdeveloped ovules or anthers.....	94
Fig. 4.7 Transcriptional network related to <i>ELFY</i> , and its expression from qPCR, in floral or floral-like buds.....	95

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
Fig. 4.8 Gene expression of floral development genes at the same level or upstream of <i>ELFY</i> in the flowering induction pathway for <i>ELFY</i> -FM and non-mutant flowering control events.	96
Fig. 4.9 Gene expression of organ identity genes downstream of <i>ELFY</i>	97
Fig. 5.1 Timeline of tissue sampling in the South African plantation.	107
Fig. 5.2 Images of tissues sequenced.....	108
Fig. 5.3 Summary of methods for differential expression and gene ontology (GO) analysis.....	110
Fig. 5.4 Genes identified during differential expression analysis with high expression in pollen compared to leaf.....	116
Fig. 5.5 Genes identified during differential expression analysis with high expression in flowers or early capsule compared to leaf	117
Fig. 5.6 Genes identified during differential expression analysis with high expression in late capsule compared to leaf	118
Fig. 5.7 Expression levels in cluster 1 of the capsule development cluster analysis	120
Fig. 5.8 Expression levels in cluster 2 of the capsule development cluster analysis	121
Fig. 5.9 Expression levels in cluster 3 of the capsule development cluster analysis	122
Fig. 5.10 Expression levels in cluster 4 of the capsule development cluster analysis ...	123
Fig. 5.11 Expression levels in the three clusters of the flowers-only cluster analysis ...	124
Fig. 5.12 Examples of the expression levels in the three clusters of the flowers-only cluster analysis	125

LIST OF FIGURES (continued)

<u>Figure</u>	<u>Page</u>
Fig. 5.13 Validation of RNA-seq expression with qPCR data	130

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 3.1 Numbers of mutants and rates of mutagenesis according to target gene, sgRNA, and clone	61
Table 3.2 Mutation types.	65
Table 4.1 CRISPR mutation rates on a per-event and per-allele basis	88
Table 5.1 Homologs in Arabidopsis of genes upregulated in flowers, capsules, and pollen when expression in compared to mature leaf.....	113
Table 5.1 Homologs in Arabidopsis of genes upregulated in flowers, capsules, and pollen when expression in compared to mature leaf.....	113
Table 5.2 Well-characterized homologs in <i>Arabidopsis</i> of upregulated genes in the seed capsule development cluster analysis	119
Table 5.3 Well-characterized homologs in <i>Arabidopsis</i> of upregulated genes the flowers-only cluster analysis	119
Table 5.4 Number of associated GO terms and annotated genes in AgriGO	126
Table 5.5 Heatmap representing expression of lignin-biosynthesis specific homologs in Eucalyptus.....	127

LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
Fig. S2.1 Putative pollen observed in 2007	179
Fig. S2.2 Abundant viable pollen from the non-transgenic control in 2009.....	180
Fig. S2.3 Non-transgenic control trees had longer catkins than transgenic trees in 2009	181
Fig. S2.4 Transgenic catkins were heavier than control catkins.....	182
Fig. S2.5 Transgenic catkins were curved and dark in tone (less red).....	183
Fig. S3.1 Diversity in putative amino acid modifications to the WT peptide sequence in 717.....	199
Fig. S4.1 Examples of partial peptide alignment of the N-terminal motif in mutants observed	215
Fig. S4.2 Leaf phenotypes of potted plants in WT trial.....	216
Fig. S4.3 Stem growth was reduced in plants that flowered precociously due to <i>AtFT</i> overexpression, but did not differ due to <i>ELFY</i> mutagenesis	217
Fig. S4.4 Developmental sequence of flower formation in the greenhouse	218
Fig. S4.5 Leaf phenotypes of potted plants in <i>FT</i> trial	219
Fig. S4.6 Flower buds and flowers of <i>AtFT</i> -only and FM events in a greenhouse trial at the University of Pretoria in South Africa	221
Fig. S4.7 3D representation of X-ray projections of inflorescences.....	222
Fig. S4.8 Underdeveloped organs appeared occasionally in organless mutant plants....	224
Fig. S4.9 Peptide alignment of the N-terminal domain in LFY and orthologous transcription factorsExamples of the expression levels in the three clusters of the flowers-only cluster analysis	225

LIST OF APPENDIX FIGURES (continued)

<u>Figure</u>	<u>Page</u>
Fig. S5.1 Poisson distance clustering of tissues.....	233
Fig. S5.2 Examination of variation among tissues including pollen and leaf.....	234
Fig. S5.3 Examination of variation among tissues excluding pollen and leaf	235
Fig. S5.4 Venn diagram of upregulated genes in flowers and capsules (LFC cutoff > 1, FDR < 0.05)	236
Fig. S5.5 Venn diagram of downregulated genes in flowers and capsules (LFC cutoff < -1, FDR < 0.05)	237

LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
Table S2.1 Genetic constructs used in this study.....	184
Table S2.2 Primers used in Polymerase Chain Reaction (PCR).....	185
Table S2.3 ANOVA table (one-way ANOVA) for the non-transgenic control trees that were distributed between the four transgenic constructs in the reporter trial for 2001	186
Table S2.4 ANOVA tables (one-way ANOVA) for the non-transgenic control trees that were distributed between the four transgenic constructs in the reporter trial for 2003	187
Table S2.5 ANOVA table (2-way ANOVA) with “Event” and “Block” as main effects for the sterility trial model	188
Table S2.6 Dunnett’s test for the sterility trial.....	189
Table S2.7 ANOVA table (one-way ANOVA) with “Construct” as main event for the reporter trial model for 2001	190
Table S2.8 ANOVA table (one-way ANOVA) with “Construct” as main event for the reporter trial model for 2003.....	191
Table S2.9 Dunnett’s test table for the reporter trial data from 2001	192
Table S2.10 Dunnett’s test table for the reporter trial data from 2003	193
Table S2.11 ANOVA tables (one-way ANOVA) with “Event” as main effect for each construct in the reporter trial for 2001	194
Table S2.12 ANOVA tables (one-way ANOVA) with “Event” as main effect for each construct in the reporter trial for 2003	195
Table S2.13 Dunnett’s test for the sterility trial data comparing catkin mean length of control to that of transgenic events for catkin collection from March 10, 2009.....	196
Table S2.14 Dunnett’s test for the sterility trial data comparing catkin mean length of control to that of transgenic events for catkin collection from March 17, 2009.....	197
Table S2.15 Catkin angle per event.	198

LIST OF APPENDIX TABLES (continued)

<u>Table</u>	<u>Page</u>
Table S3.1 Partial genetic sequence of the target genes and the off-target sites	200
Table S3.2 Table of primers, their sequence, and their specific use.....	203
Table S3.3 Lack of mutations on target sites in empty vector controls	205
Table S3.4 Mutation spectra of the different gene-sgRNA combinations with only one sgRNA.....	206
Table S3.5 Results table for the proportion comparison of all mutation spectra.....	207
Table S3.6 Mutation spectra generated by the same CRISPR Cas9 nuclease in the <i>PLFY</i> gene in two different hybrid poplar clones	208
Table S3.7 Results table for the proportion comparison of the mutation spectra of <i>LFY</i> -sg1sg2 in two different poplar clones	209
Table S3.8 Mutation spectra generated by the same CRISPR Cas9 nuclease in the <i>PAG1</i> gene in two different hybrid poplar clones	210
Table S3.9 Results table for the proportion comparison of the mutation spectra of <i>AG1</i> -sg1sg2 in two different poplar clones	211
Table S3.10 Mutation spectra generated by the same CRISPR Cas9 nuclease in the <i>PAG2</i> gene in two different hybrid poplar clones	212
Table S3.11 Results table for the proportion comparison of the mutation spectra of <i>AG2</i> -sg1sg2 in two different poplar clones	213
Table S3.12 Off-target sites studied for rate of mutagenesis.....	214
Table S4.1 Primers used for genotyping and sequencing	226
Table S4.2 Gene names and IDs for qPCR experiments	227
Table S4.3 Predicted loss-of-function (LOF) rates based on the number of frame-shifts, large deletions (i.e. ≥ 222 bp), and deletions of essential amino acids	228
Table S4.4 Phenotypes seen in FM (flowering mutant) events kept in the GH.....	229

LIST OF APPENDIX TABLES (continued)

<u>Table</u>	<u>Page</u>
Table S5.1 The 20 most upregulated genes in early flower vs leaf.	238
Table S5.2 The 20 most upregulated genes in late flower bagged vs leaf.	239
Table S5.3 The 20 most upregulated genes in late flower unconfined vs leaf	240
Table S5.4 The 20 most upregulated genes in early capsule vs leaf.	241
Table S5.5 The 20 most upregulated genes in late capsule vs leaf.	242
Table S5.6 The 20 most upregulated genes in mature pollen vs leaf	243
Table S5.7 The 20 most upregulated genes in cluster 1 when comparing amongst the flower and fruit tissues.	244
Table S5.8 The 20 most upregulated genes in cluster 2 when comparing amongst the flower and fruit tissues.	245
Table S5.9 The 20 most upregulated genes in cluster 3 when comparing amongst the flower and fruit tissues.	246
Table S5.10 The 20 most upregulated genes in cluster 4 when comparing amongst the flower and fruit tissues.	247
Table S5.11 The 20 most upregulated genes in cluster 1 when compareing EF, LFB, and LFU.	248
Table S5.12 The 20 most upregulated genes in cluster 2 when compareing EF, LFB, and LFU.	249
Table S5.13 The 20 most upregulated genes in cluster 3 when compareing EF, LFB, and LFU.	250
Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis	251
Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis	256

LIST OF APPENDIX TABLES (continued)

<u>Table</u>	<u>Page</u>
Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis	262
Table S5.17 Gene ontology significant terms for cluster four in the seed capsule development cluster analysis	270
Table S5.18 Gene ontology significant terms for cluster one in the flowers-only cluster analysis.....	274
Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis.....	277
Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis.....	285
Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast.....	290

DEDICATION

I dedicate my thesis to my grandfather Alejandro Elorriaga Aurrecoechea, who showed me the beauty of biology early on and whose love of trees I believe to have inherited.

Eskerrik asko Aitita maitea.

1 Introduction

Humans have selected, crossed, and grown plants for millennia. Many of the plants we consume today have been drastically changed compared to their wild ancestors (e.g., broccoli, tomato, corn, etc.). As humans, we have done the same selective breeding to our companion animals and the animals we raise for food (e.g., dogs, pigs, cows, etc.). Traditional selective breeding takes a long time and requires multiple generations. Introgression followed by backcrossing is a critical method, yet is unavailable in many types of plants, including forest trees, due to the time required and intolerance of inbreeding. Genetic engineering is a modern breeding technique that allows a biotechnologist to insert or modify a specific locus in the DNA with modest effects on the rest of the genome compared to sexual breeding. Genetic engineering however requires a delivery system to add the new piece of DNA and a culture system to regenerate an organism.

1.1 The origin of genetic engineering

Genetic engineering involves the addition of DNA from evolutionary unrelated or related sources, resulting in the generation of a genetically modified organism (GMO), which is either transgenic or cisgenic, respectively. In 1944, Avery et al. (1944) made the first GMO by inserting the DNA of encapsulated type II *Pneumococcus* to the unencapsulated type II *Pneumococcus*. The transformed type II *Pneumococcus* became encapsulated. During the following three decades, the ability of bacteria to uptake and integrate chromosomal DNA fragments from other bacteria was recognized as a natural mechanism of genetic recombination. Around that time, Cohen et al. (1973) constructed synthetic plasmids using restriction enzymes and transformed them into *Escherichia coli* (*E. coli*).

The first transgenic animals, mice expressing thymidine kinase from the herpes virus, were created in 1981 (Brinster *et al.*, 1981). The first transgenic rabbits and pigs expressing human growth hormone (hGH) were produced soon after using the same technique, direct microinjection into eggs cells (Hammer *et al.*, 1985). The first

transgenic plants were antibiotic resistant tobacco lines generated using *Agrobacterium*-mediated transformation in 1984 by the collaboration between the Schell and van Montagu labs (De Block *et al.*, 1984). Today, there are more than twenty five different transgenic crops grown in 24 countries, and additional 43 countries import GM foods for food, feed, and processing (ISAAA, 2017). Based on a literature meta-analysis done by Klümper and Matin (2014), transgenic crops had on average increased yield by 22%, decreased pesticide usage by 37%, decreased pesticide cost by 39%, and increased profits for GM farmers by an average of 68%.

1.2 The generation of transgenic plants

The first transgenic plants were antibiotic resistant tobacco generated using *Agrobacterium*-mediated transformation (De Block *et al.*, 1984). The first field trials took place in 1986 in both France and the USA with herbicide resistant tobacco (James & Krattiger, 1996). In 1987, the Belgian company founded by Marc Van Montagu and Jeff Schell, Plant Genetic Systems, was the first company to engineer insect resistant plants (Vaeck *et al.*, 1987). Golden Rice, the first food product engineering with increased nutrient value, was first developed in 2000 (Ye *et al.*, 2000).

1.2.1 *Agrobacterium*-mediated transformation

Agrobacterium-mediated plant transformation was the first transformation method developed. The work necessary to transform plants using dis-armed *Agrobacteria* took almost a century. In the late 19th century, “crown gall” was the term coined to describe the outgrowths that were appearing in the roots of fruit trees (Smith, 1894). By the early 1900s, large numbers of fruit trees in nurseries were getting infected with crown gall disease. Smith and Townsend (1907) proved that crown gall was caused by a bacterium; a controversial hypothesis for the time since plants were not believed to be hosts to bacterial diseases.

Based on the fact that plants infected with *Agrobacteria* would not die or wilt (instead they grow crown galls), in 1947 Braun introduced the idea that the *Agrobacterium* transferred the ability of cell-proliferation and tumor-formation to the plant (Braun,

1947). He called this transferred ability the tumor-inducing principle (TIP). In the 1960s, DNA from the bacteria was found in DNA isolated from sterile cultures of crown-gall *Nicotiana* cells and bacterial genes were identified in crown-gall tissues using bacterial antigens (Schilperoort *et al.*, 1967). These findings provided evidence that DNA was being transferred from the bacteria to the plant. In 1974, Zaenen *et al.* (1974) published work showing that the pathogenicity of *Agrobacterium* was caused by a large supercoiled plasmid. They named the plasmid the Ti (i.e., tumor-inducing) plasmid.

In the 1980s, the T-DNA border sequences were identified and shown to define the section of the plasmid that would be inserted into the nuclear (not the mitochondrial or chloroplastic) genome of the plant (i.e., the T-DNA) (Chilton *et al.*, 1980; Zambryski *et al.*, 1980; Willmitzer *et al.*, 1980; Lemmers *et al.*, 1980). During the same year, the functional genetic organization of two Ti plasmids (i.e., nopaline plasmid pTiC58 and octopine plasmid pTiB6S3) was deciphered using transposon-insertion mutagenesis (Holsters *et al.*, 1980; De Greve *et al.*, 1981). Mutations that eliminated the plasmid's oncogenicity mapped to T-DNA sections homologous to both plasmids indicating that these regions are essential for tumor induction. Other regions non-essential to oncogenicity also mapped to the T-DNA, including segments involved in nopaline biosynthesis and octopine biosynthesis. Transposon-mediated mutations in the Ti plasmid that did not map to the T-DNA helped map the *virulence* (*vir*) genes (Hernalsteens *et al.*, 1978; Dhaese *et al.*, 1979; Holsters *et al.*, 1980). The *vir* genes were determined to be responsible for the processing, delivery, and insertion of the T-DNA. The first non-oncogenic plasmid (pGV3850) was created by Zambryski *et al.* (1982).

For transfer of DNA to happen, both the T-DNA border repeat sequences and the *vir* genes are needed, and as long as the *vir* genes are present in the same *Agrobacterium* cell as the T-DNA, the T-DNA will be delivered to the host's genome. This finding was published by two labs in May 1983: Framond *et al.* (1983) and Hoekema *et al.* (1983). They transformed *Agrobacteria* with two separate plasmids: a shuttle plasmid containing the T-DNA flanked by the borders, and a helper plasmid containing the *vir* genes.

1.2.2 Biolistics or the gene gun

In biolistics, DNA-coated metal (usually gold or tungsten) beads are propelled into cells. This technique was first used in 1987 to transiently transform onion cells (Klein *et al.*, 1987). At the time, *Agrobacterium* was thought to only infect dicots, making biolistics an important method for transforming monocots. The first biolistics-mediated stably transformed plants were generated from meristems of immature soybean seeds (McCabe *et al.*, 1988) and from leaf tissue and suspension cultures of tobacco (Klein *et al.*, 1988). Since then, biolistics has also been used to transform bacteria (Smith *et al.*, 1992), fungi (Toffaletti *et al.*, 1993) and animal cells (Johnston *et al.*, 1991). In the past, biolistics was broadly used to transform cereal crops, including economically-important crops such as rice, maize, and wheat.

1.2.3 Other techniques

There are other methods available for generating transgenic plants. However, they tend to be more laborious and less efficient. In 1984, the first cauliflower mosaic virus (CaMV) mediated-transformed plants were generated (Brisson *et al.*, 1984). However, the CaMV transformation system has an upper limit of 250bp for transgene size. This size limitation effectively ended scientific research in this system. Protoplast isolation followed by plant regeneration were methods established in the 1970s (Takebe *et al.*, 1971). Protoplasts are plant cells that have had their cell wall removed enzymatically. Methods available to bacteria and animal cells can be used with protoplasts. Microinjection of protoplasts was established in 1985 by using a glass capillary to insert DNA into the protoplasts (Lawrence & Davies, 1985). Between 1985 and 1986, electroporation of protoplasts was also established as a plant transformation method (Fromm *et al.*, 1985, 1986).

1.3 Plant tissue culture

Plant tissue culture involves the *in vitro* aseptic growth of cells, tissues, organs, and whole plants under controlled environmental and nutritional conditions. Tissues, organs, and whole plants are kept in artificial media with growth regulators. Henri Louis-

Duhumel du Monceau, while studying wound-healing in plants, noticed the formation of callus (Monceau, 1758). After multiple microscopic studies on calli, independently Schleiden (1838) and Schwann (1839) formulated the theory that cells are the smallest unit of an organism and they are autonomous.

In 1902, physiologist Gottlieb Haberlandt wanted to prove that plants cells remained totipotent even after differentiation. Thus he attempted to induce cell division of fully differentiated cells with the goal of generating whole plants. He isolated palisade cells from the leaves of *Lamium purpureum*, the petioles of *Eichhornia crassipes*, the glandular hairs of *Pulmonaria mollissima*, and the stamen hairs of *Tradescantia*, and suspended them in a Knop's salt solution with sucrose. The cells accumulated starch, increased in size, and were alive for a month (Haberlandt, 1902). However, they failed to divide (Haberlandt, 1902). Haberlandt also believed that embryos could be formed from vegetative tissues. In other words, plant cells could be reprogrammed if given the correct stimulus. Despite his lack of success in the lab, he is considered the father of plant tissue culture because of his revolutionary thinking.

Molliard (1921) in France, Kotte (1922) in Germany, and Robbins (1922) in the USA, successfully cultured meristematic cells from embryos, root tips, and shoot tips respectively. Between 1932 and 1934, White started root tips cultures from tomato that were maintained until his death in 1968 (White, 1934b). White's media contained inorganic salts, yeast extract, and sucrose. White replaced yeast extract with B vitamins in his media recipe a short time (White, 1937a,b). Went (1926) isolated indole acetic acid, a form of auxin. In 1939, Gautheret added indoleacetic acid and B vitamins into his media, and was successful at establishing continuous carrot root cambium cell cultures (Gautheret, 1934, 1939). In the same year, following Gautheret's recipe, White (1939) and Nobecourt (1939) established continuous cultures from tumor tissue in *Nicotiana* and carrot, respectively. The recipes used today in tissue culture of differentiated tissue are all based on the work of Gautheret, White, and Nobecourt.

Embryo culture began in the early 1900's. In 1904, Hannig (1904a,b) cultured embryos from *Brassica* and in 1906 Brown (1906) cultured embryos from barley. Hannig and

Brown both removed nearly mature embryos and developed them to maturity on simple media. In 1925, Laibach (1925) was successful at obtaining whole plants of rare lily hybrids after rescuing embryos that would be aborted in nature. In 1934, Tukey (1934) succeeded at implementing embryo rescue on cherry plants high rates of underdeveloped embryos. Since then embryo rescue has become a useful tool for breeders, especially for hybrid crosses with post-zygotic sexual incompatibility.

White (1934a), Limasset, and Cornuet (1949) observed that root and shoot cultures derived from virus-infected tissues would sometimes become free of viruses.. However, it was Morel who realized the potential of rapid propagation using this method. He revolutionized the orchid industry when he realized that using shoot cell culture would allow for the virus-free propagation of plants. In 1960, he was able to generate almost four million orchids from one single bud in one year (1960). Since then, shoot tip culture with/without chemotherapy or thermotherapy has become the primary means to generate virus-free plants.

Miller *et al.* (1955b,a) isolated the first cell division hormone, kinetin. In 1957, they deciphered that auxin and cytokinin interplay was needed for organ development (Skoog & Miller, 1957). Relative higher concentrations of cytokinin lead to shoot formation while relative higher concentrations of auxin induce root formation.

Today, there are well established tissue culture systems for many plant species, including many woody plants (Raghavan, 1986; Krishnaraj & Vasil, 1995; Lakshmanan & Taji, 2000; Jain *et al.*, 2013). The main component in media are inorganic salts, plant growth regulators, vitamins, and a source of carbon. The main minerals found in the inorganic salts used in tissue culture are boron, calcium, cobalt, copper, iron, magnesium, molybdenum, nitrogen, potassium, sodium, and zinc. The main plant growth regulators are abscisic acid, auxins, cytokinins, and giberellins. The most common vitamins used are thiamine (B₁), nicotinic acid (B₃), and pyroxidine (B₆). The most commonly used carbon source is sucrose, however different carbon sources (i.e., carbohydrates) can have distinct effects on the morphogenesis of different tissues and/or different plant species

(Yaseen *et al.*, 2013). For example, sorbitol seems to be a better carbon source for some genotypes of different species of the Rosaceae family.

1.4 Regulation of genetically modified plants

1.4.1 History

The Cartagena Protocol on Biosafety to the Convention on Biological Diversity is an international agreement drafted and signed by the United Nations in 2000. This agreement took effect on September 11 2003. This agreement specifies that products from new technologies must follow the precautionary principle before becoming available to the public. The precautionary principle was one of 27 principles in the Rio Declaration on Environment and Development, a short document produced by the United Nations during the 1992 Earth Summit. The precautionary principle (i.e., principle 15) states that "In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation."

1.4.2 USA

GMO plants are regulated by three different government agencies in the USA; the US Environmental Protection Agency (EPA), the US Food and Drug Administration (FDA), and the US Department of Agriculture (USDA). The EPA protects the environment and the health of the American public following the Federal Insecticide, Fungicide, and Rodenticide Act (National Academies of Sciences *et al.*, 2016). The FDA is responsible for the safety of the food and feed for human and animal consumption respectively under the Federal Food, Drug, and Cosmetic Act. The USDA has the responsibility of keep the American agriculture safe from pest and diseases through the Biotechnology Regulatory Services of the Animal and Plant Health Inspection Service (APHIS). Thus, APHIS must decide when a GMO plant poses such a risk (National Academies of Sciences *et al.*, 2016).

It is expected that varieties of plants generated with new breeding techniques including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) Cas nucleases that do not contain the transgene (i.e., null segregants) will not be regulated by USDA in the USA as traditional GMOs (Waltz, 2018). Already, multiple food products have been exempted from regulation by USDA including high-oil producing camelina, herbicide-resistant canola, high-amylopectin starch content corn, leaf blight resistant corn, low-phytate corn, non-browning button mushroom, non-browning potato, delayed-flowering *Setaria viridis*, drought- and salt-tolerant soybean, high oleic acid soybean, and powdery mildew resistant wheat (Waltz, 2016, 2018; Kumar *et al.*, 2020). EPA regulation of gene edited crops with plant pest tolerance traits remains unclear (Friedrichs *et al.*, 2019).

1.4.3 Canada

Two agencies from the Canadian governments regulate all novel agricultural plant products. The Canadian Food Inspection Agency (CFIA) is in charge of regulating plants and feed with novel traits and Health Canada is responsible for regulating food products for human consumption. Following both the Seeds Act and the Feeds Act, CFIA and Health Canada must assess the safety of crops with novel traits (“novelty” being the trigger) in agriculture and human consumption, respectively.

1.4.4 European Union

The regulatory framework of the EU differentiates GM plants that will be released into the environment (i.e., deliberate release) from plants being used for food and feed. New GM plants intended for human consumption or animal feed submit application for regulatory approval to the respective authority of the member state. The authority forwards the application to the European Food Safety Authority (EFSA) which evaluated the product for safety. If the product is to be cultivated, the EFSA

1.4.5 Other countries

In Argentina, the National Service of Agricultural and Food Health and Quality (SENASA), the National Advisory Commission on Agricultural Biotechnology (CONABIA), the National Direction of Agricultural Food Markets (DNMA), and the National Institute of Seeds (INASE) are the four regulatory agencies under the Secretariat of Agriculture, Livestock, Fisheries and Food (SAGPyA) that are responsible of regulating GM plants. SENASA assesses the safety of GM plants for human and animal consumption. CONABIA and DNMA appraise the impacts of GMOs in the country's agriculture and the economy respectively. INASE regulates the registration and commercialization of GM seeds. In Brazil, all evaluations on the safety of GM plants in the environment, human food, and feed are conducted by the National Technical Commission of Biosafety (CTNBio) under the Brazilian National Biosafety Council (CNBS). In India, two governmental agencies, the Ministry of Environment, Forests and Climate Change and the Department of Biotechnology, are responsible for implementing the 1989 Biosafety Rules. These rules regulate all of genetic engineering, from application of GM plants to derivative products. The Genetic Engineering Appraisal Committee is the main authority that regulates the manufacture, sale, import, and export of all GMOs in the country.

1.5 Commercial applications of genetically modified plants

In 1992 China introduced virus-resistant tobacco, the first GM plant for public use. The first GM crop approved for commercial sale in the USA was the *FlavrSavr* tomato introduced in 1994 (Kramer & Redenbaugh, 1994). This tomato had a longer shelf-life than conventional tomato. In 1994 Europe also approved its first GM crop, herbicide-resistant tobacco. Many products were approved for human consumption in 1995: *Bacillus thuringiensis* (*Bt*) Potato, canola oil (Calgene), *Bt* maize (Ciba-Geigy), herbicide-resistant cotton (Calgene), *Bt* cotton (Monsanto) herbicide-resistant soybeans (Monsanto), virus-resistant squash (Asgrow), and other tomatoes with long shelf-life (DNAP, Zeneca/Peto, and Monsanto). Among other transgenic plants available commercially at some point, are non-browning apples (Okanagan Specialty Fruits),

purple and lavender carnations (Florigene Ltd. and Suntory Ltd.), improved-vase life carnations (Florigene Ltd.), virus-resistant common bean (Embrapa), *Bt* eggplant (Mahyco, Monsanto, and Cornell University), delayed-ripening melon (Agritope Inc.), virus-resistant papaya (University of Hawaii and Cornell University), orange petunia (Planck Institute for Plant Breeding Research), extra sweet pink flesh pineapple (Del Monte Fresh Produce Company), virus-resistant plum (USDA), *Bt* potato (Centre Bioengineering, Russian Academy of Science), non-browning potato (Simplot), *Bt* rice (Agricultural Biotechnology Research Institute, Huazhong Agricultural University), hypoallergenic rice (National Institute of Agrobiological Sciences), purple rose (Florigene Ltd. and Suntory Ltd.), drought-resistant sugar cane (Persero), and virus-resistant sweet pepper (Beijing University) (reviewed in Hallerman & Grabau, 2016; Baranski *et al.*, 2019).

For the most planted biotech crops, i.e., canola, cotton, maize, and soybean, countries where farmers can choose between conventional and transgenic seeds, GM plants have superseded conventional ones for the last two decades. GM varieties are more likely to be planted 93 to 100% of the time in Argentina (soybean), Brazil (soybean), Canada (oilseed rape), China (cotton), India (cotton), and the USA (cotton, maize, and soybean) (ISAAA, 2017). Transgenic varieties are preferred because they reduce pesticide use, ease weed control, and in consequence reduce labor and increase profits (Klümper & Matin, 2014). As an added bonus, *Bt* varieties have reduced the amount of pesticide used by farmers. In 2017, 189.8 million hectares of GM crops were planted in 24 countries by 17 million farmers (ISAAA, 2017).

1.5.1 Forest tree applications

Forest tree plantations provide timber, pulp, food, shade, and shelter, among others to humans and other animals in the area. *Pinus* and *Eucalyptus* are the most planted gymnosperm and angiosperm genera in the world, respectively. Eucalypts have been an important wood crop and ornamental species since the early nineteenth century (Turnbull, 2000). Genetically engineered forest trees have shown improved wood quality, improved biotic or/and abiotic stress tolerance, higher photosynthetic efficiency, higher

biomass production efficiency, among others (Harfouche *et al.*, 2011; Porth, 2014; Etchells *et al.*, 2015; Ault *et al.*, 2016; Zhou *et al.*, 2017). However, there are only three deregulated GMO forest trees in production— two lines of transgenic poplar expressing the *Bt* toxin in China (James, 2015) and one line of transgenic *Eucalyptus* with 15-20% faster growth in Brazil (Anonymous, 2015).

1.6 Risks of genetically modified plants

The benefits provided since their advent by glyphosate-resistant and *Bt* crops are being threatened by lack of weed- and insect-resistance management. Weeds and target insects have evolved resistance to glyphosate and *Bt* toxin respectively. Glyphosate has become one of the most popular herbicides since the appearance of glyphosate-resistant crops in 1996 (Duke & Powles, 2008). A total of 38 glyphosate-resistant weed species have been discovered in 37 countries (Heap & Duke, 2018). Weeds have become resistant by one or w combination of target gene mutation, target gene duplication, enhanced metabolism, decreased absorption, decreased translocation to other tissues, and sequestration of the active ingredient (Heap & Duke, 2018). Glyphosate will remain sustainable if effective weed management practices are applied. Other herbicide resistance traits have been developed in the industry. Also, RNAi technology holds promise for herbicide-resistant weed control. However, if weed management practices are not followed, weeds resistant to these new herbicidal traits and to RNAi will also evolve. Weed managements must involve: applying the recommended dose level to avoid low-level resistant weeds to survive and reproduce, herbicide mixtures and rotations, and mechanical deweeding (Duke & Powles, 2008; Heap & Duke, 2018). Techniques including weed identification by image analysis, precision herbicide application, and robotic de-weeding are expected to replace chemical herbicides in the future (Heap & Duke, 2018).

Five lines of *Bt* cotton were introduced in the USA between 1996 and 2015: Bollgard, Bollgard II, WideStrike, TwinLink, and WideStrike 3 (Fleming *et al.*, 2018). Bollgard has one *Bt* transgene, while Bollgard II, WideStrike, and TwinLink have two and WideStrike 3 has three. Between the years 2010 and 2015, cotton yield benefits and efficacy of Bollgard II and WideStrike decline over time and declined over time. Insect

resistance to *Bt* cotton has also surged in India, and resistance to *Bt* corn has been noted in Argentina, Brazil, South Africa, and the USA (Tabashnik & Carrière, 2017). Target insects evolve resistance to *Bt* crops by avoiding toxin activation in their gut, by mutating the genes involved in *Bt* toxin receptor development genes, or by modulating their immune system (Xiao & Wu, 2019). Resistance can be delayed by having abundant refuges of non-*Bt* plants, planting *Bt* varieties with one toxin gene separate from varieties with two or three genes, and by pyramiding RNAi and *Bt* toxin genes (Carrière *et al.*, 2004; Tabashnik *et al.*, 2013; Tabashnik & Carrière, 2017)

1.7 Society's perception of GMOs

According to recent polls, the large majority of the non-scientific community is concerned about the use of GMOs in agriculture and food production (Inc, 2007) and supports stringent labeling ('Strong Support for Labeling Modified Foods - The New York Times'). According to a decade worth of safety research, the scientific record shows that GMOs consumed by humans and animals are no more harmful for health than other kinds of foods (Nicolia *et al.*, 2014). However, creating proactive measures to mitigate gene flow might help relieve some of society's concerns with GMOS. A way to avoid most unwanted gene flow is to completely remove the plant's ability to reproduce sexual offspring; i.e., genetic containment. This is a greater issue for crops with extensive wild or feral relatives, such as many forest trees (Strauss *et al.*, 2009a, 2009b). Where eucalypt gene flow is a social or ecological concern, engineering desired varieties to be sterile may allow for easier adaptation and acceptability in industry. However, as shown by Terminator technology (discussed below), social acceptance may not be simple.

1.8 Genetic use restriction technologies or terminator technology

Genetic use restriction technologies (GURTs) were envisioned as intellectual property protection technology (IPP) (Van Acker *et al.*, 2007). GURTs can be either trait-based GURTs (T-GURTs) or variety-based GURTs (V-GURTs). T-GURTs control the expression of the trait of choice by means of an inducible promoter. V-GURTs control the use of a specific crop variety by hampering its reproductive ability. One of the first

GURTs proposed was a V-GURT known as ‘Technology Protection System’ or ‘GeneSafe Technologies’ (Oliver *et al.*, 1998, 1999). When this technology is transformed to plants, the seeds lose the ability to germinate (Oliver & Hake, 2012). The critics, who referred to the technology as ‘Terminator technology’, were successful in pressuring Monsanto to abandon plans of creating GeneSafe seeds (Niiler, 1999) and the technology has not been used commercially to date. The critics argued that conventional varieties open-pollinated with the GeneSafe crops would have reduced viable seed rates. In addition to the protection of intellectual property, V-GURTs would by default reduce the chances of transgene flow (Sang *et al.*, 2013). However, one important difference between designing a GURT that guarantees farmers buy seed every year and a GURT for reliable and full genetic containment is the difference between the high rates of inviable seeds that may be adequate for industrial purposes, and the 100% rate that may be required for stringent containment (Sang *et al.*, 2013).

1.9 Genetic containment techniques

The main categories of genetic containment technologies are: (1) tissue ablation, where reproductive tissues are destroyed during development by a cytotoxin; (2) transgene excision, where the transgene is removed prior to commercial release; (3) delayed flowering, where the onset of flowering is delayed; (4) gene suppression, when gene expression of an essential flowering or reproductive gene is significantly reduced; (5) gene knockout, when the DNA sequence of one or more essential flowering or reproductive genes is mutated, rendering the gene non-functional (reviewed in Brunner *et al.*, 2007).

For ablation, the cytotoxins used in ablation techniques are one of the following: RNAses, protein synthesis inhibitors, DNAses, proteases, glucanases, or lipases, usually under the expression of a tissue-specific floral promoter (Mariani *et al.*, 1990; Kuvshinov *et al.*, 2001; Cho *et al.*, 2001; Guerineau *et al.*, 2003; Skinner *et al.*, 2003; Lee *et al.*, 2003; Chrimes *et al.*, 2005; Luo *et al.*, 2005; Höfig *et al.*, 2006). One of the most common ways to avoid unwanted gene flow is to create male sterile plants via tissue ablation. The lack of loose pollen ensures that these varieties must be hand pollinated.

Excision of the transgenes usually involves the use of a recombinase such as Cre/*lox* from bacteriophage P1 or FLP/*FRT* from *Saccharomyces cerevisiae*. The T-DNA is designed such that it is flanked by the recombinase's recognition sites and the recombinases are expressed only in floral tissues or during induction by an external stimulus.

The constitutive or selective expression of floral repressor genes can delay the onset of flowering significantly among and within trees (Borner *et al.*, 2000; Scortecci *et al.*, 2001; Kotoda & Wada, 2005; Danilevskaya *et al.*, 2010).

Gene suppression can be achieved using either RNA interference (i.e., RNAi), where the stable transformation of transgenes containing inverted repeat or hairpin sequences that match transcripts of endogenous floral genes result in the cleavage and reduce translation of such transcripts (Zhang *et al.*, 2001; Goetz *et al.*, 2001; Mou *et al.*, 2002; Yui *et al.*, 2003; Klocko *et al.*, 2016a; Lu *et al.*, 2018a); or dominant negative proteins, where a dominant non-functional form of an endogenous transcription factor or signal transduction protein reduces WT gene dosage enough to eliminate fertility. The dominant negative protein can be a translational fusion of a transcriptional factor and a repression motif (Mamun, 2007).

Last, gene knockout involves the use of site-directed nucleases such as CRISPR nucleases or TALENs to induce mutations in gene of interest that render them non-functional after incorrect DNA repair (Zou *et al.*, 2017; Chen *et al.*, 2018b; Okada *et al.*, 2019).

1.10 CRISPR technology and applications

The potential of site-directed mutagenesis in plants has advanced significantly in the past seven years thanks to the arrival of site-specific nucleases (SSNs) (Chen and Gao, 2014). Mutagenesis in specific chromosomal sections has not been readily available in plants – only random mutagenesis by chance using gamma rays and ethyl methanesulfonate (EMS) – mostly due to their repair mechanisms (Voytas, 2013; Weinthal *et al.*, 2010). After the advent of the first SSN technology, Zinc Finger Nucleases (ZFNs), we have witnessed the emergence of two other nuclease technologies that show much promise in

genetic mutation, genetic therapeutics, and also crop improvement (Chen and Gao, 2014; Gaj et al., 2013). These technologies are transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas-mediated RNA-guided DNA endonucleases (CRISPR Cas nucleases for short). Using CRISPR Cas nucleases for directed mutagenesis of multiple genes is easier, more affordable, and usually much more efficient than using ZFNs or TALENs (Sander and Joung, 2014). And, unlike ZFNs and TALENs, CRISPR Cas nucleases are not sensitive to DNA methylation (Hsu et al., 2013).

1.10.1 CRISPR Cas9 nucleases

The first synthetic CRISPR Cas nuclease system developed (i.e., the CRISPR Cas9 system) was based on the type II CRISPR Cas RNA-guided system naturally found in *Streptococcus pyogenes*, which the bacterium uses as adaptive immunity against invading phages and plasmids (Barrangou, 2013). In *S. pyogenes*, the association of Cas9 enzyme with a short guide sequence (~20 nucleotides), known as CRISPR RNA (crRNA), and a trans-activating CRISPR RNA (tracrRNA), allows the system to find and cleave foreign invading DNA. The complex, formed by Cas9, tracrRNA, and crRNA, finds the sequence complementary to that of the crRNA and induces a double stranded break (DSB). Cas9 will only cleave the target DNA when followed by the sequence 5'-NRG (Hsu et al., 2013) which is known as the protospacer adjacent motif (PAM). DSBs are typically induced three to four nucleotides upstream of the PAM site (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). This system is considered adaptive because it allows the bacterium to acquire and store short sequences from invading phages and plasmids in its CRISPR locus (or loci). The bacterium is protected during future infections because the stored plasmid or phage sequences are processed into crRNAs during CRISPR Cas activated immunity (Barrangou, 2013). The acquired DNA sequences located in the locus are known as 'spacers,' while the matching DNA sequences in the phages or plasmids are known as 'proto-spacers.' In the locus, the spacers are separated by palindromic (for forming hairpins) repeat sequences known

simply as ‘repeats.’ CRISPR loci have been found in approximately 45% of bacterial and 90% of archaeal genomes (Barrangou, 2013).

CRISPR Cas9 nucleases became a popular synthetic biology tool after Jinek *et al.* (2012) demonstrated that a chimeric tracrRNA-crRNA synthetic fusion, known as “single guide RNA” (sgRNA), can be as effective as the wildtype crRNA and tracrRNA dimer in interacting with Cas9 and inducing Cas9-directed cleavage of target DNA. Since 2013, genes of more than 26 different plant species have been modified with CRISPR Cas9 nucleases (Bewg *et al.*, 2018a; Xu *et al.*, 2019; Ghogare *et al.*, 2019; Manghwar *et al.*, 2019). Mutation efficiencies of endogenous genes vary between 0.1% and 100%. Mutation efficiencies of 100% have been reported in cassava (Odipio *et al.*, 2017), eucalypts (Elorriaga *et al.*, under review), grapevine (Ren *et al.*, 2016), maize (Lee *et al.*, 2019), poplar (Zhou *et al.*, 2015; Wang *et al.*, 2017), rice (Shen *et al.*, 2017), tomato (Ueta *et al.*, 2017), and Wanjincheng orange (Peng *et al.*, 2017). Inheritable mutations have been reported in *Arabidopsis*, maize, potato, rice, and wheat (Feng *et al.*, 2014; Zhang *et al.*, 2014; Wang *et al.*, 2014; Zhou *et al.*, 2014a; Ma *et al.*, 2015; Butler *et al.*, 2015; Zhu *et al.*, 2016).

1.10.2 Modified Cas9 systems

Cas9 has had one or both (i.e., dead Cas9 or dCas9) of its nuclease domains (i.e., RuvC and HNH domain) inactivated to generate Cas9 nickases (i.e., only one strand is cleaved), to modify the PAM site, to reduce offtargeting, or just to repress transcription (reviewed in Mitsunobu *et al.*, 2017). dCas9 has also been fused to the FokI nuclease domain, cytosine and adenine deaminases, repressor domains, activator domains, fluorescent protein sequences, and histone demethyling domains to create more specific nucleases, site-directed base editors, transcription repressors (i.e., CRISPRi), transcription activators, site-directed visual markers, and epigenetic repressors respectively (Mitsunobu *et al.*, 2017; Schindele *et al.*, 2019). Last, split and inducible Cas9 variants have been engineered for viral-vector delivery and spatiotemporal- or dosage-controlled targeting respectively (reviewed in Mitsunobu *et al.*, 2017).

1.10.3 Other CRISPR Cas systems

Other Cas enzymes have recently been characterized. Cas12 and Cas13 are Class 2 Cas enzymes as Cas9 but with major differences (Shmakov *et al.*, 2017). For starters, neither of them requires a tracrRNA to interact with the crRNA (Fonfara *et al.*, 2016; Swarts *et al.*, 2017; Abudayyeh *et al.*, 2017). Cas12 is an RNA-guide DNA-targeting nucleases as Cas9. However, Cas12 enzymes have 5'-TTTN-3' or 5'-TTN-3' as their PAM sequences, the PAM sites must be on the 5' end of the protospacer, they cleave DNA staggerly, and they don't, yet they cleave on the 3' as Cas9 (Zetsche *et al.*, 2015; Gao *et al.*, 2016). Meanwhile, Cas13 is an RNA-guided RNA targeting enzyme. Cas13 appears to cleave RNAs *in-vitro* in a non-specific manner (Liu *et al.*, 2017b,a), however this promiscuous cleaving activity has not been seen with human cells (Abudayyeh *et al.*, 2017; Cox *et al.*, 2017). Both Cas12 and Cas13 with likely be engineered just like Cas9 for other functions including gene repression, gene activation, and site-directed visual localization.

1.11 Floral molecular biology

Because plants are stationary their reproductive success and survival depends mostly on them gauging their environment correctly when transitioning into critical stages in their life cycle. One of the most critical stages is the transition to flowering. Flowering must happen just at the right time to guarantee the species and specimen's genes live on (Andrés & Coupland, 2012). Inductive conditions for flowering vary between and within species. Plants can be either long- or short-day, and they might or not require prolonged exposure to cold. Annual, biennial, and perennial plants differ in the number of growing sessions required to complete a life cycle.

The onset of flowering can have effects on phyllotaxis, meristem fate, and meristem identity. Plants can have monopodial or sympodial growth (reviewed by Reinhardt & Kuhlemeier, 2002). Monopodial growth happens when the shoot apical meristem (SAM) remains indeterminate for the entire life of the plant. Sympodial growth occurs when the SAM is determine and development continues from lateral meristems. During the transition to flowering in plants with monopodial architecture like *Arabidopsis*, the shoot

apical meristem (SAM) transitions into an inflorescence meristem (IM). Then, floral meristems (FM) develop from each IM and begin the formation of floral organ primordia. *Arabidopsis* and *Antirrhinum*, two of the model plants, exhibit monopodial growth patterns. Meanwhile, tobacco and petunia, exhibit sympodial growth patterns.

Because of the importance of flowerings to plant fitness, its onset is subject to several interacting layers of control. Flowering in *Arabidopsis* is dependent on the integration of six pathways: age, ambient temperature, autonomous, gibberellin, photoperiod, and vernalization pathways (Wellmer & Riechmann, 2010; Leijten *et al.*, 2018). The age pathway involves several SQUAMOSA PROMOTER BINDING LIKE (SPL) transcription factors. Over time, the concentrations of SPLs increase, and in turn, SPLs induce expression of other transcription factors, including *FRUITFULL (FUL)*, *LEAFY (LFY)*, and *SUPPRESSOR OF CONSTANS 1 (SOC1)*. SPLs are regulated by microRNA miR-156. Concentration of miR-156 is higher in younger plants, and it decreases as the plants grow.

The ambient temperature pathway involves the activity of many transcription factors including the MADS-box transcription factor SHORT VEGETATIVE PHASE (SVP). Mutations in *SVP* lead to early flowering and insensitivity to ambient temperature in *Arabidopsis*. *SVP* regulates expression of *FLOWERING LOCUS T (FT)*. The concentration of gibberellin (GA) increases substantially in the flower meristem right before induction of flowering. *Arabidopsis* flowers after prolonged exposure to long days (the photoperiod pathway). The exposure to long days initiates a regulatory network involving GIGANTEA (GI) and CONSTANS (CO). CO is another MADS-box transcription factor that induces expression of *FT* and *TWIN SISTER OF FT (TSF)*. CO is degraded during both light and dark. During the day, a pathway started by photoreceptor Phytochrome B (PHYB) degrades CO, and in the dark, the ubiquitin ligase CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) degrades CO. GI interacts with F-box ubiquitin ligases, and the interaction stabilizes the ubiquitin ligation, which go on a degrade repressors of CO.

In both the autonomous and vernalization pathways, flowering induction happens by repression of *FLOWERING LOCUS C (FLC)*, a MADS-box transcription factor and strong repressor of flowering. FLC represses flowering by directly binding to the promoters and obstructing transcription of *FT*, *SOC1*, and *SPL15*. The absence of warmth for a prolonged period of time induces repression of *FLC* by histone modification of its chromatin. VERNALIZATION INSENSITIVE 3 (VIN3) and VERNALIZATION 2 (VRN2) are among the proteins involved in the long-term repression of *FLC*.

All of these pathways converge with three flowering pathway integrator (FPI) genes, *FT*, *SOC1* and *LFY*. *FT* is a member of the phosphatidylethanolamine-binding protein (PEBP). FT protein moves from the leaves to the SAM with help of FT-INTERACTING PROTEIN 1 (FTIP1). When FT reaches the SAM, it interacts with FLOWERING LOCUS D (FD) with help of the 14-3-3 protein, to induce expression of *SOC1*. *SOC1* is a MADS-box transcription factor that induces expression of *LFY* and *AGAMOUS-LIKE 24 (AGL24)*.

LEAFY (LFY) was one of the first ever identified flowering genes (Coen *et al.*, 1990a; Weigel *et al.*, 1992). It encodes a highly conserved plant-specific transcription factor found in all land plants (Moyroud *et al.*, 2009; Silva *et al.*, 2016) and streptophyte algae (Gao *et al.*, 2019b). *LFY* is an FPI gene but also a floral meristem identity (FMI) gene. FPI genes initiate transition into reproductive growth when the plant is ready and FMI genes convert inflorescence meristems to floral meristems by promoting flower initiation. In *Arabidopsis*, the main FMI genes are *APETALA1 (API)*, *CAULIFLOWER (CAL)*, *FRUITFUL (FUL)*, and *LFY*. In *Arabidopsis*, the four FMI genes commit the meristem to a floral fate by regulating the expression and function of *TERMINAL FLOWER1 (TFL1)*, a floral meristem repressor (Bowman *et al.*, 1993; Ratcliffe *et al.*, 1999; Liljegren *et al.*, 1999; Kobayashi *et al.*, 1999; Ferrandiz *et al.*, 2000; Parcy *et al.*, 2002; Jaeger *et al.*, 2013; Serrano-Mislata *et al.*, 2016, 2017; Goslin *et al.*, 2017). During flowering *LFY* activates many floral organ identity (FOI, or homeotic) genes including *APETALA1 (API)*, which then itself induces more *LFY* expression, generating a feed-forward loop for controlling flowering (Gramzow & Theissen, 2010; Liu & Mara, 2010).

The original homeotic model, the ABC model, was proposed because the initial homeotic mutants in *Arabidopsis* and *Antirrhinum majus* (Bowman *et al.*, 1991; Coen & Meyerowitz, 1991) were of one of three possible classes. In “A” class mutants, the sepals and petals were replaced by carpels and stamens respectively. In “B” class mutants, the petals were replaced by sepals, and the stamens are substituted by carpels. And in “C” class mutants, the stamens and carpels are replaced by petals and sepals respectively. However, after several years of research, other homeotic genes were identified in petunia and in *Arabidopsis*, the “D” class for ovule-specific identity genes (Angenent & Colombo, 1996; Pinyopich *et al.*, 2003; Favaro *et al.*, 2003), and the “E” class for the more redundant flower-specific *SEPALLATA* genes (Pelaz *et al.*, 2000; Ditta *et al.*, 2004). Flower development seems to depend on a five-class homeotic model, the ABCDE model. In this model, the first four classes of floral homeotic transcription factors (ABC and D) interact with a fifth class (E) as tetrameric protein complexes (i.e., the floral quartet model (FQM), (Theißen & Saedler, 2001)). These tetramers induce transcription by binding to the DNA of their target genes. The ABCDE model has been thoroughly studied in *Arabidopsis*, *Antirrhinum*, petunia, and tomato (reviewed in Causier, Schwarz-Sommer, and Davies 2010; Immink, Kaufmann, and Angenent 2010; Ó’Maoiléidigh, Graciet, and Wellmer 2014; Pajoro *et al.* 2014; Rijpkema *et al.* 2010).

The ABCDE model has also recently been simplified to the (A)B(C) model where class “A” is made up of the previous class “A” and class “E” genes, and likewise, class “C” is made up of the previous class “C” and class “D” genes because of gene function redundancy (Theißen *et al.*, 2016).

Most of the FOI, FPI, and FMI genes in *Arabidopsis* are MIKC^C MADS-box transcription factors. Plant MADS box genes are divided into two groups, MIKC^C and MIKC^{*}. MIKC^C genes have a MADS DNA binding domain, an intervening (I) domain, and keratin-like (K) domain, and a C-terminal domain. These highly conserved genes are found in all land plants (Gramzow & Theissen, 2010; Thangavel & Nayar, 2018) and appear to have evolved ~700 MYA (Thangavel & Nayar, 2018).

Genes in the ABCDE flowering model were identified in the most recent common ancestor (MRCA) of all seed plants.(Chen *et al.*, 2017). Both *Populus* (Chen *et al.*, 2018a) and *Eucalyptus* (Vining *et al.*, 2015a) have homologs to the homeotic genes in *Arabidopsis* with comparable function and expression. It seems reasonable to assume that most gymnosperms and angiosperms have homeotic genes that follow a model similar to *Arabidopsis*' ABCDE model.

**2 A tapetal ablation transgene induces stable male-sterility and slows field growth in
*Populus***

Estefania Elorriaga, Richard Meilan, Cathleen Ma, Jeffrey S. Skinner, Elizabeth Etherington, Amy Brunner, and Steven H. Strauss

Tree Genetics & Genomes

Springer Nature
One New York Plaza, Suite 4600
New York
NY 10004-1562, USA
(2015) 11:127

Issue 10 (2014)

Contributions of authors

Estefania Elorriaga analyzed the data and wrote the manuscript. Cathleen Ma supervised and performed plant stable transformation, regeneration, selection, propagation, transplanting, and field planting. Jeffrey S. Skinner assisted with the PCR confirmation of transgenesis. Elizabeth Etherington managed the field operations. Richard Meilan assisted with propagation, field planting, growth measurements, and catkin collections. Amy Brunner and Steven H. Strauss designed the study. Steve H. Strauss supervised the overall study.

Abstract

The field performance of genetic containment technologies—considered important for certain uses of transgenic trees in forestry—are poorly known. We tested the efficiency of a *BARNASE* gene driven by the *TA29* tapetum-dominant promoter for influencing growth rate and inducing male-sterility in a field trial of transgenic hybrid poplar (*Populus tremula* x *P. tremuloides*). When the growth of 18 transgenic insertion events with the sterility transgene were compared to non-transgenic controls after two growing seasons, they grew 40 % more slowly in stem volume, and all but one transgenic event grew significantly more slowly than the control. In contrast, when we compared the growth of transgenic trees containing four kinds of GUS reporter-gene constructs to non-transgenic trees—all of which had been produced using the same transformation method and poplar clone and grown at the same field site—there were no statistically significant differences in growth after three growing seasons. In two years where gross pollen release from catkins was monitored and found to be abundant in the control, no pollen was visible in the transgenic trees; microscopy suggested the cause was tapetal collapse, and revealed the presence of a very few normal sized pollen grains of unknown viability. In two additional years when viable, well-formed pollen was microscopically documented in controls, no pollen could be observed in any transgenic trees. We conclude that this construct resulted in robust and possibly complete male sterility that was stable over four years in the field.

Keywords *Populus*, *BARNASE*, *TA29* promoter, pollen, risk assessment, forest biotechnology, genetic containment, genetic engineering

2.1 Introduction

Genetically engineered (GE) trees in field trials have shown improved wood quality; faster growth; and markedly improved insect, disease, herbicide, and abiotic-stress resistance (Harfouche et al. 2011; Hinchee et al. 2011). However, regulations and substantial market barriers hinder research progress and commercial applications. A major obstacle to application of GE trees is concern over transgene dispersal in the environment (Strauss et al. 2009a, b). Although most pollen from wind-pollinated woody plants falls close to its point of release, a minority can travel from hundreds of meters to several kilometers. For example, paternity analyses done in two northwestern *Populus trichocarpa* populations, one in western Oregon and another in eastern Oregon, revealed that one-third to one-half of the fertilizing pollen originated from beyond 1 km and 10 km, respectively (DiFazio et al. 2012; Slavov et al. 2009). When sexually compatible, pollen can fertilize the abundant wild and feral populations of poplars that are common in many temperate-zone regions (James et al. 1998). A minority of seed—abetted by animal-, water-, or storm-associated dispersal—can also travel large distances and establish in the wild. Due to this potential for wide dispersal, and because the possible ecological effects of novel genes in the wild are difficult to predict with confidence, there has been long-term interest in the development of containment methods to prevent or strongly mitigate transgene dispersal.

There are several major genetic containment strategies that have been discussed for forest trees (reviewed in Brunner et al. 2007, Vining et al. 2012). These include fitness reduction, ablation, transgene excision, and floral gene suppression at the RNA or protein levels. Ablation methods have been most widely studied, and in the case of floral sterility rely on cell- or tissue-dominant promoters to drive expression of a cytotoxin gene to destroy tissues essential for gamete development. Previous studies using the pTA29::BARNASE construct in alfalfa (*Medicago sativa*, Rosellini et al. 2001), oilseed mustard (*Brassica juncea*, Jagannath et al. 2001), oilseed rape (*Brassica napus*, Mariani et al. 1990), tobacco (*Nicotiana tabacum*, Mariani et al. 1990), and wheat (*Triticum aestivum*, De Block et al. 1997) demonstrated this construct is effective in inducing male-

sterility by disrupting the development of tapetal cells. Jagannath et al. (2001) also found that the Arabidopsis tapetal promoter A9 was highly effective; 94 % of the transformants with the tobacco TA29 promoter and 87 % of the transformants with the A9 promoter showed stable male sterility, and none of the plants produced seed by selfing (i.e., reverted to being fertile). Mariani et al. (1990) found that 106 out of 115 transformed tobacco events showed stable male sterility. The 14 transgenic lines of wheat studied by De Block et al. (1997) had one to three copies of the barnase gene and all but one showed stable and complete male sterility that was inherited in offspring of a varietal hybrid. The constructs BpMADS1::BARNASE (Lemmetyinen et al. 2001, 2004) and BpFULL1::BARNASE (Lännenpää et al. 2005) produced male sterility in Arabidopsis, tobacco (*Nicotiana tabacum*) (Lännenpää et al. 2005; Lemmetyinen et al. 2001), and silver birch (*Betula pendula*) (Lännenpää et al. 2005; Lemmetyinen et al. 2004). The employed promoters were derived from the birch BpMADS1 gene (homologous to the Arabidopsis gene SEPALLATA3, previously known as AGL19), and the birch BpFRUITFULL-LIKE1 gene. Zhang et al. (2012) showed that a male cone-dominant promoter from Monterey pine (*Pinus radiata*), PrMC2, fused to an attenuated version of barnase gave complete and stable male sterility in multi-year field trials of a pine hybrid (*Pinus rigida* x *P. taeda*) and a eucalypt hybrid (*Eucalyptus grandis* x *E. urophylla*).

There have been several reports of vegetative abnormalities when the barnase gene was employed for sterility. In the greenhouse studies of silver birch (Lännenpää et al. 2005; Lemmetyinen et al. 2004), transformants were bushy, short, and grew more weakly than the control. Jagannath et al. (2001), studying Brassica, found that constructs with the strong constitutive cauliflower mosaic virus (CaMV) 35S promoter driving the selectable-marker gene showed more vegetative abnormalities than transformants with a weaker promoter driving the selectable-marker gene. They inferred this to be a result of 35S enhancer effects causing vegetative expression of barnase. Thus, it is important to carefully evaluate vegetative growth in transgenic plants containing barnase-based ablation transgenes.

We report that a TA29::BARNASE transgene was highly effective at inducing male sterility in poplar, and did so in the field over several years, but it also caused significant growth retardation. These results show that sterility transgenes can be highly effective at mitigating transgene dispersal in poplar, but that further technology development and testing is required to establish methods for imparting female sterility and reducing undesired effects on growth.

2.2 Materials and Methods

2.2.1 Gene constructs

2.2.1.1 Sterility

Hybrid poplars were transformed with the binary plasmid pTTM8 provided by Plant Genetic Systems (Gent, Belgium). The vector (described in Li et al. 2007) contained three transgenes within its T-DNA, including *BARNASE*; neomycin phosphotransferase II (*NPTII*), which provides resistance to kanamycin; and *bar* (selectable marker for glufosinate ammonium herbicide resistance). The *BARNASE* gene was derived from *Bacillus amyloliquefaciens* (Mariani et al. 1990) and encodes a ribonuclease.

2.2.1.2 Reporter

To evaluate whether transformation affected growth rate, we produced a number of transgenic trees that lacked the *BARNASE* gene, and contained only the β -glucuronidase (GUS) reporter gene and the selectable marker gene for kanamycin resistance. Between 1998 and 1999, 10 transgenic events were produced via the same *Agrobacterium tumefaciens* transformation methods for each of the four GUS transgene constructs: *PTD::GUS* (abbreviated 3PG), *EnACT11::GUS* (3A11G), *EnACT2::GUS* (3A2G), and *En35S::GUS* (3SG) (Supporting Information Table S2.1). The promoter from the *PTD* gene is expressed in a floral predominant manner (Sheppard et al. 2000; Skinner et al. 2003), while the other three promoters (*EnACT11*, *EnACT2*, and *En35S*) allow for approximately constitutive expression (An et al. 1996, Huang et al. 1997). In brief, constructs *En35S::GUS::E9*, *EnACT11::GUS::E9*, and *EnACT2::GUS::E9* were

generated using pMON10547 as the binary vector backbone (Supporting Information Table S2.1). The vector pMON10547 contains two copies of the 35S promoter enhancer region, the 35S promoter basal region, a multicloning site (MCS), and the E9 transcriptional terminator. Construct En35S::GUS::E9 was generated by adding the GUS reporter gene behind the constitutive 35S promoter. Both actin promoters, EnACT11 and EnACT2, were amplified from *Arabidopsis thaliana* ecotype Columbia based on the corresponding actin2 (AT3G18780) and actin11 (AT3G12110) gene sequences, using primers with flanking restriction enzyme adaptor sites. The ACT2 promoter region was amplified using primers ACT2.001 (ESM, Table S2) and ACT2.002 (ESM, Table S2), and the ACT11 promoter region using primers ACT11.001 (ESM, Table S2) and ACT11.002 (ESM, Table S2). Amplicons were subcloned into the PCR cloning vector pCR-TOPO2.1 and confirmed by Sanger sequencing. The cloned promoters were excised and subcloned into the binary vector to replace the 35S basal promoter region (i.e. the -90 fragment of Benfey and Chua 1990) while retaining the upstream double 35S promoter enhancer regions to generate constructs EnACT2::GUS::E9 and EnACT11::GUS::E9. The tandem 35S promoter enhancer regions were thus fused upstream of the respective actin promoter to augment the endogenous constitutive actin regulatory and basal promoter regions. Sanger sequencing was performed to confirm the correct assembly of the respective promoter GUS reporter gene fusion constructs EnACT2::GUS::E9 and EnACT11::GUS::E9.

2.2.2 Micropropagation

Sterile *in vitro* young micro-cuttings (shoot tips and micro-nodes) of male hybrid aspen genotype (INRA 353-38, *P. tremula* × *P. tremuloides*) were propagated on hormone-free, half-strength Murashige and Skoog medium (MS). Plants grew on these media for 4-6 weeks and then micropropagation was repeated three to five times until enough plants for field tests were produced. They were then transplanted to soil and acclimated in the greenhouse and field prior to planting in field trials.

2.2.3 Transformation

Sterile *in vitro* cultures of genotype 353-38 were used for all transformations. Internodes and leaf discs, in admixture, were used as explants for cocultivation. All transformation and regeneration was essentially as described in Filichkin et al. (2006), using *Agrobacterium tumefaciens* strain C58/pMP90 (GV3101). For the sterility construct, 18 independent transgenic events were generated and each transgenic event was micropropagated to produce three ramets for field testing (except for event 2, which had four). There were nine non-transgenic wild-type controls propagated in the same way. For the reporter gene field study, 10 events were produced using each construct and multiplied and readied for the field essentially the same as for the sterility field study.

2.2.4 Transgene confirmation

Genomic DNA was isolated from leaf tissue using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA) following the manufacturer's instructions. All transgenic plants were confirmed by polymerase chain reaction (PCR) against one or more target genes within the T-DNA. We used non-transgenic, *in vitro* grown trees produced at the same time as negative PCR controls. The *NPTII* and *BARNASE* genes were amplified from the male-sterile trees (Supporting Information Table S2.2) and the GUS gene was amplified in the reporter trees (Supporting Information Table S2.2); negative controls, using template from a non-transgenic control, were also included during all PCR analyses. For simplicity, genotype designations have been recorded in this manuscript as follows (format is: "original code/manuscript code"): 7/1, 43/2, 15/3, 102/4, 2/5, P/6, 58/7, 29/8, 39/9, 8/10, 38/11, 28/12, 16/13, 40/14, 9/15, 30/16, 41/17, and 63/18.

2.2.5 Field layout

The reporter and sterility field trials were planted adjacent to one another in a field site near the Willamette River in Benton County, OR. The sterility trial was planted in September 1995 and the reporter trial in September 2000; both trials were irrigated in the first and second growing season after planting.

The sterility trial was planted in a randomized block design with three blocks; each block contained a single ramet of each transgenic event (except for block 1 which had two ramets for event 2), and each block had three ramets for each control type. The reporter trial was planted in a completely randomized design with 10 transgenic events for each construct (except 3PG, which had 3 ramets) and two wild-type controls for each construct (except for the 3PG construct, which had 4), for a total 10 control trees. The spacing was 3 m between rows and 1.5 m within rows. As described above, for the reporter trial all transgenic events started with two ramets (except for event 3PG96, which had three). However, by 2003, one event each in 3A2G, 3PG, and 3SG, and one control ramet (out of the 10), was lost due to animal damage or unknown causes. We conducted a one-way ANOVA with constructs (including control as a “construct”) as the main effect to test if mean volume index varied among constructs for the years 2001 and 2003. The volume index of the different construct-associated control groups was not significantly different (Supporting Information Tables S2.3 and S2.4), so all controls were pooled for analysis.

2.2.6 Tree growth measurements

Tree height and/or stem diameter at 1.5 m above ground were measured for both trials. Where height and diameter were both available, they were used to calculate a volume index ($\text{height} \times \text{diameter}^2$) for each tree. Volume index data for the sterility trial is presented after two growing seasons, approximately at the onset of visible inter-tree competition. Data for the reporter trial is presented for growth after one and three growing seasons; the latter was at the onset of inter-tree competition.

2.2.7 Catkin collection and measurements

Flowering was studied only in the sterility trial, where catkins were collected every year from 2006 to 2009 starting just prior to the onset of pollen release, and then additional catkins were collected for approximately one month or until no new catkins could be seen opening on trees. Prior to the onset of flowering, two out of the three original blocks had to be removed due to other demands for that land. Catkins were placed in coolers with ice immediately after collection. Because the branches bearing catkins were

approximately 10 to 30 m above the ground, pole pruners and hydraulic lifts were used to facilitate sampling (Fig. 2.1). During the collection period in 2009, we also took data on catkin morphology (length, weight, and curvature). To quantify a visible difference in catkin curvature, we measured six visibly representative catkins, one from the control and one from five different transgenic events, during the March 17th collection, using ImageJ freeware (Schneider et al. 2012). The angle made by a line projected from the catkin base with that from the catkin tip was used to approximate curvature.

2.2.8 Pollen dehiscence and assessment

After catkin length and weight were measured, they were placed in a refrigerator until they completed development and began to release pollen. Full pollen release was observed after transfer to Petri dishes kept at room temperature (~20 °C). Catkins were shaken to aid in release of pollen, and the pollen then diluted in water for microscopic quantification. The numbers of well-formed (i.e., normal and uniform size and shape) grains were counted under a dissecting microscope with a hemocytometer (Supporting Information Figs. S2.1 and S2.2). For each sample, seven hemocytometer squares were counted. To estimate pollen viability in 2008 and 2009, we used 2,2,5-triphenyl tetrazolium chloride (TTC) (Cook and Stanley 1960); only control pollen was stained because the transgenic trees did not produce any detectable pollen. Photos of catkins in Petri dishes were taken with an Axio camera model AxioCam ICc 1 and a ZEISS Stemi SV11 Apo stereomicroscope, and the photos were examined using Carl Zeiss Vision/Axiovision software (release 4.8.1, November 2009). Images of catkins from transgenic events and controls releasing pollen in Petri dishes were taken for all the years of study (2006-2009). Images of some catkins were previously published in review papers by Brunner et al. (2007) and Vining et al. (2012), and thus not shown here.



Fig. 2.1 Field trial during early growth and catkin collection. a, Photographed in July 1997, two growing seasons after planting. Note the person (~1.8 m) just visible within the trees (center bottom). b, the trees in November 1998, after the third growing season after planting. c, collecting catkins with a lift while using a pole pruner during March 2009. d, collecting catkins by hand with a Swedish ladder and a pole pruner in February 2007.

2.2.9 Anther microscopy

To observe anther structure, catkins were placed in a formalin–acetic acid–alcohol (FAA) fixative solution and vacuum infiltrated for 1–2 h, then stored at 4 °C in the dark. For the histological images taken for anthers collected in 2006 and 2009, samples were fixed, dehydrated, embedded in glycol GMA methacrylate plastic, sectioned, and mounted on slides. Sections were stained in 0.5 % Toluidine Blue O in citrate buffer. Photographs were taken using a DFC 290 camera with a Leica DM5000 microscope at both 10X and 25X magnifications, and analyzed using the Leica Application Suite software on the

camera. Histology images from 2006 were published previously and not presented here (Brunner et al. 2007; Dalton et al. 2013).

2.2.10 Quantitative analyses

Statistical analyses were conducted using the R statistical computer language and environment version 3.0.1 (R Core Team, 2013) using the R packages lsmeans (Lenth 2013), multcomp (Hothorn et al. 2008), and nlme (Pinheiro et al. 2013). Because of heteroscedasticity observed after inspection of residuals, all models were fit using generalized least squares (Hothorn et al. 2008; Lenth 2013; Pinheiro et al. 2012) with variances allowed to differ among constructs or events. Graphical checks of the residuals then showed that variances were approximately uniform and normal in distribution. For the sterility trial, we conducted a two-way ANOVA with blocks and events (including controls) as main effects to test if mean volume index varied significantly. Following ANOVA, a Dunnett's multiple comparison test was used to compare the mean volume index of each transgenic event against the control. For the reporter trial, we carried out a one-way ANOVA to test if mean volume index varied among constructs or controls for the years 2001 and 2003 separately. We also conducted a one-way ANOVA for each construct to test if mean volume index varied among events within a construct. Following ANOVA, a Dunnett's multiple-comparison test was used to compare the mean volume index of lines from each construct against the control.

In 2009, catkins were collected from events 7, 9, 12, 14, 17, and control trees on March 10th and March 17th to measure their length and weight (no catkins were collected from event 17 on March 10th). During the first collection, the number of catkins collected were 47, 37, 54, 36, and 60, corresponding to events 7, 9, 12, 14, and control, respectively. On the second collection, the number collected were 6, 31, 14, 13, 9, and 35, corresponding to events 7, 9, 12, 14, 17, and control, respectively. Mean length was calculated by averaging the length of all the catkins from a single event or control. A Dunnett's multiple-comparison test was used to compare the mean catkin length of each transgenic event against that of the control for each collection date. Mean weight was determined by averaging the total weight of all the catkins belonging to an event by the number of

catkins. A single-sample Student's t-test was used to compare the mean catkin weight of each transgenic event against that of the control for each collection date.

2.3 Results

2.3.1 Slowed early growth of transformants with a BARNASE transgene

Statistical analysis of early growth in the sterility trial showed that blocks and events were significant sources of variance ($F_{2,46} = 52$, $p < 0.0001$; Supporting Information Table S2.5). All of the transgenic events showed lower mean volume indexes than the control (Fig. 2.2), and all but one of the 18 events was statistically different (smaller) than the control based on Dunnett's test (Fig. 2.2; Supporting Information Table S2.6). On average, the sterility transgenic events cumulatively grew 40 % more slowly than the control trees over the two seasons of the growth trial.

2.3.2 Absence of evidence for growth impairment due to transformation

Because expression of the GUS reporter gene in transgenic plants is widely known to be essentially free of pleiotropic effects on plant phenotypes (e.g., Gilissen et al. 1998), we employed a population of these plants for helping to assess whether the growth loss we observed in our sterility trial could have been due to transformation itself. In the reporter trial, the mean volume indexes of the transgenic constructs were not significantly different from each other or the controls in 2001 (one year of growth) ($F_{4,36} = 0.212$, $p = 0.930$; Fig. 2.3; Supporting Information Table S2.7) or in 2003 (three years of growth) ($F_{4,36} = 0.263$, $p = 0.900$; Supporting Information Table S2.8). The lack of differences among constructs was also revealed by Dunnett's test for both 2001 and 2003 (Supporting Information Tables S2.9 and S2.10). Nonetheless, events within construct

type were a statistically significant source of variance (ESM, Tables S11 and S12), likely due to an unusually fast- or slow-growing event (data not shown).

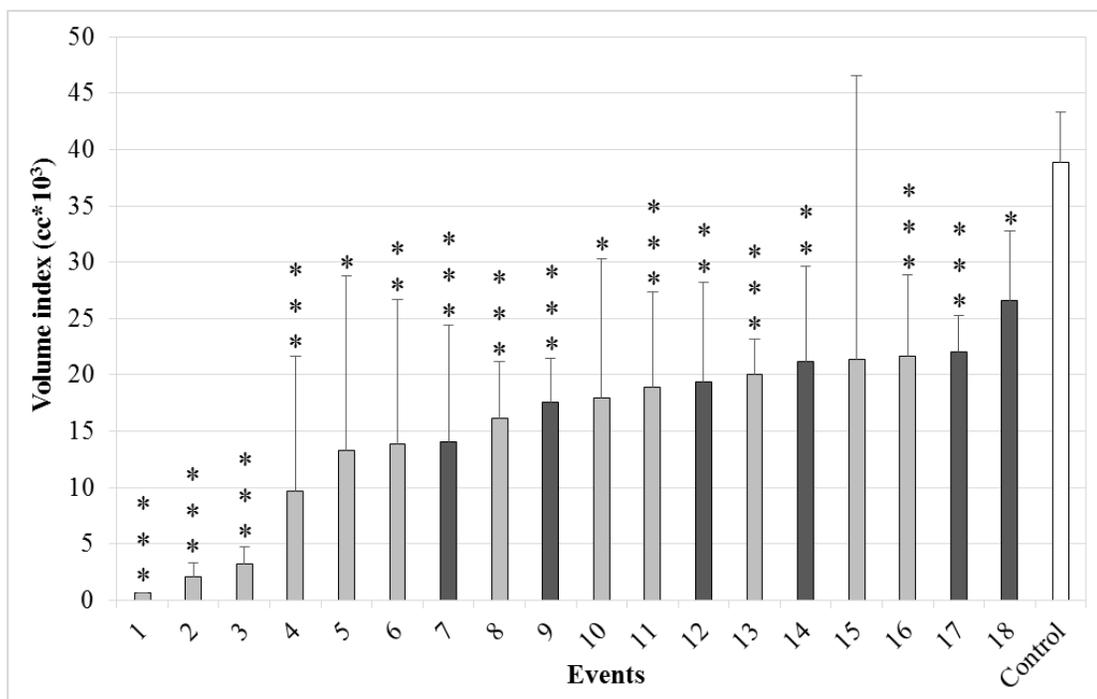


Fig. 2.2 Transgenic trees showed reduced growth when compared to non-transgenic control. The heights and diameters of all the trees were measured in fall 1997, two growing seasons after planting. Each bar identifies an individual gene insertion event or control. The brackets represent 95 % confidence intervals. The asterisks indicate whether the volume index of the specific event was significantly different than the control based on a Dunnett's test (three asterisks: $P < 0.001$, two asterisks: $P < 0.01$, and one asterisk: $P < 0.05$; all rounded up). The events with darker bars were also studied for their pollen sterility; cc, cubic centimeters.

2.3.3 Transgenic catkins showed distinct size and morphology

Measurements of catkin length from two collections in 2009 showed that transgenic catkins were often significantly shorter than control catkins (Supporting Information Fig. S2.3; Tables S2.13 and S2.14). Measurements of weight on the same catkins showed that, for the first collection, weight was similar between transgenic and control catkins ($p = 0.116$); however, for the second collection, when catkins were more fully mature, the weight of the control catkins was significantly lower than that of the transgenic catkins ($p = 0.009$) (Supporting Information Fig. S2.4). When we measured the angle projected

from the two tips of the catkin toward one another, the transgenic catkins were distinctly curved ($91^{\circ} \pm 10.3$), whereas control catkins all appeared perfectly straight (180°) (Supporting Information Fig. S2.5; Table S2.15).

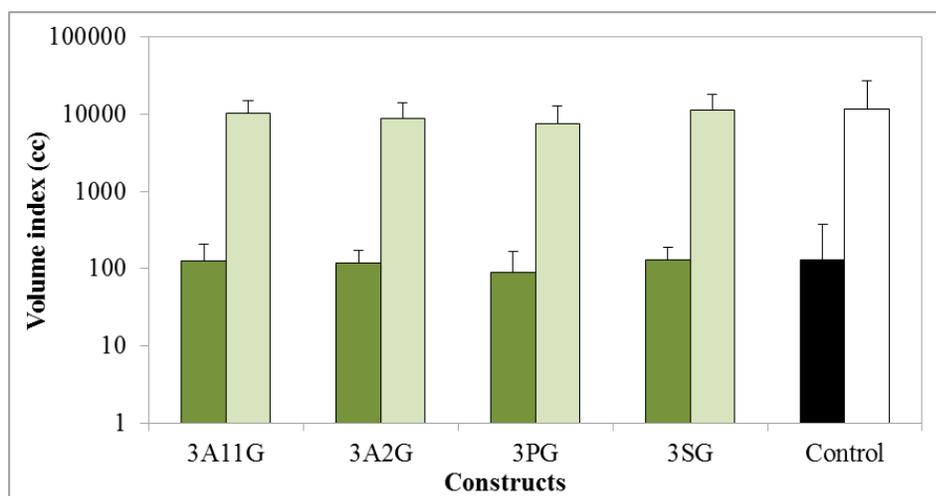


Fig. 2.3 Reporter and non-transgenic trees grew at similar rate. The volume index of each transgenic construct was not significantly different from the controls (all $P > 0.60$, see Supporting Information Tables S2.7 and S2.8). Brackets represent 95 % confidence intervals. Darker bars show data from 2001 and lighter bars show data from 2003; cc, cubic centimeters.

2.3.4 Absence of pollen during visual inspection of transgenic catkins

Based on visual inspection of whole Petri dishes after manual agitation, we found that none of the transgenic trees released significant amounts of pollen during any of the years of study (Fig. 2.4). After microscopic inspection, control trees released an average of 73,000 and 85,000 pollen grains per catkin in years 2006 and 2007, respectively, whereas only a few possible pollen grains (based on similarity in size and shape to wild type pollen) were observed from transgenic trees in 2006 (Brunner et al. 2007) and 2007 (Supporting Information Fig. S2.1). In 2008 and 2009, no pollen, viable or otherwise, was macro- or microscopically detected from the transgenic trees, though it was abundantly produced by the control trees (Supporting Information Fig. S2.2). For the years 2008 and 2009, the control trees released approximately 775,000 and 2,700,000 pollen grains per catkin, respectively. The abundant pollen that was produced on the

control catkins were also highly visible when whole catkins with mature anthers were examined; the control catkins had swollen anthers, but the anthers of transgenics were shrunken (Fig. 2.5).

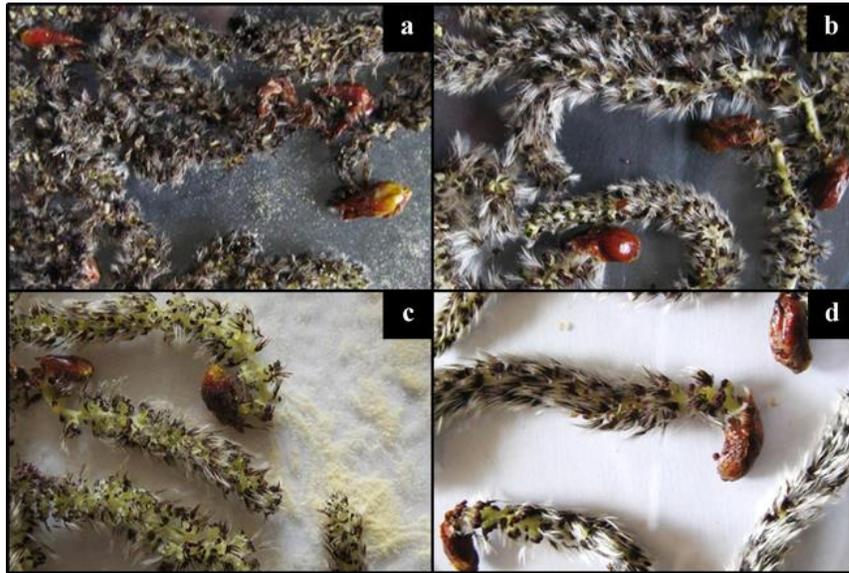


Fig. 2.4 Absence of visible pollen release from transgenic catkins. a and c, show control catkins with released pollen. b and d, show transgenic catkins without visible released pollen. b, catkins correspond to event 17. d, catkins correspond to event 14. a and b, catkins were collected in 2007. c and d, catkins were collected in 2009. Images of catkins from 2006 and 2008 were previously published in Brunner et al. (2007) and Vining et al. (2012).

2.3.5 Transgenic anthers showed a collapsed tapetum and absence of pollen

In 2009, we selected catkins from one of the sterile lines (event 12) to study in further detail the structure and morphology of the anther sacs, including the tapetum.

Microscopic analyses showed that transgenic anthers had significant developmental abnormalities compared to control anthers (Fig. 2.6). The tapetum of the transgenic anthers appears to have collapsed and no pollen grains were observed inside the pollen sacs. It was difficult to differentiate the tapetum from the endothecium, but it appears that the tapetum was completely ablated and the endothecium appeared thicker than in wildtype.

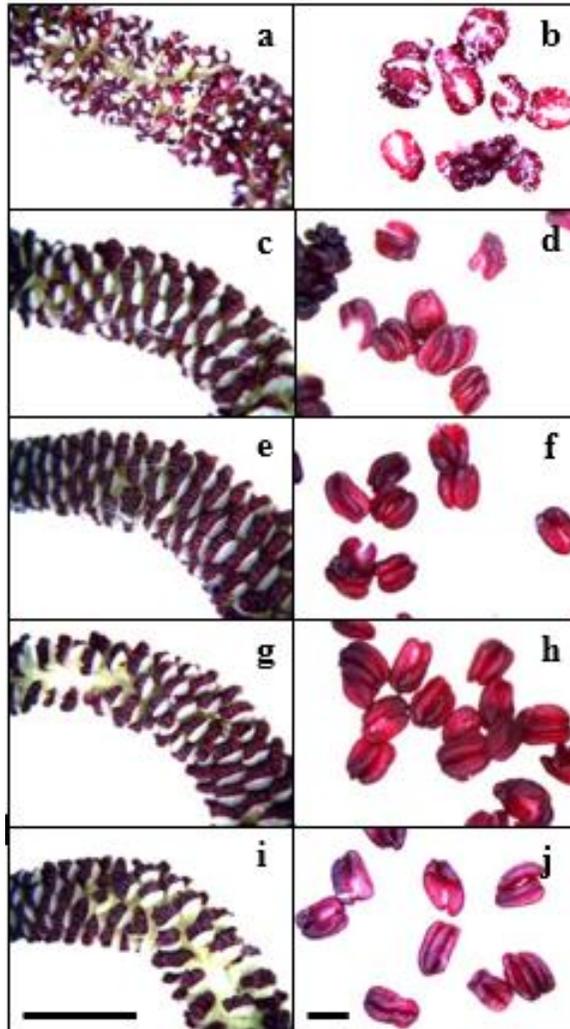


Fig. 2.5 Transgenic catkins lacked visible pollen. a and b, catkin and stamens are from a control tree. c and d, catkin and stamens are from transgenic event 12. e and f, event 9. g and h, event 14. i and j, transgenic event 7. The black bar in the catkin image i corresponds to 1 cm and the black bar in the stamen image j corresponds to 1 mm. Photos were taken during March 2009. Histology photos from 2006 can be found in Brunner et al. 2007 and Dalton et al. 2013.

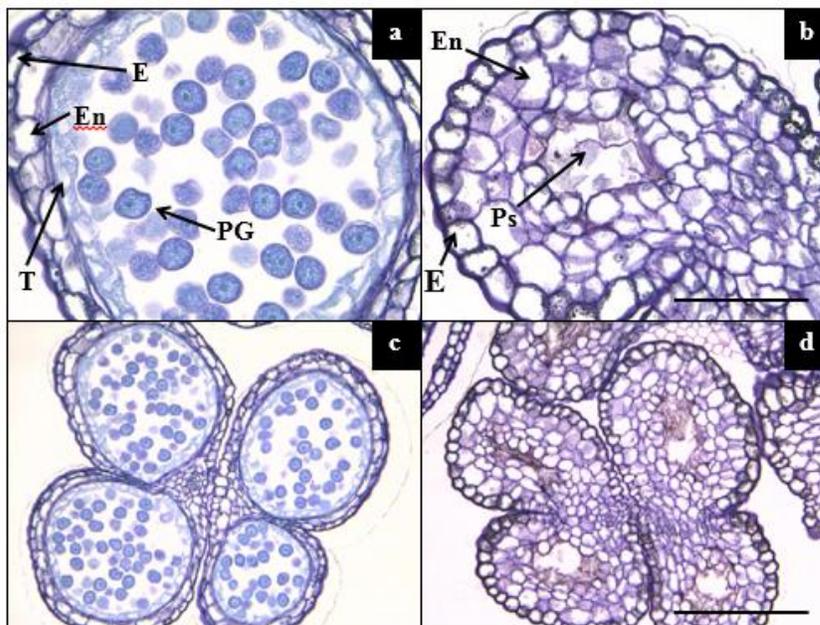


Fig. 2.6 Transgenic catkins lacked visible pollen. a and b, catkin and stamens are from a control tree. c and d, catkin and stamens are from transgenic event 12. e and f, event 9. g and h, event 14. i and j, transgenic event 7. The black bar in the catkin image i corresponds to 1 cm and the black bar in the stamen image j corresponds to 1 mm. Photos were taken during March 2009. Histology photos from 2006 can be found in Brunner et al. 2007 and Dalton et al. 2013.

2.4 Discussion

Based on macroscopic inspection of dehiscing catkins, all of the flowering transgenic trees were male-sterile in all four years of study. In the two years where pollen viability was also determined and analyzed microscopically, only the control produced viable pollen; no pollen was detected from the transgenics. In the review paper by Brunner et al. (2007), preliminary data based on hemocytometer counts were presented that showed extremely little pollen production in 2006; of six events studied, two produced no detectable pollen, and the other four events produced a mean of six pollen grains per catkin, consistent with the very rare grains we observed in 2007 (Supporting Information Fig. S2.2). However, viability was not determined in either 2006 or 2007. Thus, it remains unclear if transgenic trees in this study could produce any viable pollen.

Our results showed strong and consistent, but possibly not absolute, male sterility. This finding is similar to several other studies. Lemmetyinen et al. (2004) saw no pollen in transgenic birch. Similarly, no pollen was observed on any transformed tobacco plants (Mariani et al. 1990) nor on transgenic wheat (De Block et al. 1997). Moreover, none of the wheat produced seed from selfing; seed was only produced when cross-pollination was performed. Studying *Brassica*, Jagannath et al. (2001) found that all of the transformed lines lacked pollen, and none of the plants produced seed by selfing or reverted to being fertile. However, Jagannath et al. (2001) also produced semi-sterile plants and they noted that many of them had both sterile and fertile anthers, the former of which eventually reverted to being fully fertile. Rosellini et al. (2001) noted “traces” of pollen in three out of five transformed alfalfa lines, while two lines showed no pollen at all. The three transgenic pollinating lines were selfed and produced just a few seeds; however, the control produced greater than six-fold more seed than the transgenic lines.

In addition to pollen sterility, we found alterations in catkin morphology in the transgenic trees. The catkins from the transformants were smaller and curved when compared to the control catkins (Supporting Information Figs. S2.3 and S2.5). In one collection, we were surprised to find that the transgenic catkins were heavier than the control catkins (Supporting Information Fig. S2.4). We hypothesize that this is because the control catkins were weighed after most of their pollen had been released. The control catkins were significantly longer than the transgenic catkins (Supporting Information Fig. S2.3); it is unlikely that pre-dehiscence catkins would also be lighter. Male-sterile *Brassica* transformants also had smaller flowers than controls (Jagannath et al. 2001).

The pollen sacs in our transformants were collapsed, with the tapetum and pollen grains absent. Similar results were reported by Mariani et al. (1990), De Block et al. (1997), and Rosselini et al. (2001) in tobacco, wheat, and alfalfa, respectively. In transgenic alfalfa deterioration of the tapetum was evident during premiotic development (Rosselini et al. 2001).

Because of its potent and nearly indiscriminant degradation of cellular RNA, barnase is toxic in both prokaryotes and eukaryotes (Ulyanova et al. 2011). All of the transgenic

trees we studied had inferior growth to the control trees. This growth difference had also been briefly reported earlier in a preliminary report from our laboratory (Skinner et al. 2000). The GUS enzyme is one of the most widely used reporter genes in transgenic plants, partly because it is not toxic to transformed cells (Gilissen et al. 1998, Miki and McHugh 2004). Transgenic GUS plants have shown no growth effects (Gilissen et al. 1998) or pleiotropic changes in gene expression (Ouakfaoui and Miki 2005) when compared to non-transgenic control plants. Moreover, in a randomized greenhouse study Lemmetyinen et al. (2001) saw no significant difference in growth between their non-transgenic control line and a transgenic line containing *BpMADS5::GUS*.

Based on the lack of evidence for growth effects of GUS reporter constructs in the literature, and the current results where GUS transgenic trees did not differ from non-transgenic trees in volumetric growth, the data implicate the barnase cytotoxin transgene as the probable cause of the slowed early tree growth seen in the male-sterile transgenic poplars. Reduced growth associated with barnase expression was also seen in greenhouse studies of silver birch (Länneppää et al. 2005; Lemmetyinen et al. 2004). In the study with the *BpFULL1* promoter, one third of their 12 non-flowering lines showed severely reduced growth and small narrow leaves (Länneppää et al. 2005), and in the study with the *BpMADS1* promoter 38 of the 45 non-flowering transformants studied were weaker, shorter, and showed abnormal dichotomous branching. Vegetative impairment was seen in all *Brassica* transformants (Jagannath et al. 2001) where the CaMV 35S promoter drove expression of the *bar* selectable marker gene in a construct containing the TA29 promoter driving expression of the *barnase* gene. The authors attributed the morphological abnormalities in their transformants to the unintended expression of barnase in vegetative tissues associated with 35S enhancers, as has been reported in other studies (Yoo et al. 2005). Interestingly, in the Jagannath et al. (2001) study no morphological abnormalities were seen in transformants with a 5-kb spacer between the 35S promoter and the *barnase* gene driven by the TA29 promoter. Our barnase construct, however, did not include the 35S promoter. Instead, the nopaline synthase gene promoter (NOSp) drove the *NPTII* gene and the promoter from the ribulose-1,5-bisphosphate carboxylase small subunit (*rbcS*) gene from *Arabidopsis*,

atS1A, drove the *bar* gene. Both of these promoters are generally considered to be expressed in numerous tissue types, however, the NOS promoter has shown significantly less expression (at least 30-fold) of marker genes compared to the 35S promoter in both petunia and moss (Horstmann et al. 2004; Sander et al. 1987). The alfalfa *rbcS* promoter also showed much weaker activity than the 35S promoter in young leaves, old leaves, shoot tips, and nodules of alfalfa (Samac et al. 2004). Nonetheless, we believe that the NOSp, together with the effects of “random” transgene integration, were a likely cause of unintended barnase expression in vegetative tissues, and thus growth inhibition. It is also likely that the TA29 promoter has imperfect tissue fidelity, especially in the taxonomically distant dicot *Populus* (it was isolated from tobacco).

The deleterious effects of barnase observed in this study may have been exacerbated by growth in a field environment. No adverse effects on growth were observed in the greenhouse studies of oilseed rape (Mariani et al. 1990), tobacco (Mariani et al. 1990), or wheat (De Block et al. 1997). These plants showed normal height, leaf size, tillering, and/or high vigor based on causal observation. They also found no abnormalities in any of the flowering organs except for the tapetum. In the randomized greenhouse study on *Arabidopsis* and tobacco by Lemmetyinen et al. (2001), there were no differences in growth of transgenic vs. control lines until flowering began. Likewise, Wei et al. (2006) reported normal growth and morphology of greenhouse-grown trees expressing *barnase* under the poplar *LEAFY* promoter, whereas their field-grown transgenic trees had highly abnormal morphology and reduced growth. Surprisingly, this occurred in spite of co-expression of the barnase inhibitor *barstar* (Wei et al. 2006). In contrast, Zhang et al. (2012) did not observe any growth impairment in their field-grown, male-sterile pines and eucalypts, possibly because of the reduced toxicity of the barnase variant employed. They did not, however, present any data or statistical analysis in support of this observation. These results suggest that barnase toxicity can vary widely depending on species and growth environment, and the need for steps to reduce barnase toxicity due to mis-expression. These could include the use of spacers (Jagannath et al. 2001), separation of subunits among plants following crossing (Burgess et al. 2002; Bihao et al. 2012), or attenuated versions of the protein (Zhang et al. 2012).

The male sterility we observed was expressed over four years in the field. Similarly, Zhang et al. (2012) reported complete male sterility over four years in pine and over two years in eucalypts. Tobacco and silver birch transformed with the BpMADS1::*BARNASE* and the BpFULL1::*BARNASE* constructs were highly sterile for three and two consecutive years, respectively (Lännenpää et al. 2005). Commercial male-sterile *Brassica* has been authorized for use since 1996 in the USA and Canada (CERA 2013). Thus, it appears that barnase expression can be a highly reliable means for generating male sterility.

There remain a number of research needs if barnase technology is to become a general tool for transgene containment in forest trees. These include statistically robust and long-term field tests evaluating the level of sterility and impacts on vegetative growth, and the development of new promoters that are more specific in their expression patterns and/or employ less toxic forms of cytotoxin genes. Female sterility will also be needed for many forest trees due to animal, wind, and water dispersal of seeds. This is particularly true for poplars, which have seeds that are very small and can “float” in air and on water over large distances. To our knowledge, no genes that cause bisexual or female sterility appear to have been field-tested in any plant species.

Field tests are essential for determining the efficacy and stability of transgenic sterility. Unfortunately, the stringent regulation of all forms of direct genetic modification that are in place around the world make even small field trials very difficult, and in many cases impossible, to carry out (Viswanath et al. 2012). The development of robust containment technology would therefore benefit not only from additional laboratory research, but from more discriminating regulatory systems that are based on trait risks and benefits, not on a presumed hazard due to use of recombinant DNA methods (Strauss et al. 2010; Meilan et al. 2012).

Acknowledgements

We thank Plant Genetic Systems (Belgium) for providing the barnase construct for testing, Monsanto for providing the vector backbone for the reporter constructs, Jim Etherington for help in catkin harvesting, and Kathy Cook (Oregon State University) for conducting the anther histological studies.

Data archiving statement

Data used in this manuscript will be made publicly available through DRYAD (<http://datadryad.org/depositing>).

3 Variation in mutation spectra among CRISPR Cas9 mutagenized poplars

Estefania Elorriaga, Amy L. Klocko, Cathleen Ma, and Steven H. Strauss

Frontiers Media SA

Avenue du Tribunal Fédéral 34

1005 Lausanne

Switzerland

Tel +41 (0)21 510 17 00

Fax +41 (0)21 510 17 01

Volume 9 (2018)

Contributions of authors

Estefania Elorriaga designed the study with the help of Amy L. Klocko and Steven H. Strauss. Cathleen Ma performed the plant transformation, regeneration, and selection. Estefania Elorriaga designed the molecular constructs, sequenced the target sites, analyzed the data, and wrote the manuscript. Steven H. Strauss supervised the study with the help of Amy L. Klocko.

Abstract

In an effort to produce reliably contained transgenic trees, we used the CRISPR Cas9 system to alter three genes expected to be required for normal flowering in poplar (genus *Populus*). We designed synthetic guide RNAs (sgRNAs) to target the poplar homolog of the floral meristem identity gene, *LEAFY* (*LFY*), and the two poplar orthologs of the floral organ identity gene *AGAMOUS* (*AG*). We generated 557 transgenic events with sgRNA(s) and the Cas9 transgene and 49 events with Cas9 but no sgRNA, and analyzed all events by Sanger sequencing of both alleles. Out of the 684 amplicons from events with sgRNAs, 474 had mutations in both alleles (77.5%). We sequenced both *AG* paralogs for 71 events in INRA clone 717-1B4 and 22 events in INRA clone 353-53, and found that 67 (94.4%) and 21 (95.5%) were double locus knockouts. Due partly to a single nucleotide polymorphism (SNP) present in the target region, one sgRNA targeting the *AG* paralogs was found to be completely inactive by itself (0%) but showed some activity in generating deletions when used in a construct with a second sgRNA (10.3% to 24.5%). Small insertion/deletion (indel) mutations were prevalent among mutated alleles of events with only one sgRNA (ranging from 94.3% to 99.1%), while large deletions were prevalent among alleles with two active sgRNAs (mean proportion of mutated alleles was 22.6% for small indels vs. 77.4% for large indels). For both *LFY* and *AG*, each individual sgRNA-gene combination had a unique mutation spectrum ($p < 0.001$). An *AG*-sgRNA construct with two sgRNAs had similar mutation spectra among two poplar clones ($p > 0.05$), however, a *LFY*-sgRNA construct with a single sgRNA gave significantly different mutation spectra among the same two clones ($p < 0.001$). The 49 empty vector control events had no mutations in either allele, and 310 potential “off-target” sequences also had no mutations in 58 transgenic events studied. CRISPR Cas9 is a very powerful and precise system for generating loss-of-function mutations in poplars, and should be effective for generating reliably infertile trees that may promote regulatory, market, or public acceptance of genetic engineering technology.

Keywords *Populus*, CRISPR Cas9, site-directed-mutagenesis, *LEAFY*, *AGAMOUS*

3.1 Introduction

Demand for forest products is expected to increase considerably with the projected population growth in the next few decades (FAO *et al.*, 2012). We harvest forest products from wild and cultivated forests, yet clearing of wild forests comes at a high cost to natural ecosystems (Gamfeldt *et al.*, 2013; Pimm *et al.*, 2014). Meanwhile, plantation forests provide more timber per area than natural forests and provide some of the same ecosystem services as wild forests (Brockerhoff *et al.*, 2008). Plantation forests only comprise 5% of the forested land but they provide about 35% of the world's forest products (FAO, 2010). Based on numerous field studies, it appears that wood yield from intensively grown plantation forests could be improved by the use of genetic engineering (GE) techniques (Strauss *et al.* 2017), and may be particularly important given the rapid growth of biotic and abiotic stresses on forests (Strauss *et al.* 2015). GE may thus lessen the effects that human demand is causing to wild forests and their ecosystems (Strauss *et al.*, 2017). Unfortunately, regulatory and market obstacles greatly limit the ability to use GE methods, even for field research, in many parts of the world, and concerns over gene flow and resulting adventitious presence are major reasons for these obstacles. A reliable genetic containment system might be a key, enabling tool for many applications.

Site-directed mutagenesis has not been readily available in vascular plants, as in other organisms including yeast, *Drosophila*, mouse and human cells, until the advent of site specific nucleases (Weinthal *et al.*, 2010; Voytas, 2013; Chen & Gao, 2014). The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas gene editing system is revolutionizing reverse genetics studies in all systems including trees (Belhaj *et al.*, 2015; Quétier, 2016; Montenegro, 2016; Song *et al.*, 2016). It has made site-directed mutagenesis attractive and attainable in plants because of its relatively low cost, ease of use compared to other methods such as ZFNs and TALENs, and its high mutagenesis efficiency (Samanta *et al.*, 2016; Demirci *et al.*, 2017), including in poplar (*Populus* species) (Fan *et al.*, 2015; Zhou *et al.*, 2015). It should therefore enable the directed mutation of genes essential for sexual fertility—many of which are known from studies in *Arabidopsis* and other model plant species—potentially enabling the production of

predictably and reliable sterile trees (reviewed in Brunner et al., 2007; Vining et al., 2012). Because intensively grown plantation forest trees such as poplar are often vegetatively propagated, and seed as well as pollen dispersal are of concern in most tree species, we chose two types of gene targets whose loss of function is expected to give bisexual sterility.

We targeted the poplar homologs of two genes essential to flower formation and morphology, *LEAFY* (*LFY*) and *AGAMOUS* (*AG*). Flowers form on the edge of shoot apical meristems (SAMs) because of the action of the meristem identity genes *LFY*, *APETALA 1* (*AP1*), and *CAULIFLOWER* (*CAL*) (Parcy, 2005; Diggle et al., 2011). *LFY* encodes a transcription factor that regulates the expression of floral organ identity genes. The precise spatial and temporal expression of the floral organ identity genes determines the generation of the flower and is largely explained by the ABCDE model (previously known as the ABC model) (Coen & Meyerowitz, 1991; Mendoza et al., 1999; Rijpkema et al., 2010). *AG* is a class C gene that encodes a MADS box transcription factor essential for stamen, carpel, and ovule formation (Theissen et al., 2000; Krizek & Fletcher, 2005). Strong homozygous *LFY* mutants in *Arabidopsis* are completely male sterile, and their female fertility is significantly reduced (Schultz & Haughn, 1991; Weigel et al., 1992). Homozygous *FLORICULA* (ortholog of *LFY*) mutants in snapdragon and homozygous *FALSIFLORA* (ortholog of *LFY*) mutants in tomato show complete sexual sterility (Coen et al., 1990a; Molinero-Rosales et al., 1999). The *LFY* homolog in poplar, *PLFY*, is a single copy gene that shows strong expression in developing inflorescences and weak expression in vegetative tissues (Rottmann et al. 2000). Targeting of poplar *LFY* by RNA interference (RNAi) led to female trees with completely sterile flowers and apparently normal growth in the field (Klocko et al., 2016c).

Homozygous *AG* mutants in *Arabidopsis* completely lose their third and fourth whorl identities, and also lose determinacy of the floral meristem (Bowman et al., 1989a). Due to a relatively recent partial genome duplication, there are two *AG* orthologs in poplar, *PAG1* and *PAG2*, located on two different chromosomes (Brunner et al. 2000). They both have a similar expression pattern to that of *AG* in *Arabidopsis* and they share 89% amino

acid identity with each other. Strong RNA suppression of both *AG* genes and *AG-like11* leads to healthy trees with completely sterile flowers in a field trial (Lu *et al.*, 2018b).

We designed four sgRNAs to test the mutagenesis efficiency of the CRISPR Cas9 nuclease system by targeting the poplar orthologs to *LFY* and *AG*. We created six plant-expression plasmids; four expressing the sgRNAs individually and two expressing them in pairs, and transformed them along with a Cas9-only control vector. We were successful at generating hundreds of transgenic events with altered gene sequence. We report that the CRISPR Cas9 system is highly efficient in generating floral gene knock-outs in poplar, and can be readily used to generate large as well as small deletions that should stably destroy protein function.

3.2 Materials and methods

3.2.1 Plant Materials

Leaf, stem, and petiole explants from *in vitro* grown hybrid poplar, INRA clone 717-1B4 (female, *Populus tremula* × *P. alba*; hereafter 717) and INRA 353-38 (male, *P. tremula* × *P. tremuloides*; hereafter 353), which have been grown in our lab for numerous transgenic studies (e.g., Strauss *et al.* 2004; Zhang *et al.* 2010b), were used for *Agrobacterium*-mediated plant transformation. Both clones, abbreviated as 717 and 353, were re-established from field grown material into sterile culture in 2012.

3.2.2 Target Gene Sequencing

Partial sequencing of the *LFY* ortholog, *PLFY* (GenBank accession number U93196, Potri.015G106900), and two *AG* paralogs, *PAG1* and *PAG2* (GenBank accession numbers AF052570 and AF052571, Potri.004G064300 and Potri.011G075800) (Rottmann *et al.*, 2000; Brunner *et al.*, 2000), in 717 and 353 was done previously (Lu *et al.*, 2016). For this study, further sequencing of all genes was done to find natural allelic variants outside of the target region (gene sequence between both target sites) to certify that both alleles for each gene were amplified by PCR (Supporting Information Table S3.1). Several amplicons covering the promoter region, the first exon, the first intron, and

part of the second exon in *PLFY* were sequenced with various pairs of primers (Supporting Information Table S3.2). Most of the first exon in both *PAG* genes was amplified with several PCR reactions (Supporting Information Tables S3.1 and S3.2).

3.2.3 CRISPR Cas9 Target Site Selection

We chose two different target sites for each gene (Fig. 3.1), *PLFY*, *PAG1*, and *PAG2*, with the help of the sgRNA design online tool ZiFit ((Sander *et al.*, 2007, 2010; Mali *et al.*, 2013; Hwang *et al.*, 2013), <http://zifit.partners.org/ZiFiT/>). The same target sites were selected for *PAG1* and *PAG2* to allow for dual gene targeting. Based on the partial sequence we had for each gene (Lu *et al.*, 2016), we selected highly conserved sites that had no known sequence variants. However, we renewed plant material before this study in 2012 and discovered a SNP in the *PAG2* gene that was not detected there in previous work.

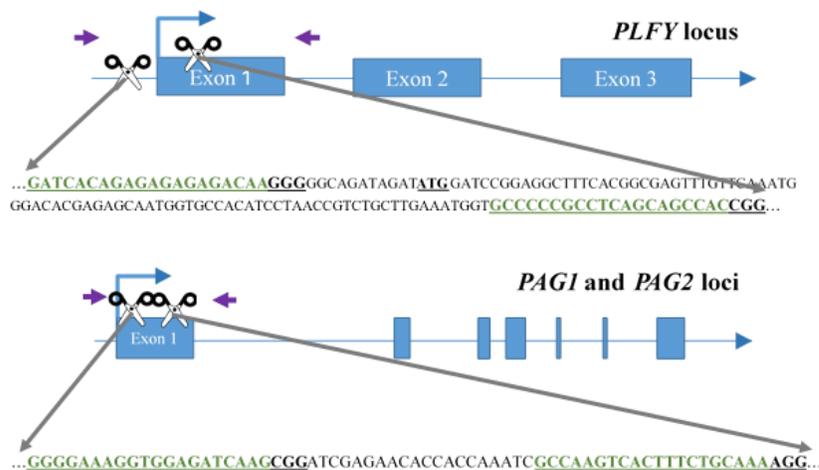


Fig. 3.1 CRISPR Cas9 sgRNA design and mutation detection in *LFY* and *AG* paralogs. Schematic representations of the target sites and the PCR assay for Sanger Sequencing. Exons and introns are represented by blue boxes and blue lines, respectively. The scissors indicate the target sites for each Cas9 nuclease. The purple arrows indicate the approximate location of the primers for sequencing. The target sites are colored in green inside the partial gene sequence. The underlined ATG in *LFY* indicates the location of the translation start codon.

For each target gene, we chose one target site either in the promoter region or at the beginning of the coding region, and the second target site tens to hundreds of bases 3' in the first exon (Fig. 3.1). The purpose was to choose targets far enough from each other to

create a large deletion when both sgRNAs were present. The target sites selected had a ‘G’ as their first base to function as the RNA polymerase start site and were followed by ‘NRG’ given *Streptococcus pyogenes* Cas9 preference for that sequence as the Protospacer Adjacent Motif (PAM).

3.2.4 CRISPR Cas9 Construct Assembly

To implement the CRISPR Cas9 system in *Populus*, we selected vectors (AtU6-26SK and 35S-Cas9-SK) that had previously been proven highly active in *Arabidopsis* (Feng *et al.*, 2013). We chose a double 35S promoter to drive the Cas9 to guarantee high expression and a human-codon optimized Cas9 because it is shown to be highly efficient in plants (Belhaj *et al.*, 2013). We assembled seven CRISPR Cas9 constructs; three to target *PLFY*, three to target both *PAGs* genes, and an empty-vector control for expression of Cas9 in the absence of sgRNAs (Fig. 3.2). Out of each three constructs targeting a specific gene or genes, two constructs contained only one sgRNA and the last construct had both sgRNAs together. The AtU6-26SK and 35S-Cas9-SK intermediary vectors were used to assemble all the CRISPR Cas9 constructs (Feng *et al.*, 2013). Final constructs were assembled as previously described (Feng *et al.*, 2013). In brief, two single-stranded 24 bp oligos were purchased from IDT (Coralville, IA, USA) for each sgRNA, where oligo 1 was of the form: bps “GATT” followed by 20 bps matching the target site and oligo 2 was of the form: bases “AAAC” followed by 20 bps matching the reverse complement of the target site). Each pair of oligos corresponding to a sgRNA was phosphorylated and annealed together in a reaction using T4 Polynucleotide Kinase (T4 PNK, NEB BioLabs, Beverly, MA) and an oligo concentration of 100 μ M (thermocycler parameters: 37°C for 30 min, 95°C for 5 min, then ramp down to 25°C by decreasing 5°C every minute). The AtU6-26SK was then digested with BbsI (NEB). Each pair of annealed oligos was ligated into the digested AtU6-26SK vector using T4 ligase (NEB). For the construct with two sgRNAs, the AtU6-26SK vector with the second sgRNA was used as template in a PCR reaction (Mullis *et al.*, 1986) and the section containing the promoter, the sgRNA, and the terminator was amplified with primers (IDT) containing 5'-KpnI and 3'-EcoRI sites. The PCR amplicon and the AtU6-26SK vector with the first

sgRNA were digested with KpnI-HF (NEB) and EcoRI-HF (NEB) and ligated together using T4 ligase (NEB). Next, the promoter, sgRNA, and terminator cassettes (with one or two sgRNAs) in the modified AtU6-26SK vectors and the 35S-Cas9-SK vector were digested with HindIII (NEB) and ligated together using T4 ligase (NEB). Then, the plant expression vector pK2GW7 was digested with KpnI-HF (NEB) and ZraI (NEB). The entire piece containing the sgRNA expression cassette(s) and the Cas9 expression cassette in the modified 35S-Cas9-SK vector was digested with KpnI-HF (NEB) and SmaI (NEB) and ligated into the KpnI and ZraI sites in the already digested pK2GW7 using T4 ligase (NEB). For the empty-vector control construct, the Cas9 cassette was digested using KpnI-HF (NEB) and SmaI (NEB) from the 35S-Cas9-SK vector and ligated into the pK2GW7 already digested with KpnI and ZraI with T4 ligase (NEB). All restriction enzyme digestions were incubated for one hour at 37°C. After incubation each digestion reaction was run on a 1% agarose gel, extracted, and purified using the Zymoclean Gel DNA Recovery kit (Zymo Research). All ligation reactions were incubated at 16°C for 12 hours. After each ligation, in house-made DH5 α *Escheria coli* cells were transformed, plated in antibiotic solid *Luria-Bertani* media with agar (Bertani, 1951), and grown overnight for further cloning.

3.2.5 Agrobacterium-mediated transformation

pK2GW7 constructs with CRISPR Cas9 cassettes (one or two sgRNAs and the Cas9 enzyme sequence) were transformed into *Agrobacterium tumefaciens* AGL1 using the freeze and thaw method (Weigel and Glazebrook, 2006). Each CRISPR Cas9 construct was transformed into hybrid poplar using standard methods (Filichkin et al., 2006). In brief, leaf, petiole, and stem explants from 353 and 717 in-vitro grown plants were cocultivated with each strain of AGL1 (containing one CRISPR Cas9 construct) for 48 hours in callus induction media (CIM) in the dark. Following this, the explants were washed and then moved to CIM with antibiotic for 3 weeks of culture in the dark. After significant calli could be seen with the naked eye, the explants were moved to shoot induction media with antibiotic for six to eight weeks, subculturing at 3- to 4-week intervals. After shoots became visible, explants were moved to shoot elongation media

with antibiotic for two to three weeks. Last, shoots were moved to rooting media with antibiotic for three to four weeks. Individual transgenic events were confirmed at this point and further micropropagated.

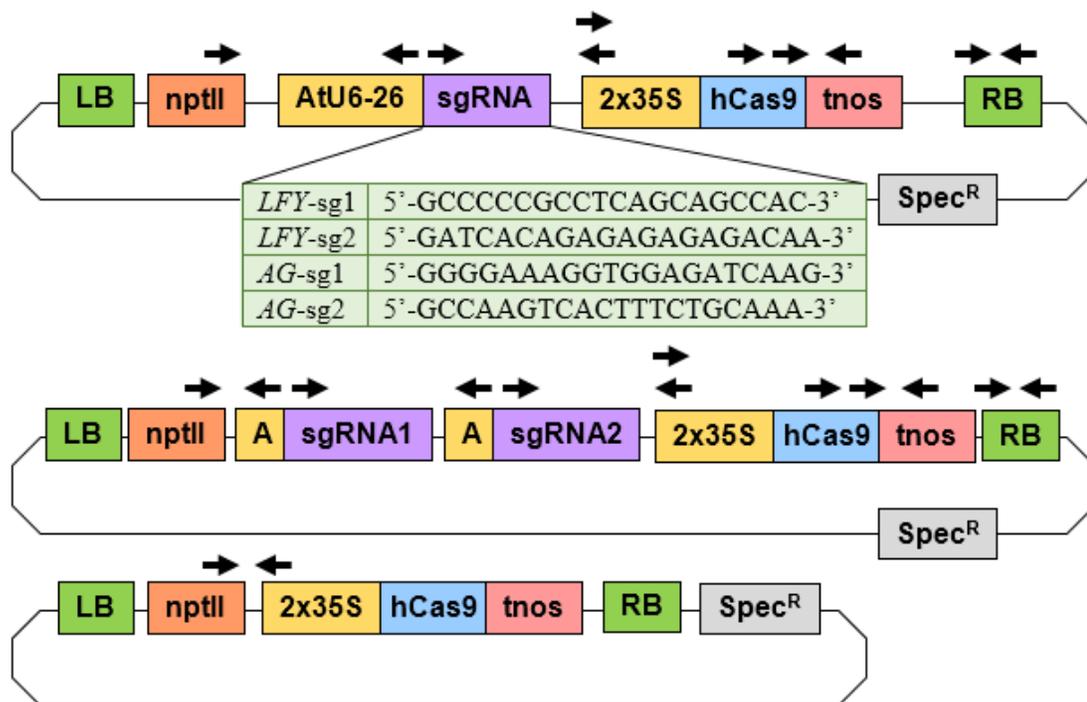


Fig. 3.2 Experimental constructs targeting one or two loci simultaneously. The construct at the top was used to target a single site in the target gene(s). The table below shows the specific sequence of each sgRNA. The plasmid on the middle was used to target two loci in same the gene(s). The plasmid on bottom was the Cas9 control plasmid with no sgRNA. The arrows indicate the primers used to verify the genetic sequence of the plasmids and to determine if the independent insertion events were transgenic. 2X35S, double *Cauliflower mosaic virus* (CaMV) 35S gene promoter; AtU6-26 or A, *Arabidopsis thaliana* U6-26 gene promoter; hCas9, human codon-optimized Cas9 gene sequence from *Streptococcus pyogenes*; LB, left T-DNA border; nptII, neomycin phosphotransferase II gene sequence for kanamycin resistance; RB, right T-DNA border; sgRNA, gene-specific sgRNA sequence; Spec^R, spectinomycin resistance gene sequence; tnos, termination region of the nopalene synthetase gene from *Agrobacterium tumefaciens*.

3.2.6 DNA Isolation and Transgene Confirmation

Shoot tip and leaf tissue from *in vitro* propagated 717 and 353 individual shoots were harvested for genomic DNA extraction according to Crowley et al. (Crowley *et al.*, 2003). Genomic DNA concentration and purity for some of the events was determined

using a Nanodrop 2000 spectrophotometer (www.nanodrop.com). The presence of the transgene was verified using PCR (Mullis *et al.*, 1986) with Econotaq DNA Polymerase (Lucigen, Middleton, Wisconsin, USA) and two sets of primers (IDT); one set near the left T-DNA border (AtU626_F1 and sgRNA_R1, Supporting Information Table S3.2), and another set near the right T-DNA border (Cas9_end_F2 and tnos_R2) (Fig. 3.2, Supporting Information Table S3.2).

3.2.7 Mutation Identification

We used PCR (Mullis *et al.*, 1986) to amplify the genomic region flanking all of the target sites. We amplified the promoter and the entire first exon in *PLFY* in order to identify as many mutation types as possible. The farthest forward and reverse primer were 229 bp upstream of *LFY*-sg2 and 333 bp downstream of *LFY*-sg1 respectively (LFY_seq_F7 and LFY_R2; product size 702 bp). For *PAG1* and *PAG2*, we amplified most of the first exon from both genes. In *PAG1*, our forward primer was 73 bp upstream of *AG*-sg1 and 138 bp downstream of *AG*-sg2 (AG1_seq_F1 and AG1_seq_R4; product size 323 bp). In *PAG2*, our forward primer was 81 bp upstream of *AG*-sg1 and 344 bp downstream of *AG*-sg2 (AG2_seq_F1 and AG2_seq_R5; product size 529 bp). Individual amplicons from each transgenic event were run on agarose gels. Bands were excised using a clean razor and DNA extracted using the QIAEX II Gel Extraction kit (Qiagen, Hilden, Germany) or the Zymoclean Gel DNA Recovery kit (Zymo Research) following the manufacturer's instructions. The pairs of primers used for sequencing *PLFY* were LFY_seq_F1 or LFY_seq_F7 and LFY_R2 (Supporting Information Table S3.2). The primers used for sequencing *PAG1* were AG_seq_F1 or AG1_seq_F1 and AG1_seq_R4 (Supporting Information Table S3.2). The primers used for sequencing *PAG2* were AG2_seq_F1 and AG2_seq_R5. The primers used for allelic-specific PCR when sequencing *PAG1* in clone 717 were AG1I_F1 (allele one) or AG1II_F2 (allele two) and AG1_seq_R4. The primers used for allelic-specific PCR when sequencing *PAG2* in clone 717 were AG2_seq_F1 and AG2I_R4 (allele one) or AG2II_R4 (allele two). The primers used for allelic-specific PCR when sequencing *PAG1* in clone 353 were AG1I_353_F1 (allele one) or AG1II_353_F1 (allele two) and AG1_seq_R4. The primers used for

allelic-specific PCR when sequencing *PAG2* in clone 353 were AG2_seq_F1 and AG2I_353_R2 (allele one) or AG2II_353_R2 (allele two). The sequence of each purified PCR product was defined using Sanger Sequencing by the Center for Genome Research and Biocomputing (CGRB) at Oregon State University. Individual sequences were aligned to the wild type (WT) sequences using MEGA6 (Tamura *et al.*, 2013). Partial amino acid sequences were translated using MEGA6 to determine the severity of the mutation on the predicted final peptide sequence (Supporting Information Fig. S3.1).

3.2.8 Haplotype Validation

We identified six natural SNP variants in *PLFY* in 717, two in *PAG1*, and eight in *PAG2* (Supporting Information Table S3.1). The two haplotypes are CGCTTG and TATCGA for *PLFY*, AG and GA for *PAG1*, and AATGCCCT and GCCATTTTC for *PAG2*. For clone 353, we identified five SNP variants in *PLFY*, one in *PAG1*, and five in *PAG2* (Supporting Information Table S3.1). In clone 353, the two haplotypes are ATTCC and GCCTT for *PLFY*, A and C for *PAG1*, and CATGT and AGCTA for *PAG2*. We used these SNP variants and the haplotypes they defined to ensure that both alleles had been amplified for each target gene.

3.2.9 Allele Characterization

We started our analysis of mutations by simultaneously amplifying both alleles of our insertion events in each PCR product. Given that most of the events with guide RNAs had different genotypes on each allele, our trace files showed double peaks. Initially to obtain an approximate ratio between biallelic (two altered alleles) and heterozygous (one altered allele and one WT allele) events, we amplified the promoter and first exon of *PLFY* for a randomly selected group of events, subcloned the allele-specific amplicons into pCR4-TOPO vector (www.invitrogen.com), and transformed DH5alpha *E. coli* cells. We included a few of randomly selected homozygous mutants to certify that both alleles indeed had the same mutation. The separation of alleles allowed us to determine the specific natural haplotypes of WT 717. We also used TOPO cloning to determine the sequences of the alleles of *PAG2* for a selected group of events that were homozygous

mutants in *PAG1*. We amplified both alleles simultaneously for all of our empty-vector control events because we did not expect to have different genotypes at each allele.

As we found that many events had different alleles, we utilized the online tool DSDecode (Liu *et al.*, 2015) to genotype events with chromatograms that showed heterozygous sequences. The ab1 file with the sequence information for each event and the WT sequence of the corresponding gene were uploaded to the DSDecode online tool. Last, results were manually confirmed by locating the double peaks in the ab1 files and by ensuring that the cleavage sites were in the target regions of the sgRNAs.

For a quarter (27.9%) of our transgenic events, we used allele-specific PCR (Newton *et al.*, 1989; Cha *et al.*, 1992) to identify the mutations in both alleles in both *PAG1* and *PAG2*. Allele-specific primers were designed based on the natural allelic variants in each allele (Supporting Information Table S3.1).

3.2.10 Characterization of mutation spectra

We compared mutation types with a prevalence higher or equal to 4.5% in most gene-sgRNA combinations (i.e. *LFY*-sg1, *LFY*-sg2, *AG1*-sg1, and *AG2*-sg2) using Pearson's Chi Square Test of Independence to test for equality of proportions (Supporting Information Table S3.6). We also employed the same test to determine if the same gene-double sgRNA combination (i.e. *LFY*-sg1sg2, *AG1*-sg1sg2, and *AG2*-sg1sg2) had the same profile in both hybrid clones (Supporting Information Tables S3.6, S3.7, and S3.8). All analyses were performed in R 3.4.1 (R Core Team, 2017) using the `chisq.test` function from the MASS package (Venables & Ripley, 2002). Monte Carlo simulation of 2,000 replicates were done when the sample sizes were less than 100. When referring to small indel mutation, we summed the number of small deletions and small insertions.

We used the Probe Search from the sPta717 Genome (Xue *et al.*, 2015; Zhou *et al.*, 2015) and the Cas-OFFinder online algorithm (Bae *et al.*, 2014) to identify genes that contained putative off-target sites in their coding region and had two or less mismatches when compared to the 'seed section' of the target site (last 12 bps of the sgRNA sequence) (Sternberg *et al.*, 2015) (Supporting Information Table S3.12). We selected two genes

with off-target sites that matched 17 and 16 of the 20 bases in *LFY*-sg1 and three genes with sites that all matched 17 out of the 20 bases in *AG*-sg2. The genes that partly match *LFY*-sg1 were Potri.001G254500 and Potri.009G049600 and matched all but 2bp in the seed sequence and all but 3 and 4bp in the entire sgRNA sequence respectively. The three genes that partly matched *AG*-sg2 were Potri.005G156900, Potri.013G104900, and Potri.019G077200, and they had only two mismatches in the seed region and three mismatches in the entire 20bp sequence.

Potri.001G25450/Potri.009G049600 and Potri.013G104900/ Potri.019G077200 are pairs of paralogs and share 88.8% and 93.8% of amino acid similarity with each other respectively. Potri.001G254500 and Potri.009G049600 encode proteins similar to *Arabidopsis* *UBIQUITIN-CONJUGATING ENZYME 19 (UBC19)* and *UBIQUITIN-CONJUGATING ENZYME 20 (UBC20)*. Potri.013G104900 and Potri.019G077200 encode a MADS box transcription factor homologous to *SEEDSTICK (STK)*, also known as *AGL11*, gene id *AT4G09960*) in *Arabidopsis*. Potri.005G156900 encodes for *UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 36/42 (USP36)* similar to *UBIQUITIN-SPECIFIC PROTEASE 16* in *Arabidopsis*. None of the off-target sites had allelic variants in the sgRNA target sites (i.e., natural SNPs). We sequenced 19 events that had mutations in *PLFY* and 39 events that had mutations in *PAG1* and *PAG2*; plants were sampled for DNA extraction after 4 to 10 months of *in vitro* propagation. Between three to five PCR products were isolated together from gel using either the QIAEX II Gel Extraction kit (Qiagen) or the Zymoclean Gel DNA Recovery kit (Zymo Research). Sequences were defined by the Sanger Sequencing service at the CGRB. To estimate maximum off-target rates, we calculated the rates as $1/(N\text{-alleles})$, and then the standard error using binomial expectation of: $\text{square root}[(pq)/(2N)]$.

3.3 Results

3.3.1 High knockout rates in *PLFY*

For *LFY*-sg1, out of 114 independent events, 103 had mutations in at least one allele and 90 events had both alleles defined by sequencing (Table 3.1). Out of the 90 defined

events, 15 had the same mutations in both alleles (homozygous mutants), 54 had a different mutation in each allele (biallelic mutants), two were chimeric with three mutant alleles observed, eight had one mutated allele and one WT allele (heterozygous mutants), and the remaining 11 had two WT alleles (Table 3.1). In summary, 71 of 114 independent events had all alleles altered making the potential total knockout rate 62.3%.

For *LFY*-sg2, out of the 45 independent events, 42 had mutations in at least one allele and 38 had both alleles defined (Table 3.1). Out of the 38 defined events, twelve were homozygous mutants, 22 were biallelic mutants, one was a heterozygous mutant, and three had no mutations on both alleles (Table 3.1). Given the location of *LFY*-sg2 in the promoter region and all of the mutations being small indels, we did not expect to get any knockout phenotypes in this group.

We generated transgenic independent events with two sgRNAs in both 717 and 353 hybrid clones. For *LFY*-sg1sg2 in 717, we generated 87 independent events and found 84 had mutations in at least one allele and 73 that had both alleles defined by sequencing (Table 3.1). Out of the 73 defined events, six were homozygotes, 58 were bi-allelic mutants, three were chimeric with all altered alleles, three were WT chimeras (two mutated alleles and a third WT allele), one was a heterozygote, and three had two WT alleles (Table 3.1). Thus, there were 67 of 87 independent events with both alleles altered and the putative knockout rate was 77.0%.

For *LFY*-sg1sg2 in 353, we sequenced 33 transgenic events, 30 had at least one allele mutated and 26 had both alleles defined by sequencing (Table 3.1). Out of the 26 events, seven were homozygous mutants, 15 were biallelic mutants, one was a chimera with all altered alleles, and three had two WT alleles (Table 3.1), summing to 23 of 33 independent events with altered alleles and a putative knockout rate of 69.7%.

3.3.2 High double knockout rates in *PAG* genes

Poplars have two orthologous genes to *Arabidopsis*' *AG* gene. The second *PAG* gene was generated during a recent partial genome duplication that happened between 35 and 18 million years ago (MYA) (Tuskan *et al.*, 2006). Thus, we were simultaneously targeting

four gene copies with two sgRNAs. For analysis of the first guide RNA in *AG1*, (*AG1*-sg1), we sequenced 64 independent transgenic events and none of them had any mutations (Table 3.1). For analysis of the same guide RNA in the *AG2* gene (*AG2*-sg1), we sequenced eight of the 64 independent transgenic events from the *AG1*-sg1 group and saw no mutations (Table 3.1). In summary, analysis of the sg1 guide RNA in both *AG* genes (*AG1*-sg1 and *AG2*-sg1), there were no events with altered alleles and the putative knockout rate was 0.0%.

For analysis of the second *AG* guide RNA in the *AG1* gene (*AG1*-sg2), we sequenced 61 events, and 58 had mutations in at least one allele and 59 had both alleles defined by sequencing (Table 3.1). Out of the 59 events, six were homozygous mutants, 48 were biallelic mutants, two were heterozygous mutants, and three had no mutations in either allele (Table 3.1), equating to 54 of 61 independent events with altered alleles and a putative knockout rate of 88.5%. For *AG2*-sg2, we sequenced 64 events (61 events with *PAG1* sequenced plus three more); 61 had mutations in at least one allele and 59 had both alleles defined (Table 3.1). Out of the 59 events, six were homozygous mutants, 47 were biallelic mutants, one was a chimera with all altered alleles, one was a heterozygous mutant, and four had no mutations in either allele (Table 3.1), equating to 54 events with altered alleles and a putative knockout rate of 84.4%. Out of the 64 events with *AG*-sg2 for which we sequenced *PAG2*, two had only one allele defined (both mutations) and 52 had both alleles altered in *PAG1*. Thus, 52 (81.3%) of 64 events were putative double knockouts in *PAG1* and *PAG2*.

For the *AG* construct with two guide RNAs, we started with analysis of the *AG1* gene (*AG1*-sg1sg2) in clone 717. We generated 118 independent events and found that 103 of them appeared to have mutations in at least one allele; in 89 of these both alleles were defined by sequencing (Table 3.1). Out of the 89 defined events, eight were homozygotes, 67 were bi-alleles, one was a WT chimera, three were heterozygotes, and ten had two WT alleles (Table 3.1), totaling 75 of 118 independent events with altered alleles and a putative knockout rate of 63.6%. For *AG2*-sg1sg2 in 717, we sequenced 24 (out of the 118 we sequenced for *AG1*-sg1sg2) transgenic events; 22 had mutations in at

Table 3.1. Numbers of mutants and rates of mutagenesis according to target gene, sgRNA, and clone. The events with both alleles defined were used to calculate mutation rates and to separate events according to putative phenotype (knock-out or WT). We described an event as a “knock-out” in none of its alleles had WT sequence and “WT” if one or more of its alleles had WT sequence. A “chimeric” knock-out had three mutated alleles. A chimeric WT had two mutated alleles and one WT allele. A, altered allele; Heteroz., heterozygote; Homoz., homozygote; W, WT allele. Different numbers in the subscript of the alleles stand for different alleles.

Gene-sgRNA	Clone	Total events (N)	Events w/both alleles defined (N)	Events with all alleles altered			Events with one or more WT alleles		
				Homoz. (A ₁ /A ₁)	Bi-allele (A ₁ /A ₂)	Chimera (A ₁ /A ₂ /A ₃)	Chimera (A ₁ /A ₂ /W)	Heteroz. (A ₁ /W)	WT (W/W)
<i>LFY</i> -sg1	717	114	90	15 (13.2%)	54 (47.4%)	2 (1.8%)	0 (0.0%)	8 (7.0%)	11 (9.6%)
<i>LFY</i> -sg2		45	38	12 (26.7%)	22 (48.9%)	0 (0.0%)	0 (0.0%)	1 (2.2%)	3 (6.7%)
<i>LFY</i> -sg1sg2		87	73	6 (6.9%)	58 (66.7%)	3 (3.4%)	3 (3.4%)	1 (1.1%)	2 (2.3%)
<i>AG1</i> -sg1		64	64	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	64 (100.0%)
<i>AG2</i> -sg1		8	8	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (100.0%)

Table 3.1. Numbers of mutants and rates of mutagenesis according to target gene, sgRNA, and clone (continued).

<i>AG1</i> -sg2		61	59	6 (9.8%)	48 (78.7%)	0 (0.0%)	0 (0.0%)	2 (3.3%)	3 (4.9%)
<i>AG2</i> -sg2		64	59	6 (9.4%)	47 (73.4%)	1 (1.6%)	0 (0.0%)	1 (1.6%)	3 (4.7%)
<i>AG1</i> -sg1sg2		118	89	8 (6.8%)	67 (56.8%)	0 (0.0%)	1 (0.8%)	3 (2.5%)	10 (8.5%)
<i>AG2</i> -sg1sg2		24	20	2 (8.3%)	13 (54.2%)	2 (8.3%)	0 (0.0%)	1 (4.2%)	2 (8.3%)
<i>LFY</i> -sg1sg2	353	33	26	7 (21.2%)	15 (45.5%)	1 (3.0%)	0 (0.0%)	0 (0.0%)	3 (9.1%)
<i>AG1</i> -sg1sg2		31	30	1 (3.2%)	25 (80.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (12.9%)
<i>AG2</i> -sg1sg2		35	35	4 (11.4%)	26 (74.3%)	0 (0.0%)	0 (0.0%)	1 (2.9%)	4 (11.4%)
Total		684	591	67 (9.8%)	375 (54.8%)	9 (1.3%)	4 (0.6%)	18 (2.6%)	117 (17.1%)
Total (w/out AG-sg1)		612	519	67 (10.9%)	375 (61.3%)	9	4	18 (2.9%)	45 (7.4%)

					(1.5%)	(0.7%)		
--	--	--	--	--	--------	--------	--	--

least one allele and 20 had both alleles defined (Table 3.1). Out of the 20 defined events, two were homozygotes, thirteen were bi-alleles, two were chimera with all altered alleles, one was a heterozygote, and two had no mutation in either allele (Table 3.1), summing to 17 of 24 events with alleles altered and a putative knockout rate of 70.8%. Out of the 24 events with *AG*-sg1sg2 for which we sequenced *PAG2*, one had only one allele amplified in *PAG1*, one had both WT alleles, and 15 were putative knockouts. Therefore, 15 (62.5%) of 24 events were putative double knockouts in *PAG1* and *PAG2*.

For *AG1*-sg1sg2 in 353, we sequenced 31 transgenic events, 27 had at least one allele mutated and 30 had both alleles defined by sequencing (Table 3.1). Out of the 30 events, one was a homozygote, 25 were biallelic mutants, and four had two WT alleles (Table 3.1), totaling 26 of 31 events with both copies altered and a putative knockout rate of 83.9%. For *AG2*-sg1sg2 in 353, we sequenced 35 transgenic events and all of them had both alleles defined (Table 3.1). Out of the 35 events, four were homozygous mutants, 26 were biallelic mutants, one was a WT chimera, and four had two WT alleles (Table 3.1), summing to 30 of 35 events with altered alleles and a putative knockout rate of 85.7%. Out of the 30 events with both alleles altered in *PAG2*, 22 were sequenced in *PAG1*, of which one had only one allele defined and 21 had all four gene copies altered making the putative double knockout rate 95.5%.

3.3.3 No mutations detected in Cas9-only transgenic controls

A total of 49 empty vector control events that had only the Cas9 gene sequence had no mutations in both alleles of *PLFY*, *PAG1*, and *PAG2* (totaling 294 different gene amplicons) (Table S3.3). Out of the 49 independent events, 32 were in 717-1B4 and 17 were in 353-53 (Table S3.3).

3.3.4 Mutation types correspond to activity and number of sgRNAs

Events generated with one active sgRNA had mostly small deletions (60.9% to 79.5%, Table 3.2) and secondly small insertions (17.0% to 33.3%, Table 3.2). Meanwhile, events with two active sgRNAs targeting the same gene (i.e. *LFY*-sg1sg2) had mainly large deletions (64.1% to 90.7% in 717 and 353 respectively, Table 3.2, Fig. 3.3A) and

Table 3.2. Mutation types. Rates of major classes of mutations from each gene-sgRNA combination. Undefined refers to insertion lines whose alleles were difficult to define by DSDecode. The most prevalent mutation type is highlighted in green and bold and the second most prevalent type in yellow and italics. Small refers to mutations of 15bp or less. Invers., inversion; N, number; subs., substitution; Undef., undefined.

Gene-sgRNA	Clone	Alleles defined (N)	Mutation in each allele							
			Small deletion	Small insertion	Small subs.	Large deletion	Large insertion	Invers.	Large subs.	Undef.
<i>LFY</i> -sg1	717	174	106 (60.9%)	<i>58</i> (33.3%)	0 (0.0%)	2 (1.1%)	5 (2.9%)	0 (0.0%)	0 (0.0%)	3 (1.7%)
<i>LFY</i> -sg2		76	53 (69.7%)	<i>20</i> (26.3%)	1 (1.3%)	1 (1.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (1.3%)
<i>LFY</i> -sg1sg2		153	<i>31</i> (20.3%)	15 (9.8%)	0 (0.0%)	98 (64.1%)	0 (0.0%)	8 (5.2%)	0 (0.0%)	1 (0.7%)
<i>AG1</i> -sg1		64	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
<i>AG2</i> -sg1		8	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

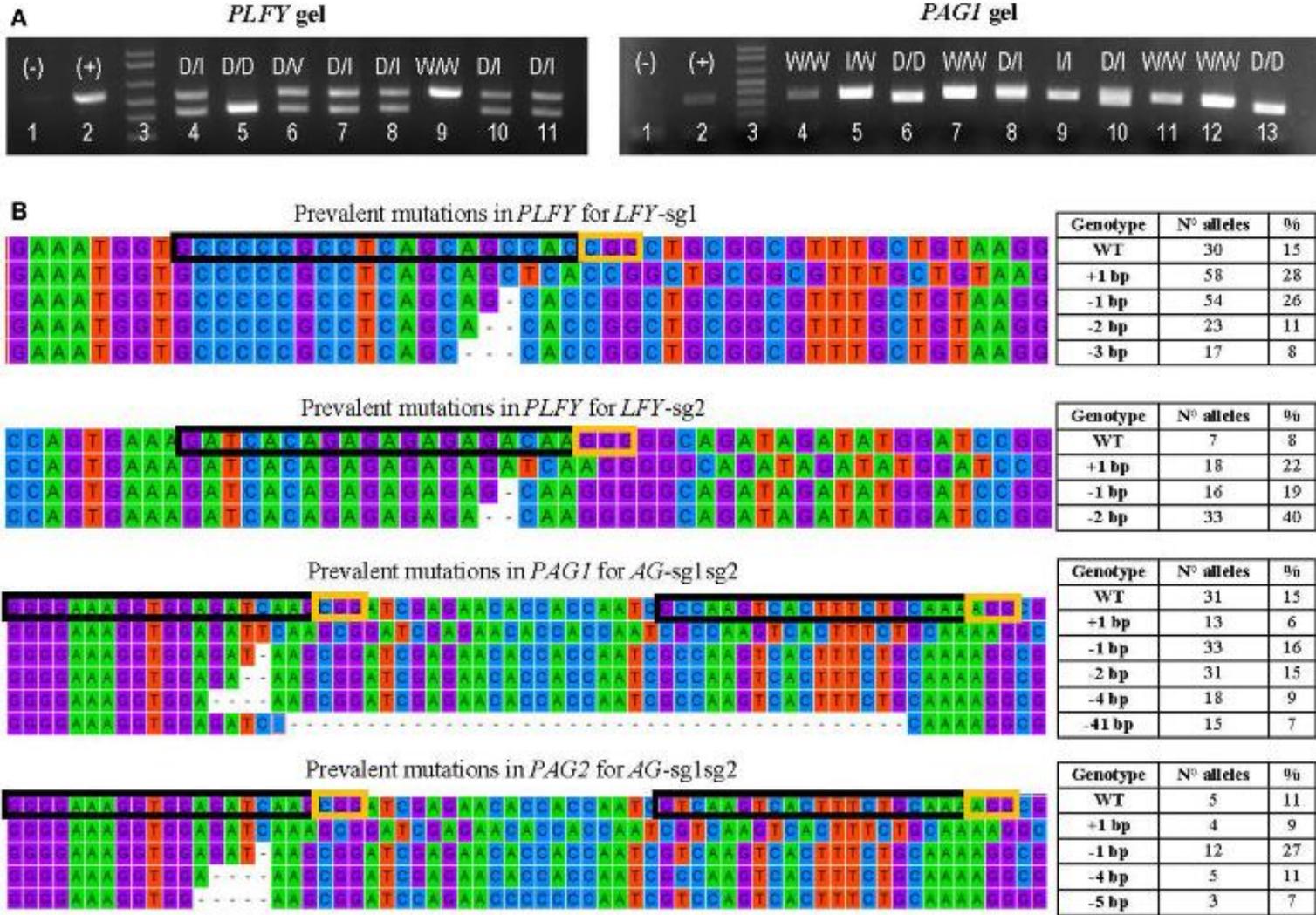


Fig. 3.3 Transformation event genotyping of *LFY* and *AG* paralogs.

Fig. 3.3 Transformation event genotyping of *LFY* and *AG* paralogs. (A) Example of gels of *PLFY* and *PAG1* PCR for insertion events with two sgRNAs. Symbols above each lane indicate the sequencing results of the DNA band(s). (+), positive control (-), negative control D, large deletion (16 or more base pairs); I, indel (insertion or deletion of 14 or fewer base pairs); V, inversion; W, wildtype. (B) Examples of the mutation types seen in alleles from mutants with one sgRNA in *PLFY* and two sgRNAs in *PAG1* and *PAG2*. The top alignment shows the partial sequence of *PLFY* flanking *LFY*-sg1 in the coding region. The second from the top alignment shows the partial sequence of *PLFY* flanking *LFY*-sg2 in the promoter region. The third from the top alignment shows the partial sequence of *PAG1* between *AG*-sg2 and *AG*-sg1. The bottom alignment shows the partial sequence of *PAG2* between *AG*-sg2 and *AG*-sg1. The protospacer sequence (i.e. target site) is surrounded by a black box. The PAM sites are surrounded by a yellow box. The dashes indicate deleted base pairs. The tables on the right indicate the mutation seen in each row, the number of alleles with that mutation, and the percentage that the number represents in each group.

secondly small indel mutations (5.6% to 30.1% in 353 and 717 respectively, Table 3.2). Events with both sgRNAs targeting *PAGs* had mostly small indels (75.5% to 86.9%, Table 3.2) but large deletions were also seen (10.3% to 24.5%, Table 3.2). Events with a SNP in their target did not have any mutations (i.e. *AG1*-sg1 and *AG2*-sg1, Table 3.2).

3.3.5 Mutation spectra varies among sgRNA targets

After defining 1,159 alleles in 561 events (Table 3.2), we suspected that there might be distinct mutation spectra for each gene-sgRNA combination (Supporting Information Table S3.4). The combinations *LFY*-sg1, *LFY*-sg2, *AG1*-sg2, and *AG2*-sg2 in 717 all had significantly different mutation spectra (χ^2 : 105.05, 15 degrees of freedom (df), $p < 0.001$; Supporting Information Table S3.5). Among the 171 separate mutated alleles belonging to *LFY*-sg1, 33.9% had a 1bp insertion, 31.6% had a 1bp deletion, 13.6% had 2bp deletion, 9.9% had a three bp deletion, 2.3% had a 4bp deletion, and 8.8% had one of nine other possible mutations (Fig. 3.3B, Supporting Information Table S3.4).

Meanwhile, from the 75 alleles sequenced belonging to *LFY*-sg2, 44.0% had 2bp deletion, 24.0% had 1bp insertions, and 21.3% had 1bp deletions (Fig. 3.3B, Supporting Information Table S3.4). Among the 112 alleles belonging to *AG1*-sg2 and the 116 alleles belonging to *AG2*-sg2, most alleles had a 1bp deletion (20.5% for in *PAG1* and 35.3% in *PAG2*) (Fig. 3.3B, Supporting Information Table S3.4). Yet, for the rest of the alleles in *AG1*-sg2, 18.8% had a 4bp deletion, 16.1% had a 1bp insertion, and 12.5% had

a 2bp deletion (Supporting Information Table S3.4). Meanwhile, for the remaining alleles in *AG2*-sg2, 18.1% had a 1bp insertion, 14.7% had a 4bp deletion, and 8.6% had a 3bp deletion (Supporting Information Table S3.4). Nonetheless, the spectrum from *AG1*-sg2 is not significantly different from that of *AG2*-sg2 (χ^2 : 8.15, 5 df, $p > 0.05$) (Supporting Information Table S3.5). All other pair comparisons of mutation spectra differed significantly ($p < 0.001$, Supporting Information Table S3.5).

Given the difference in activity between *LFY*-sg1sg2 and either *AG1*-sg1sg2 or *AG2*-sg1sg2, we did not consider it meaningful to compare their mutation spectra.

Nonetheless, we decided to compare the mutation spectrum of *LFY*-sg1sg2 in 717 and in 353 (Supporting Information Tables S3.6) and the mutation spectrum of both *AG1*-sg1sg2 and *AG2*-sg1sg2 in 717 and 353 (Supporting Information Tables S3.8 and S3.10). Events with *LFY*-sg1sg2 in 717 and in 353 had a significantly different mutation spectrum ($p < 0.001$, Supporting Information Table S3.7). Meanwhile, 717 and 353 events with either *AG1*-sg1sg2 or *AG2*-sg1sg2 did not have significantly different mutation spectra ($p \gg 0.05$, Supporting Information Tables S3.9 and S3.11).

3.3.6 Absence of mutations detected in off-target sites

A concern in using site-directed mutagenesis is the possibility of off-target mutations. We identified two potential off-site target sites that were similar to the target sites of *PLFY*, and three that were similar to the target sites of the *PAG* genes (Table S12). We selected events for analysis in which the desired target sites were mutated, indicative of a functional CRISPR Cas9 locus. In total, we genotyped 310 alleles for off-target mutations, but saw no mutations in any of these sequences. Specifically, we found no mutations in either allele of 19 transgenic events with mutations in *PLFY* in both of the selected genes, and also saw no mutations in either allele of the 39 transgenic events with mutation in the selected *PAG1* off-target genes (Table S12). Thus, the off-target mutation rate is expected to be less than about 5% for the *PLFY* off-targets ($2.6 \pm 1.8\%$) and less than about 2% for the *PAG1* off-targets ($1.3 \pm 0.9\%$).

3.4 Discussion

The purpose of this work was to examine the mutagenesis efficiency and pattern produced by CRISPR Cas9 nucleases directed at endogenous floral genes of poplar. Because poplars are naturally outcrossing species with high levels of heterozygosity, it was essential to characterize both alleles at each locus using allele-specific primers or by cloning and sequencing PCR products using conserved, primer sites. Initially, we amplified both alleles together, and used the DSDecode software to analyze difficult heterozygous samples (Ma *et al.*, 2016). However, for 717 events with AG-sg2 and the 353 events with AG-sg1sg2, we amplified and sequenced separate alleles using allele-specific primers for both *PAG1* and *PAG2*. A few mutated lines had both alleles amplified together that were difficult to genotype with certainty by DSDecode, and we labeled them as “undefined” (Table 2).

A minor goal of this research work was to determine the prevalence of off-target mutations. We did not detect any mutations in 155 amplicons from specific loci (total of 310 alleles), corresponding to five off-target sites. These potential targets were similar to either of our *PLFY* or *PAG* target sites, differing in only three or four bases out of 20 base pairs of the sgRNA. The events surveyed, which included the entire CRISPR Cas9 locus, had been growing in Magenta boxes for 6-12 months, and subcultured every 2-3 months, before tissue was sampled for DNA isolation, providing ample time for mutagenesis. A lack of off-target mutagenesis has been reported in many CRISPR Cas studies in plants (*A. thaliana*, *N. benthamiana*, hybrid poplar, rice, soybean, sweet orange, and wheat) with up to 7 mismatches (Lawrenson *et al.*, 2015; Schiml & Puchta, 2016; Sauer *et al.*, 2016; Wolt *et al.*, 2016). They have also not been detected in three genome scale studies (Feng *et al.*, 2013; Peterson *et al.*, 2016). However, off-target mutagenesis has been detected in a few plant studies, with rates ranging from 1.6% to 13.0% with one or two mismatches in the last 12bp of the sgRNA (Xie & Yang, 2013; Jacobs *et al.*, 2015; Lawrenson *et al.*, 2015; Sauer *et al.*, 2016) and with rates ranging between 1.6% and 9.7% with one to three mismatches in the first eight bp (Upadhyay *et al.*, 2013; Zhang *et al.*, 2014; Xu *et al.*, 2015). One case that is of interest found mutations in T1 rice plants that had constitutive Cas9 and sgRNA expression, similar to our own studies (Xu *et al.*

2015). Clearly, off-target rates appear to be low, but additional studies are needed, especially in systems such as trees where CRISPR Cas expression may continue for many months or even years.

No mutations were seen in either allele of the three target genes, *PLFY*, *PAG1*, and *PAG2*, in 49 empty vector control events that were transformed with the Cas9 gene sequence but no sgRNA. Thus, as expected the CRISPR Cas9 system requires both a nuclease and fully functional RNA components for specific mutagenesis, and shows that somaclonal variation associated with *in vitro* culture and *Agrobacterium* transformation had a negligible influence by comparison. Given our large sample size, we were able to characterize mutations according to type for each sgRNA. The specific class of mutation seen depended on the number of sgRNAs present in the binary vector. As in other plant studies, most of the events with one active sgRNA had small deletions or single base insertions (reviewed by (Bortesi *et al.*, 2016). Meanwhile, lines with two active sgRNAs targeting the same gene, i.e. *LFY*-sg1sg2, had mainly large deletions (between 64.1% and 90.7%) removing the DNA between the sites, many indels (between 5.6% and 30.1%), and some inversions (between 1.9% and 5.2%). This is the third study on plants that reports inversions. Large deletions and inversions have also been reported in *Arabidopsis* (Zhang *et al.*, 2017) and rice (Zhou *et al.*, 2014b; Liang *et al.*, 2016) when using two sgRNAs separated between 200bp or 245 kb. However, our independent events transformed with two sgRNAs that were not of comparable activity, i.e. *AG*-sg1sg2, had mainly small deletions like those lines transformed with only one sgRNA.

The most common peptide modifications expected from translating the altered alleles with only one sgRNA (i.e. *LFY*-sg1 and *AG*-sg2) or two sgRNAs with one inactive (i.e. *AG*-sg1sg2) included removal of essential amino acids (see -3bp deletion with *LFY*-sg1 in Supporting Information Fig. S3.1), early stop codons, and frame-shifted proteins (Supporting Information Fig. S3.1). We occasionally saw insertions leading to predicted peptides with extra amino acids (data not shown). We did not translate the peptide sequence for *LFY*-sg2 because this sgRNA targeted the promoter, so we do not expect it to modify the *PLFY* protein sequence. With two active sgRNAs, we mainly predicted truncated or frame-shift proteins.

In this study, we characterized a large number of events (684) and alleles (1,159) by direct Sanger Sequencing. From this data, we noticed that most of the gene-sgRNA combinations had a unique mutation spectrum, suggesting that their distinct sequences or the adjacent chromosome region affect the character of the resulting mutations.

Van Overbeek et al. (2016) first described such an effect in a study done on 223 CRISPR Cas9 target sites within human cells. They found that the specific mutation seen for each target sequence were likely due to the local adjacent sequence and not due to the guide RNA sequence per se or the genomic region.

Another goal was to select sgRNAs that would be able to induce mutations in more than one gene to get a complete loss-of-function mutant. For *PAG* we needed to alter four gene copies, the two alleles of *PAG1* and the two alleles of *PAG2*, as these two *AG*-like genes appear to share protein function (Brunner *et al.*, 2000). Successful multi-gene targeting has been previously documented in pig, mouse, and moss (Wang *et al.*, 2013; Yang *et al.*, 2015; Lopez-Obando *et al.*, 2016). The sgRNA *AG*-sg2 had high mutation rates in both *PAG1* and *PAG2*, generating several potential complete *PAG* loss-of-function (i.e., double putative knockout) mutants. Out of 54 events transformed with *AG*-sg2 with both *PAG1* and *PAG2* defined, 52 of 64 (81.3%) events were confirmed putative double knockouts in both *AG* genes. The *AG*-sg1sg2 sgRNA was also highly active. Out of the 24 events transformed with *AG*-sg1sg2 in clone 717 with both *PAG1* and *PAG2* defined, 15 (62.5%) were double putative knockouts. In addition, out of the 22 events transformed with *AG*-sg1sg2 in clone 353 with both *PAG1* and *PAG2* defined, 21 (95.5%) were double putative knockouts.

A major goal was to study the rate at which the system produced complete knockouts (i.e. loss-of-function) events for each of our target genes. The *AG*-sg1 nuclease however, induced no mutations in either *PAG1* or *PAG2*. This lack of mutation was likely in part due to the presence of a SNP in *PAG2* in our new 717 stock (Zhou et al. 2015), and possibly also low activity by the sgRNA. Nonetheless, when this guide RNA was present in a construct with a second, active guide RNA, we observed several deletions with an endpoint at the target of this otherwise inactive sgRNA, indicating it may have retained some level of Cas9 guide activity.

Three of the four CRISPR Cas9 nucleases, i.e. *LFY-sg1*, *LFY-sg2*, and *AG-sg2*, generated high rates of mutagenesis in their corresponding target gene(s) when acting individually, creating many putative loss-of-function lines. Of all the events with either *LFY-sg1* or *LFY-sg1sg2* in 717, 62.3% and 77.0% respectively, are putative proteins knockouts. In 353, 69.7% of the events are also putative protein knockouts, and like in 717, they had mainly truncated and/or frame-shifted proteins. Clearly, CRISPR Cas9 is a very powerful technology that, for the first time, can readily generate loss of function mutations at single loci as well as at the paralogous gene families that are so prevalent in poplar (Tuskan et al. 2006) and many other plant species.

Acknowledgements

We thank Dr. Jian-Kang Zhu at Purdue University and Dr. Yanfei Mao at Shanghai Center for Plant Stress Biology for providing us with the vectors AtU6-26SK and 35S-Cas9-SK. We thank Xinmin An for checking several of our events for transgene presence. We thank many undergraduate students for their help, including Gillian Bergmann, Clark Embleton, Ruchira Agarwal, Melissa Meyhoff, and Analeslie Martinez for helping with media preparation, plant transformation, and plant selection. We thank Gilles Pilate and Lise Jouanin of INRA, France for use of poplar clones 353 and 717. We also thank the members of the Tree Biosafety and Genomics Research Cooperative (TBGRC) at OSU, the United States Department of Agriculture (USDA award 2011-68005-30407, System For Advanced Biofuels Production From Woody Biomass In The Pacific Northwest), the USDA Biotechnology Risk Assessment (grants 2011-68005-30407 and 2010-33522-21736), and the NSF I/UCRC Center for Advanced Forestry (grant 0736283).

4 CRISPR disruption of *LEAFY* function in *Eucalyptus* gives sterile indeterminate inflorescences but normal vegetative development

Estefania Elorriaga, Amy L. Klocko, Cathleen Ma, Marc du Plessis, Xinmin An, Alexander A. Myburg, and Steven H. Strauss

New Phytologist

Bailrigg House
Lancaster University
Lancaster LA1 4YE, UK
(2019) 222: 923–937

In review

Contributions of authors

Estefania Elorriaga designed the study with the help of Amy L. Klocko and Steven H. Strauss. Estefania Elorriaga sequenced the target gene; designed and constructed the vectors; gathered, analyzed, and interpreted the data; and wrote the manuscript. Cathleen Ma performed the plant transformation, regeneration, selection, and transplanting for the greenhouse study. Amy L. Klocko helped with vector construction. Marc du Plessis created Fig. 4.3. Xinmin An sequenced several independent transgenic events. Alexander A. Myburg supervised the work at University of Pretoria. Steven H. Strauss with help from Amy L. Klocko obtained funding for the study and supervised the overall study.

Summary

- *Eucalyptus* is among the most widely planted taxa of forest trees worldwide. However, its spread as an exotic or genetically engineered form can create ecological and social problems.
- To mitigate the risk of gene flow via pollen and seeds, we mutated the *Eucalyptus* ortholog of *LEAFY* (*LFY*) by transforming a wild type *Eucalyptus grandis* × *urophylla* hybrid and two *Flowering Locus T* (*FT*) overexpressing (i.e., early flowering) lines of the same genotype with CRISPR Cas9 constructs targeting *LFY*.
- We achieved high rates of *lfy* biallelic knock-outs, often approaching 100% of transgenic insertion events. Frameshift mutations in early-flowering, *AtFT*-overexpression backgrounds had strong floral alterations including indeterminacy in floral development and an absence of viable male or female gametes, and did not differ statistically in vegetative growth rate or leaf morphology from transgenic controls in greenhouse trials. Genes upstream or near to *LFY* in the floral development pathway were hyperexpressed, whereas floral organ identity genes downstream of *LFY* were severely depressed, showing an inability to progress towards floral organ differentiation.
- We conclude that disruption of *LFY* function appears to be capable of efficient genetic containment while exhibiting no detectable effects on vegetative growth rate or morphology.

Keywords: *LEAFY*, *Eucalyptus*, flowering, CRISPR, containment.

4.1 Introduction

Forest plantations cover about 7% of the world's forests and one quarter of these are comprised of non-native species and interspecific hybrids (FAO, 2010). Although non-native trees are often preferred because of their high productivity (Richardson, 1998; Dodet & Collet, 2012), their success also results from extensive genetic improvement programs and intensive silvicultural practices (Gonçalves *et al.*, 2013; Crous *et al.*, 2019). Plantation forestry often introduces exotic trees over vast areas, which in some cases can lead to encroachment and/or genetic admixture into native ecosystems (Wilson *et al.*, 2009; Donaldson *et al.*, 2014).

Eucalyptus (family Myrtaceae) is among the most widely planted genera of forest trees, with the largest areas of plantation occurring in Brazil (5.7 million ha), China (4.5 million ha), and India (3.9 million ha) (FRA *et al.* 2018). Although large commercial plantations are often established using vegetative propagules (Nakhoda & Jain, 2016), “feral” eucalypts spread mostly by seed, and then once “naturalized,” pollen flow can enable larger scale movement. As a means for containment of exotic or genetically engineered trees, elimination of sexual reproduction would greatly reduce the potential for spread and invasiveness, while retaining desirable vegetative growth and adaptability traits inherent to the modified genotypes, including their ability to be clonally propagated.

A flower is an angiosperm structure that often has sterile showy organs (usually a perianth, the combination of sepals and petals) to attract pollinators, and stamens and/or carpels to enable sexual reproduction. While floral development involves many highly conserved gene families and key regulators, there are important differences among families, genera, and species. For example, *Eucalyptus* trees (family Myrtaceae) have distinctive bisexual flowers with a modified perianth. The word *Eucalyptus* comes from the Greek words ‘eu’ meaning ‘well’ and ‘calyptos’ meaning ‘covered’ (ευκάλυπτος), referring to the opercula that covers the flower. *Eucalyptus* flowers do not have a traditional perianth. Instead, their whorls of sepals and petals are replaced by two opercula, the calicine (outer) operculum and the coroline (inner) operculum.

Flowers can be single or compound (i.e., a cluster or group of flowers). The cluster of flowers is known as an inflorescence. How, when, and where flowers are found in the plant also depends on the family, genus, and species. When *Arabidopsis* (a model herbaceous plant) transitions to flowering, the apical indeterminate meristem transforms to an inflorescence meristem and eventually to a determinate floral meristem. In most eucalypts, the apical meristem remains vegetative (indeterminate) throughout the life of the plant, whereas axillary meristems may become shoots or (in mature trees) inflorescences. Each inflorescence consists of a single flower bud or a cluster (i.e. umbel) of three to 15 flower buds. Early in development, the flower cluster is covered by one or two bracts, and each bud is usually covered by a pair of bracteoles. Understanding the genetic mechanisms controlling the transition to reproductive competency, flower initiation, and flower development in *Eucalyptus* would both shed light on floral evolution, and identify targets for reproductive containment.

The molecular mechanisms that regulate flowering in *Arabidopsis* have received much attention. The floral pathway integrator (FPI) genes perceive environmental and endogenous signals and initiate the transition into reproductive growth, the floral meristem identity (FMI) genes convert inflorescence meristems to floral meristems, and the floral organ identity (FOI) genes regulate expression of genes that produce floral organs (reviewed in Pajoro et al., 2014). The key floral regulator *LEAFY* (*LFY*) is an FPI and an FMI gene and was one of the first flowering genes identified (Coen *et al.*, 1990b; Weigel *et al.*, 1992). It encodes a highly conserved plant-specific transcription factor found in all land plants, including non-flowering plants (Moyroud *et al.*, 2009; Silva *et al.*, 2016), and streptophyte algae (Gao *et al.*, 2019a).

LFY has high expression in floral meristems in both *Arabidopsis* and *Antirrhinum* (Coen *et al.*, 1990b; Weigel *et al.*, 1992). In other angiosperms, *LFY* is mainly expressed in floral meristematic and primordial organs, yet some vegetative expression has also been seen (Hofer *et al.*, 1997; Molinero-Rosales *et al.*, 1999; Rottmann *et al.*, 2000; Ahearn *et al.*, 2001). *ELFY*, the homolog in *Eucalyptus*, in particular has high expression in the tips of leaf primordia and in flower meristems (Dornelas *et al.*, 2004). *Pinus radiata* and other gymnosperms have two *LFY* homologs; *PRFLL* and *NEEDLY* (*NLY*) (Mouradov *et*

al., 1998; Mellerowicz *et al.*, 1998). Both *PRFLL* and *NLY* are expressed in cone primordia, but also in vegetative meristems (Mouradov *et al.*, 1998; Mellerowicz *et al.*, 1998; Vázquez-Lobo *et al.*, 2007). Likewise, the two *LFY* homologs in the lycophyte *Isoetes* L. are expressed in both vegetative and reproductive tissues (Yang *et al.*, 2017). In contrast, the two *LFY* homologs in moss, fern, and streptophytic algae are expressed solely in embryonic and meristematic tissues (Tanahashi *et al.*, 2005; Plackett *et al.*, 2018; Gao *et al.*, 2019a).

The orthologs of *LFY* are present as single-copy genes in most land plants, except gymnosperms (Vázquez-Lobo *et al.*, 2007; Moyroud *et al.*, 2010). Loss-of-function mutations in both *LFY* (or its orthologs) alleles lead to sterile, late flowering plants in *Arabidopsis* and tomato, and flowerless plants in *Antirrhinum* (Coen *et al.*, 1990b; Weigel *et al.*, 1992; Molinero-Rosales *et al.*, 1999). Because of the high level of conservation in plants, *LFY* is a good target for genetic containment of exotic and weedy species. However, loss-of-function mutations in *LFY* have only been characterized in the herbaceous plants *Arabidopsis*, *Antirrhinum*, and tomato, and *LFY* function and expression differ among angiosperms. In addition, apart from the partial loss-of-function field studies using RNA interference against the *LFY* homolog in poplar (Klocko *et al.*, 2016b), we are aware of no in depth studies of vegetative development, nor randomized experiments, to estimate impacts on biomass growth rate and vegetative morphology. Thus, it remains unclear whether *LFY* indeed has significant vegetative functions in the species where it shows vegetative expression.

The multiple year delay of flowering in trees presents a great logistical challenge to genetic studies of floral development. Fortunately, this can be overcome by precocious floral induction using chemical or genetic treatments, including overexpression of *FLOWERING LOCUS T (FT)*. *FT* is an FPI gene whose encoded protein (“florigen”) is transported from the leaves to the shoot apical meristems where it acts as a flowering hormone (Shim & Imaizumi, 2015). Constitutive or inducible overexpression of *FT* elicits early flowering in many herbaceous and woody species, including *Eucalyptus* (Endo *et al.*, 2005; Böhlenius *et al.*, 2006; Lifschitz & Eshed, 2006; Yamagishi *et al.*, 2011; Hsu *et al.*, 2011; Lee *et al.*, 2013; Klocko *et al.*, 2016d). In this study, to

understand the effects of CRISPR mutation of *LFY* on floral structure and function, we retransformed two previously-characterized early-flowering *Eucalyptus* lines that were shown to produce viable pollen and germinable seeds (Klocko *et al.*, 2016d). We generated three CRISPR Cas9 nuclease constructs and used them to target the dimerization domain of the *LFY* homolog in *Eucalyptus*, *ELFY*. Because overexpression of *FT* also adversely affects tree form, we conducted a second greenhouse study in a CRISPR-mutated wild type (non-*FT*) background to determine if mutation of *LFY* would affect vegetative traits and/or growth. We report that sterile, floral-like indeterminate organs, or an absence of flowers, were produced in all CRISPR mutant events, but there were no detectable effects on vegetative growth rate or leaf morphology.

4.2 Materials and Methods

4.2.1 Plant materials and growth conditions

Sterile *in-vitro* shoot cultures of wild type (WT) hybrid *Eucalyptus* clone SP7 (*Eucalyptus grandis* x *urophylla*) were kindly provided by FuturaGene (<http://www.futuraGene.com/pt/>). The two *AtFT* overexpressing lines (lines 4-2 and 30-3 transformed with pCAM:409S:AtFT under Hygromycin selection, *AtFT* hereafter) were previously generated in our laboratory (Klocko *et al.*, 2016d) and grown at 25°C under a 16-h day (photon flux density of 40 $\mu\text{mol m}^{-2}\text{s}^{-1}$) in shoot multiplication medium (SMM).

4.2.2 Target gene sequencing and CRISPR Cas9 target site selection

We determined the first exon's sequence of the *LFY* (GenBank accession number NM_125579, AT5G61850) ortholog, *ELFY* (GenBank accession number KK198763, Eucgr.K02192), in SP7 by amplifying both alleles separately in *E. coli* cells using pCR4-TOPO vectors (www.invitrogen.com) following the manufacturer's instructions. As in our study in poplar (Elorriaga *et al.*, 2018), we used the sgRNA design online tool ZiFit (Sander *et al.*, 2007, 2010) to identify two different target sites in *ELFY*; one downstream of the translation start site (from bp 52 to bp 72; Fig. 4.1**a,b**) and one at the end of the first exon (from bp 290 to bp 310; Fig. 4.1**a,b**).

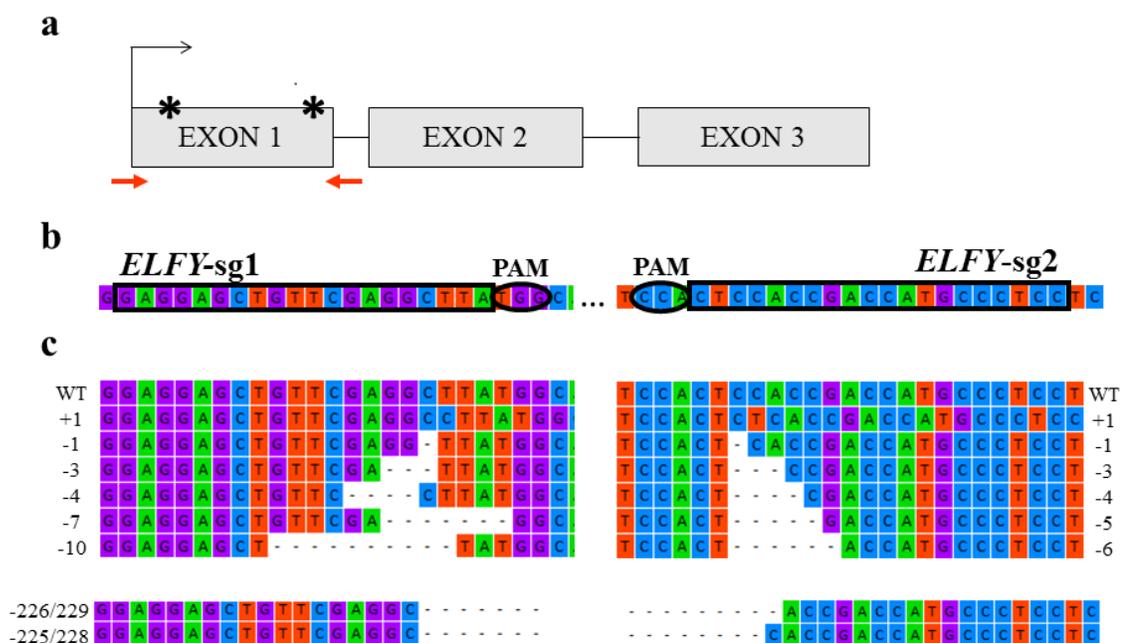


Fig. 4.1 Examples of nucleotide sequence alignments of gene-edited *ELFY* alleles. (a) Schematic of the *ELFY* gene with the two sgRNA:Cas9 targets (stars) and the sequencing primers (red arrows). (b) Nucleotide sequence proximal to the two target sites. The three periods correspond to the 213 bases (216 in *E. urophylla*) between target sites. The sequences matching the sgRNAs are surrounded by a black square and the protospacer adjacent motif (PAM) sequences by a black oval. (c) Common mutations seen among transgenic events modified with *ELFY-sg1* (top left), *ELFY-sg2* (top right), and *ELFY-sg1sg2* (bottom).

4.2.3 CRISPR Cas9 vector construction

We created three CRISPR Cas9 constructs to target *ELFY*. Two of the constructs contained only one sgRNA (i.e., *ELFY-sg1* and *ELFY-sg2*) and the third construct contained both sgRNAs together (i.e., *ELFY-sg1sg2*). Two intermediary vectors, AtU6-26SK and 35S-Cas9-SK, were used to assemble the three CRISPR Cas9 cassettes (Feng *et al.*, 2013). Final constructs were assembled as in our previous study (Elorriaga *et al.*, 2018) with some modifications. In brief, we phosphorylated, annealed, and cloned single-stranded 24 bp oligos (www.idtdna.com) into the BbsI sites on AtU6-26SK. Next, for the constructs with one sgRNA, we amplified and digested the PCR product that included each sgRNA transgene with KpnI and ClaI before ligating it into the KpnI and ClaI sites on the 35S-Cas9-SK vector. For the construct with both sgRNAs, we amplified

and digested the PCR product containing the second sgRNA with SalI and ClaI and ligated it into the same restriction sites on the AtU6-26SK vector already containing the first sgRNA transgene. The AtU6-26SK vector containing the two sgRNA transgenes was digested with KpnI and ClaI and the two transgenes were ligated into the KpnI and ClaI sites in the 35S-Cas9-SK vector. Last, we linearized the plant expression vector pK2GW7 using the restriction enzymes KpnI and ZraI and ligated in the Cas9 expression cassette plus either a single sgRNA transcriptional unit or both sgRNA transcriptional units after digesting with KpnI and SmaI. We used the same empty vector control as in our poplar study (Elorriaga *et al.*, 2018).

4.2.4 Plant transformation and regeneration

The three CRISPR Cas9 constructs and the empty vector control construct were transformed into WT and *AtFT SP7* using *Agrobacterium*-mediated transformation methods (Chauhan *et al.*, 2014). In brief, we wounded and cocultivated sterile WT and *AtFT SP7* leaf tissue with *Agrobacterium tumefaciens* AGL1 cells previously transformed with the constructs of interest (Weigel & Glazebrook, 2006). After 48 hours of cocultivation, we transferred the explants to callus induction medium (CIM). After one week on CIM, explants were moved to shoot induction medium (SIM) for several months. Individual shoots were collected and placed in shoot elongation medium (SEM) for several weeks. Then, shoots were sampled for DNA isolation and transgene genotyping. PCR-confirmed transgenic events were propagated in SMM. Last, individual ramets for each PCR-confirmed transgenic event were transferred to rooting media (RM). All the media contained kanamycin (75 mg/L) as the selective agent except for SMM and RM.

4.2.5 DNA isolation and transgene genotyping

DNA was isolated from shoot tip and leaf tissue of *in-vitro* propagated WT and *AtFT SP7* rooted shoots (Keb-Llanes *et al.*, 2002). We used a Nanodrop 2000 spectrophotometer (www.nanodrop.com) to determine yield and purity of individual genomic DNA samples. We verified that individual shoots were transgenic by amplifying with PCR sections of both ends of the T-DNA; a section near the left T-DNA border (AtU626_F1 and

sgRNA_R1, Supporting Information Table S4.1), and a section near the right T-DNA border (Cas9_end_F2 and RB_R2, Supporting Information Table S4.1). Events that had both T-DNA sections amplified with PCR were considered transgenic.

4.2.6 Haplotype validation and allele-specific PCR

We discovered natural allelic variants when we sequenced each *ELFY* allele separately in *E. coli*. We learned that there were three SNPs at positions 20, 339, and 346 starting from the translational start site (e.g. ATG correspond to positions 1, 2, and 3). We also found that one amino acid, H92, was present only in the allele from *E. urophylla*. We used the SNPs to design allele-specific primers. We used allele-specific PCR (Newton *et al.*, 1989; Cha *et al.*, 1992) to amplify the genomic region flanking both target sites (Egrandis_F3/Egrandis_R1 and Euro_F3/Euro_R1, product lengths 357 and 360 respectively; Supporting Information Table S4.1) and to determine the genotype of each allele for all events.

4.2.7 Mutation identification

We amplified sections of genomic DNA flanking both target sites using allele-specific PCR to determine the mutations in each allele for all events. PCR amplicons were run and excised from 1.5% agarose gels with a clean razor blade. DNA from PCR products was isolated using the Zymoclean Gel DNA Recovery kit (www.zymoresearch.com) following the manufacturer's instructions. The pairs of primers used for sequencing *ELFY* were the same used for PCR amplification (for *E. grandis* allele: Egrandis_F3 and Egrandis_R1; for the *E. urophylla* allele: Euro_F3 and Euro_R1; Supporting Information Table S4.1). Each amplicon was sequenced using the Sanger Sequencing service provided by the Center for Genome Research and Biocomputing (cgrb.oregonstate.edu/core/sanger-sequencing) at Oregon State University. Individual allele sequences were aligned to the WT allele sequences using MEGA6 (Tamura *et al.*, 2013). We used the MEGA6 translation feature to determine the partial peptide sequence to predict the severity of each mutation on the final *ELFY* protein.

4.2.8 Rooting and greenhouse conditions

Events with predicted loss-of-function mutations in both *ELFY* alleles were selected and propagated. These events were propagated to generate multiple identical ramets (trees) per independent transgenic event. Individual ramets were rooted and transferred to soil in two-inch square pots. While acclimating to soil, we kept the ramets in humid conditions in a glasshouse. After a month of acclimation, we moved the ramets to a greenhouse and transplanted each one to an eight-inch circular pot. All the *AtFT* CRISPR Cas9, *AtFT* Cas9, and *AtFT*-only events were randomized in one block with nontransgenic SP7 control ramets that were grown and propagated in tissue culture.

4.2.9 Vegetative data measurements and statistical analysis

We measured two wood yield-related traits, height and trunk diameter (or the diameter of the tallest stem for ramets with more than one stem). We also measured four leaf traits including relative SPAD value (a proxy of chlorophyll density), area, perimeter, and leaf weight. We assessed tree height from soil level to the apex of the main stem (or the highest stem, in cases where there were more than one) for each ramet. Trunk diameter was measured at four inches from soil-level with digital calipers. Height and trunk diameter were recorded a month after moving to the greenhouse, before any plants started competing for light. We used a SPAD meter (Konica Minolta) to measure relative leaf chlorophyll content. Relative leaf chlorophyll content was recorded two months after moving the plants to the greenhouse. We took three readings from two separate leaves for each ramet. The three readings from each leaf were averaged together. Three different leaves from each ramet were scanned using an HP Scanjet 8200. The leaf area and leaf perimeter of each scanned leaf were calculated using ImageJ (Schneider *et al.*, 2012). Dry leaf weight was recorded after desiccating the scanned leaves at 65°C for five days. A weight measurement was taken for each leaf, and the average of the three weights was used for the analysis. We also calculated two derived traits, stem volume index (= tree height*(trunk diameter²)) and specific leaf weight (= dry leaf weight for each leaf/leaf area for each leaf, also called leaf density). Stem volume index and

specific leaf weight are considered non-destructive proxies to biomass yield (Zianis *et al.*, 2005) and leaf photosynthesis (Criswell and Shibles 1971; Dornhoff and Shibles 1970).

We used general linear mixed effects model to determine if loss-of-function mutations had an effect on any of the vegetative traits measured. The model included genotype (FM for flowering mutant, escape, Cas9, and WT) as fixed effect, event as random effect, and residual error. Residual plots were used to check the equal variance and normality assumptions. We performed all of our statistical analyzes using R v3.6.1 (R Core Team, 2019). We used the R package nmle (Pinheiro *et al.*, 2018) to fit our data to our model. Means were estimated using the R package emmeans (<https://CRAN.R-project.org/package=emmeans>).

4.2.10 Analysis of floral morphology in FT trial

Flowering was first recorded when the ramets were moved to the greenhouse. Flower morphology was evaluated and recorded every month for twelve months. Flower buds and flowers were imaged whole and dissected using a Keyence VHX-1000 digital microscope for WT and FM plants. Buds and flowers were dissected to determine if any developing or underdeveloped reproductive organs were present.

4.2.11 RNA isolation and cDNA synthesis

We collected developing flower buds in the early afternoon of October 4th, 2018. We sampled buds from six FM events; two events were transformed with *ELFY*-sg1: 30-10 and 30-11, two with *ELFY*-sg2: 30-31 and 30-45, and the last two with *ELFY*-sg1sg2: 30-2 and 30-40. We also sampled buds from two Cas9 events, Cas9-30-14 and Cas9-30-5, and from two ramets of the *AtFT*-only insertion line 30. Two to three buds were collected from two ramets (approximately one gram of tissue in total) of the same event. The buds were sampled, frozen immediately in liquid N, and kept at -80°C until RNA isolation. RNA was extracted according to Howe *et al.* (2013). The RNA samples were treated with DNaseI (New England Biolabs) to remove any remaining genomic DNA. DNase-treated RNAs were submitted for analysis by the Agilent Bioanalyzer 2100 at the

CGRB to determine their integrity. The SuperScript III First-Strand Synthesis system (Invitrogen) was used to synthesize cDNA from the DNase-treated RNAs.

4.2.12 Gene expression and statistical analysis

Real-time quantitative PCR (qPCR) analysis was performed in a StepOnePlus Real-Time PCR system (Applied Biosystems). Each 20 μ l reaction contained 10 μ l of PowerUp SYBR Green Master Mix (ThermoFisher Scientific), 1.0 μ l (10 ng) of cDNA, 1.2 μ l of forward and reverse primers, and 7.8 μ l of water. Gene amplification was conducted under the following thermocycler conditions: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 60 s. Right after gene product amplification was completed, melt-curve analysis was performed by increasing the temperature by 0.3°C s⁻¹ between 60 and 95°C. We recorded the expression of *ELFY* and other genes in the flower development pathway (Smaczniak *et al.*, 2012; Bouché *et al.*, 2016; Theißen *et al.*, 2016; Wils & Kaufmann, 2017) that were upstream, downstream, or at the same developmental stage as *LFY* in *Arabidopsis* (Fig. 7). The specific genes were the orthologs in *Eucalyptus* of: *FLOWERING LOCUS T (FT)*, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3 (SPL3)*, *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 (SPL9)*, *CAULIFLOWER (CAL)*, *FRUITFULL (FUL)*, (there are two in *Eucalyptus*), *APETALA3 (AP3)*, *PISTILLATA (PI)*, *AGAMOUS (AG)*, *SHATTERPROOF 2 (SHP2)*, and *SEEDSTICK (STK)* (Table S2). The relative gene expression of each gene was determined using the delta-delta-Ct (ddCt) method. All reactions were done in triplicate. Expression was normalized using the *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (GAPDH)* gene as the housekeeping gene. We designed gene-specific primers by first using the PrimerQuest online tool (Integrated DNA Technologies) and then checking each pair's specificity in *E. grandis* by using Primer-BLAST (Ye *et al.*, 2012). The primer specificity was further tested using standard curve analysis by serial dilutions of cDNA (five 1:2 dilutions) for each gene in triplicate. All primers pairs had amplification efficiencies (E) between 90 and 110% and correlation coefficients (R²) higher than 98%. The DataAssist v3.01 software (Applied Biosystems) conducted a two-sample two-tailed Student's t-test to determine if expression of the control group was different to that of the FM group for each gene.

4.2.13 Peptide alignment

We used the UniProt protein database ('UniProt', 2019) to collect 19 sequences of homologs of *LFY* corresponding to 13 eudicots, two monocots, one tracheophyte, one conifer, one ginkgo, and one embryophyte. We used Clustal Omega (Madeira *et al.*, 2019) to align the peptide sequences and ESPript 3.0 (Robert & Gouet, 2014) to create the graphic.

4.3 Results

4.3.1 Mutation and knock-out rates among transgenic events were high

We generated nine and 59 transgenic insertion events with our three CRISPR Cas9 constructs for the WT trial and *FT* trial, respectively. For the WT trial, we were interested in determining if knocking out *ELFY* would have an effect on growth or morphology. For the *FT* (early flowering) trial our intent was to understand *ELFY*'s function in relation to flowering in eucalypts and thus to determine if it would be an effective target for containment. In the WT trial, all nine insertion events had mutations in both *ELFY* alleles (100% biallelic mutation rate, Table 4.1). The two empty vector controls (i.e., Cas9 only) did not have mutations on either allele of *ELFY*. In the *FT* trial, 58 out of 59 transgenic events had mutations in both *ELFY* alleles (98.3% biallelic mutation rate, Table 4.1) and the last transgenic event had a mutation only in the *E. urophylla* allele. The nine empty vector control events did not have mutations in either *ELFY* allele. Last, the mean mutation rate among all confirmed transgenic events was 98.5% (Table 4.1).

Based on their translated peptide sequence, 9 of 9 (100%, Supporting Information Table S4.3) and 53 of 59 (90%, Supporting Information Table S4.3) events in the WT trial and *FT* trial respectively, had knock-out mutations in both alleles, thus we expected they would have a flowering mutant (FM) phenotype. In the *FT* trial, we expected the remaining six of the 59 (10%, Supporting Information Table S4.3) events, including the monoallelic mutant, to have WT flowers.

Table 4.1 CRISPR mutation rates on a per-event and per-allele basis.

Population	Total events (alleles)	Alleles modified	N° events
WT LFY- CRISPR	9 (18)	Both alleles	9 (100%)
		One allele	0 (0%)
		None	0 (0%)
<i>AtFT</i> LFY- CRISPR	59 (118)	Both alleles	58 (98%)
		One allele	1 (2%)
		None	0 (0%)
All eucalypt	68 (136)	Both alleles	67 (99%)
		One allele	1 (1%)
		None	0 (0%)

4.3.2 No consistent differences between *ELFY* knock-outs and transgenic controls in vegetative traits

The purpose of the greenhouse WT trial was to determine if *ELFY* had any vegetative function that would affect growth or morphology. For this trial, we had nine CRISPR Cas9 (with *ELFY*-sg1, *ELFY*-sg2, or *ELFY*-sg1sg2) insertion events, two escapes (i.e., not detectably transgenic by PCR but were cocultivated with *Agrobacterium* and then regenerated with antibiotic selection), three Cas9 (i.e., empty vector) insertion events, and WT (i.e., not cocultivated). The nine CRISPR Cas9 insertion events had a total of 41 ramets. The escapes had 12 ramets (i.e., six ramets for each event). The three Cas9 events had 18 ramets. Last, there were seven WT ramets. In total, we monitored 78 ramets, and each insertion event had between three and six ramets. When analyzing the different traits measured, we found no significant differences in any comparisons between FM plants, escapes, empty vector controls, and WT controls in volume index,

leaf perimeter, leaf dry weight, and specific leaf weight ($P > 0.05$; Fig. 4.2, Supporting Information Fig. S2). However, leaf area varied among groups at the 5% significance level. The estimated mean leaf area was 27.18% lower in FM plants than in WT controls ($P = 0.049$; Fig. S2), but the FM group did not differ significantly from any of the transgenic control groups (Fig. S2).

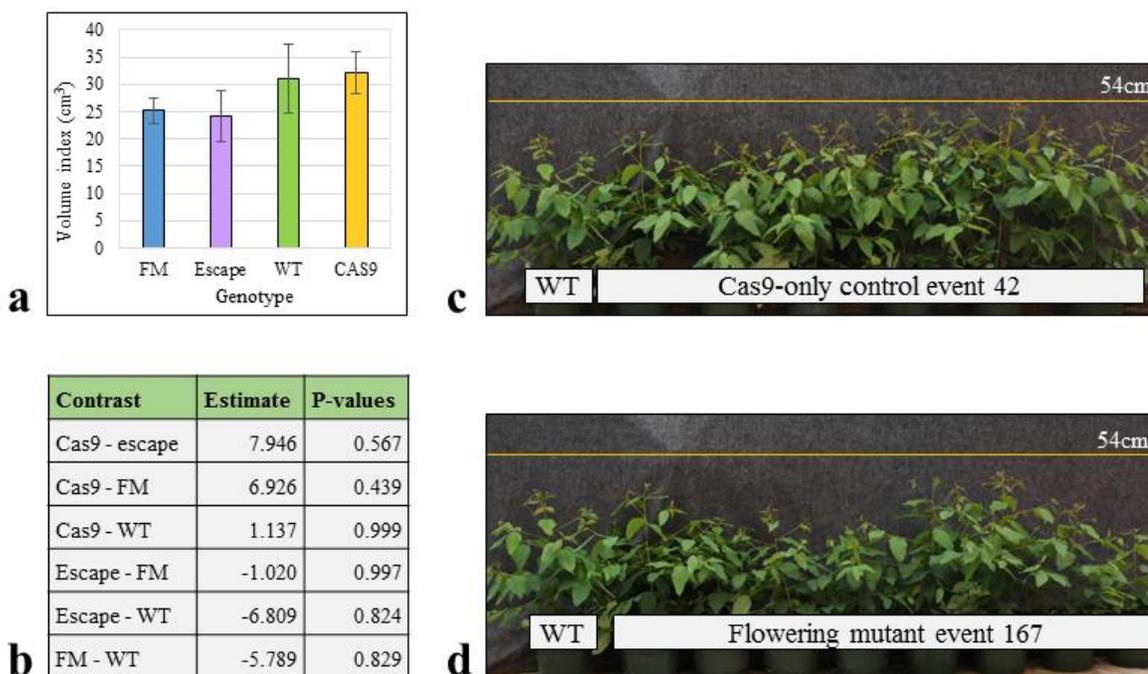


Fig. 4.2 Stem volume growth and plant form appear to be unaffected by knock-out of *ELFY*. **(a)** Mean stem volume index (height \times diameter²) for the flowering mutants and the three control groups. Error bars represent \pm SE of means. Cas9, transgenic lines that do not contain sgRNAs. Escape, non-transgenic but *Agrobacterium* cocultivated and regenerated lines. FM, flowering mutant. WT, wild type, not cocultivated or regenerated but micropropagated. **(b)** Table of estimated mean differences and p-values corresponding to a Student's t-test. **(c)** Image of potted reference WT ramet and the six ramets corresponding to Cas9 event 42. **(d)** Image of potted reference WT ramet and eight ramets of mutant event 167. The yellow lines in both photographs are at 54 cm height.

4.3.3 Most trees flowered in the FT greenhouse trial

The purpose of the *FT* (early flowering) trial was to determine the effect that modifying *ELFY* would have on floral morphology and reproductive viability in our hybrid eucalypts. For this trial, we had 42 CRISPR Cas9 insertion events (with *ELFY*-sg1,

ELFY-sg2, or *ELFY*-sg1sg2 and also *AtFT*), six Cas9 *AtFT* insertion events, and two different *AtFT* insertion events. Each one of the 42 CRISPR Cas9 insertion events had between one and seven ramets, totaling 166 ramets. There were 25 ramets that corresponded to the six Cas9 *AtFT* insertion events, 23 ramets that corresponded to the two different *AtFT* insertion events, and five WT ramets. In total, we monitored 219 ramets. All 42 CRISPR Cas9 insertion events produced biallelic mutants (Table S4). However, four events were predicted to have WT flowers based on their translated peptide sequences. The remaining 38 events were predicted to have alterations in their floral morphology (i.e., they were FM plants). None of the ramets of six of the 38 (15.8%) predicted FM events flowered during the study. Most ramets of the remaining events (i.e., 32 FM events, six Cas9 events, and two *AtFT*-only events) transitioned to reproductive stage. Two FM events, 4-7 and 4-65, had ramets that did not flower (Table S4).

4.3.4 Lack of differences between *ELFY* knock-out and controls in vegetative traits within flowering trees

We recorded and compared vegetative traits among the trees that flowered. The overexpression of *AtFT* eliminated the apical dominance in all these trees and as a result they had a bush-like form. After analyzing yield, SPAD values, and the four leaf traits (i.e., leaf perimeter, leaf area, leaf dry weight, and specific leaf weight), we found no significant differences after contrasting the means of all the genotype categories (i.e., Cas9, escape, FM, *AtFT*-only, and NM; Supporting Information Fig. S4.3, S4.5).

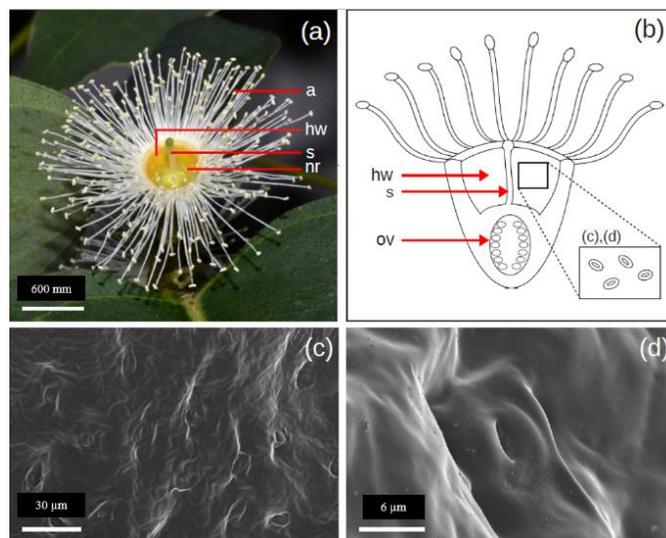


Fig. 4.3 Wildtype *E. grandis* flower images. **(a)** Opened flower showing anthers (a), style (s), hypanthium wall (hw), and nectary ring (nr). **(b)** Diagram of flower showing style (s), ovary (ov), and hypanthium wall (hw), as well as the position of the images shown in panels c and d. **(c)** and **(d)** Scanning electron microscopy (SEM) micrographs of hypanthium wall (hw) showing nectary pores at 1,150X and 7,480X magnification, respectively.



Fig. 4.4 Flower development stages in control and *ELFY* knock-outs. **(a - d)** Correspond to flowering tissues from *AtFT*-only (control) flowering-induced events. **(a)** Umbel with three flowering buds at the earliest recognizable stage. The buds have both calicine and coroline opercula. **(b)** Umbel with three flowering buds with bracts and calicine opercula shed. **(c)** Opened flower at anthesis. **(d)** One seedpod harvested about four months after anthesis. **(e - h)** Correspond to flowering tissues from knock-out events. **(e)** Umbel with three flowering buds at the earliest recognizable stage. At this stage, the flower buds from mutant events are indistinguishable from the flowering buds of flowering controls. **(f)** Umbel with four flowering buds with bracts shed. At this stage, flowering buds from mutant events are recognizably different from those of flowering controls. **(g)** Umbel with five mature buds generating and shedding layers of pedicels and bracts. **(h)** Umbel with three stacked floral-like organs showing indeterminacy and retention of senescent bracts eight to ten months after buds were discernible as in (f).

4.3.5 Knock-out mutants had either underdeveloped or no floral organs

We monitored flowering in the *AtFT* ramets for approximately 18 months. We re-sequenced the *ELFY* alleles of ten FM events to test whether the mutations seen early in development had changed because of the overexpression of Cas9 (more than three years elapsed from the first DNA extraction from tissue culture plants to resampling in the greenhouse). For this analysis, we sampled leaves from four different axillary stems. No changes in DNA sequence at the target sites were observed. Also, a greenhouse trial in University of Pretoria in South Africa with several of our *AtFT*-only, Cas9, and FM events showed floral phenotypes that were consistent with those seen in Oregon (Supporting Information Fig. S4.4, S4.6), providing further evidence that the mutations and phenotypic effects were stable.

The ramets from the four CRISPR Cas9 events predicted to have WT flowers produced flowers identical to those found in the six *AtFT* Cas9 events and the two *AtFT*-only events (Fig. 4.3, Fig. 4.4c; Supporting Information Fig. S4.4a). These flowers had a central pistil and a staminal ring at the base of the hypanthium (Fig. 4.3a, Fig. 4.4c, Supporting Information Fig. S4.4a). All the flowers appeared capable of secreting nectar through their nectary pores located on the hypanthium wall (e.g., Fig. 4.3b, c, d). We classified our 32 CRISPR Cas9 insertion events with non-WT reproductive organs as FM, even though we saw a range of sterile flower-like phenotypes. All the FM plants had “flowers” with repeated bract-like and pedicel-like organs (Fig. 4.4g, h, Supporting Information Fig. S4.4b), and the range in phenotypes went from flowers with two to three repeated layers of bract-like and pedicel-like organs with sterile underdeveloped anthers and/or underdeveloped ovules (Fig. 4.5i) to flowers with many repeated layers with no reproductive organs at all (Fig. 4.4h, Fig. 4.6, Supporting Information Fig. S4.4b).

Based solely on the phenotypes, we classified our FM plants into three categories (Supporting Information Table S4.4): “early organs,” if a few sterile underdeveloped ovules and/or underdeveloped stamens were present after only two or three layered pedicels (Fig. 4.5i); “late organs,” if a few sterile underdeveloped ovules were present after three to five layered pedicels (Fig. 4.5a thru h); and “organless,” if no reproductive organs were seen after more than five layered pedicels (Fig. 4.6, Supporting Information

Fig. S4.7b). On occasion, long-lived flowers (>5 months) of organless FM events would eventually produce underdeveloped sterile reproductive-like organs (Supporting Information Fig. S 4.8c,d) while most of the rest of the long-lived flowers never produced any reproductive organs (Supporting Information Fig. S4.8a,b). By contrast, wild type flowers usually developed over three to four months, with the seed capsules requiring an additional four to five months to mature and dehisce.

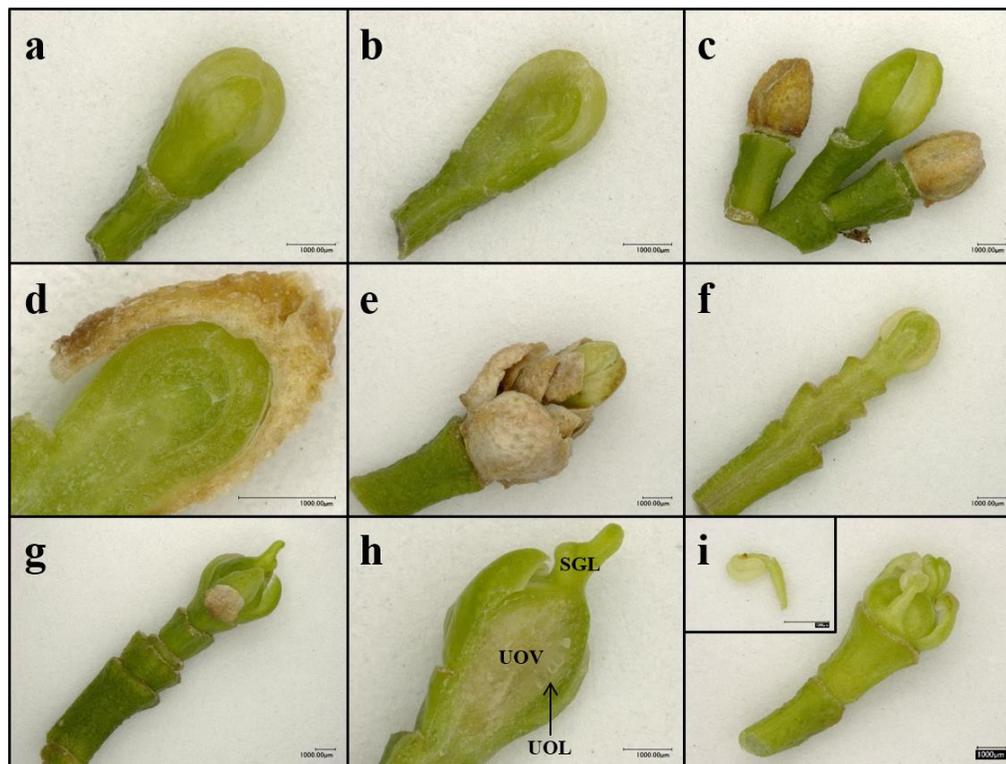


Fig. 4.5 Sterile floral-like buds with underdeveloped ovules belonging to mutant event 30-6. (a) Early bud and (b) its cross-section with no sign of reproductive organs. (c) Umbel with more developed buds than (a) and (b). (d) Cross-section of bud in (c) with no reproductive organs. (e) Late bud with four layered pedicels surrounded by many bract-like organs. (f) Cross-section of bud in (e) with no reproductive organs. (g) Late bud with four layered pedicels. (h) Cross-section of bud in (g) with underdeveloped ovules. (i) Sterile bud from FM event 30-30 with four underdeveloped stamen-like organs surrounding the hypanthium. One single stamen-like organ in photo insert. SGL, stigma-like. UOL, underdeveloped ovules. UOV, underdeveloped ovary.

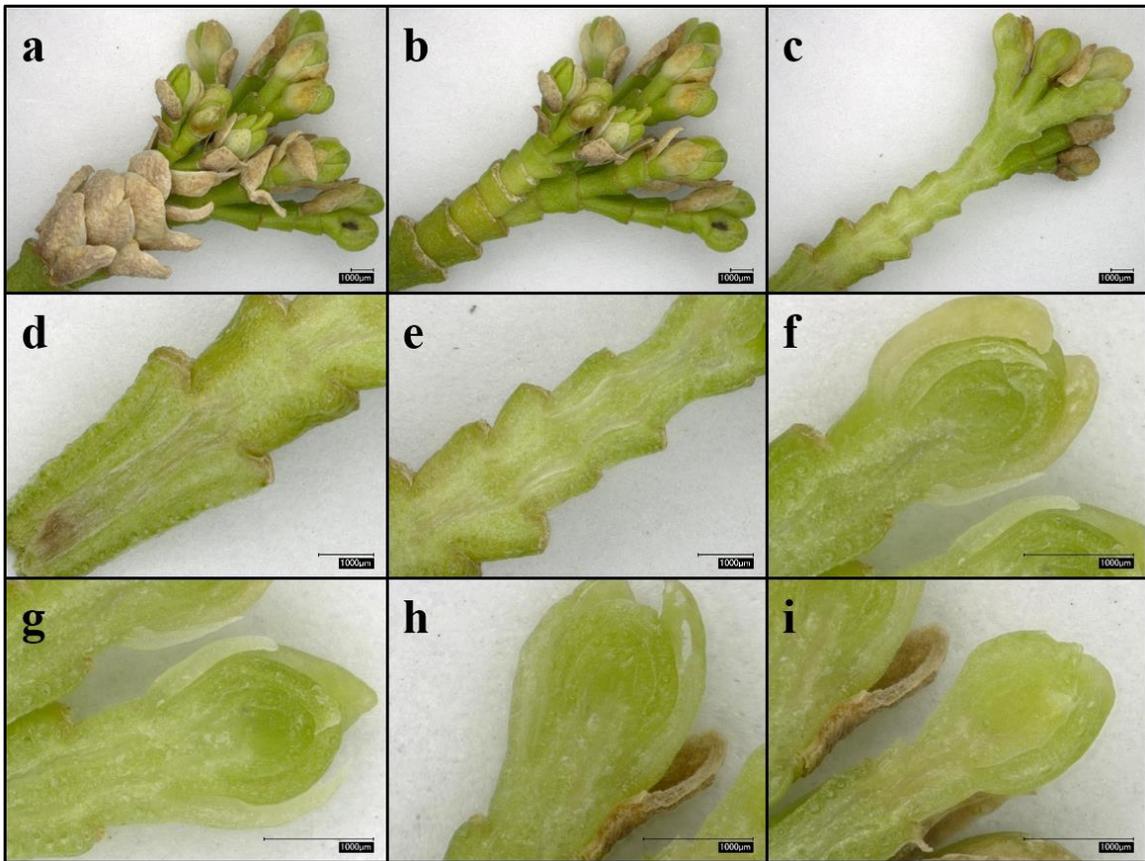


Fig. 4.6 Sterile floral-like buds from mutant event 30-10 with many repeated bract-like and pedicel-like organs, and no underdeveloped ovules or anthers. **(a)** Umbel surrounded by many bract-like organs. **(b)** Umbel with bracts removed showing three pedicel-like repeated layers before splitting and generating more pedicel-like organs. **(c)** thru **(i)** cross-sections of the buds showing the lack of discernible ovules or stamens.

4.3.6 Changes in expression of flowering genes

We selected buds from the *AtFT*-only and *AtFT Cas9* events that had just shed or were shedding their calicine operculum and were about a month away from anthesis (Fig. 4.7b). We selected buds from the FM events that were shedding or had just shed their first layer of bract-like organs (Fig. 4.7b). Differences in gene expression of twelve flowering genes including *ELFY* were analyzed for six FM events (two for each *ELFY*-sg1, *ELFY*-sg2, and *ELFY*-sg1sg2) and three flowering control events (one *AtFT*-only and two *AtFT Cas9* events). *ELFY* expression was significantly higher in the FM events than in the control events (mean of 636% higher expression than controls, $P = 0.02$; Fig. 4.7c). Expression of six genes (i.e., *EFT*, *ESPL3*, *ESPL9*, *ECAL*, *EFUL1*, and *EFUL2*)

upstream or at the same developmental time as *ELFY* was higher in the FM events than in the control events (Fig. 4.8). When comparing the expression between the control events and the FM events, the FM events had a mean fold-change in gene expression of 3.0 for *EFT* ($P = 0.003$), 4.4 for *ESPL3* ($P = 1.0E-4$), 2.9 for *ESPL9* ($P = 0$), 1.9 for *ECAL* ($P = 0.004$), 2.1 for *EFUL1* ($P = 0.002$), 2.6 for *EFUL2* ($P = 0$). Meanwhile, expression of five FOI genes that are induced by *ELFY*, directly or indirectly, (i.e., *EAP3*, *EPI*, *EAG*, *ESHP2*, and *ESTK*) was significantly lower in the FM events than in the control events (Fig. 4.9). When comparing the expression levels in the control events to the FM events, the control events had a mean fold difference in gene expression of 2,500.0 for *EAP3* ($P = 0.006$), 2.8 for *EPI* ($P = 3.0E-4$), 5.6 for *EAG* ($P = 0.009$), 6.6 for *ESHP2* ($P = 0.01$), and 178.6 for *ESTK* ($P = 0.01$).

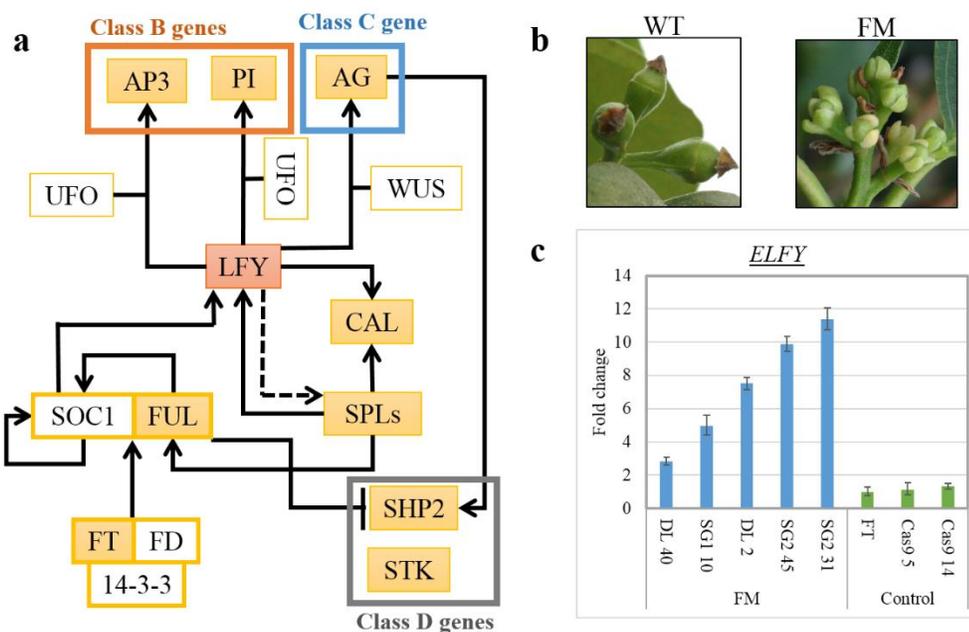


Fig. 4.7 Transcriptional network related to *ELFY*, and its expression from qPCR, in floral or floral-like buds. **(a)** Simplified genetic pathway from *Arabidopsis* (described in Methods). We performed gene expression analysis on genes with an orange (*LFY*) or yellow fill (Figs. 8 and 9). **(b)** Images of buds from *AtFT*-only event 4 and FM (floral mutant) event 30-2 showing the developmental stage at which tissues were sampled for gene expression analysis. **(c)** Gene expression seen in independent *ELFY*-FM and non-mutant flowering control events. The average fold-change in expression was calculated as a ratio to the *AtFT*-only control, which was set to 1. Error bars represent \pm SE of means. Gene expression was significantly different when comparing mean expression for the FM events to control events ($P = 0.02$, two-tailed Student's t-test).

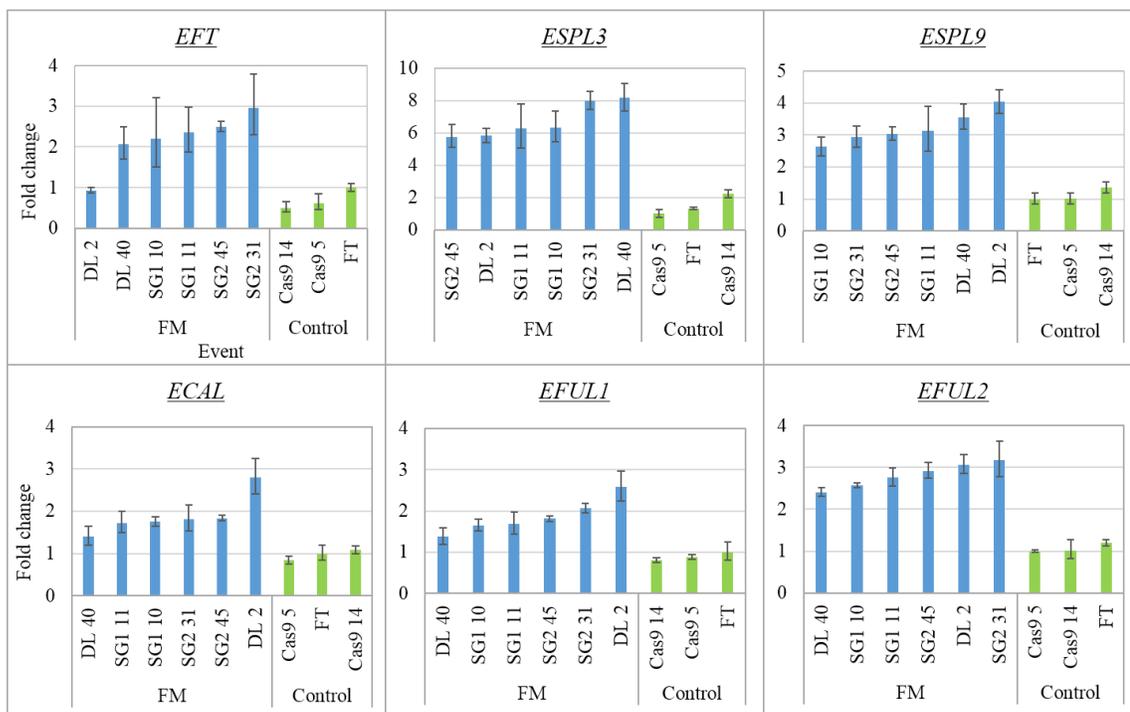


Fig. 4.8 Gene expression of floral development genes at the same level or upstream of *ELFY* in the flowering induction pathway for *ELFY*-FM and non-mutant flowering control events. The average fold change in expression was calculated as a ratio to the *AtFT*-only control, which was set to 1. Error bars represent \pm SE of means. Mean gene expression for FM vs. control groups was significantly different for all genes ($P < 0.006$ for all contrasts, two-tailed Student's t-test).

4.4 Discussion

All three vectors were highly (nearly 100%) efficient at inducing mutations on the endogenous target sites. Since the first CRISPR Cas9 gene editing studies published in tobacco, *Arabidopsis*, rice, and wheat (Li *et al.*, 2013; Nekrasov *et al.*, 2013; Shan *et al.*, 2013), the genomes of more than 25 plant species have been modified (Bewg *et al.*, 2018b; Xu *et al.*, 2019; Ghogare *et al.*, 2019). Mutation efficiencies of endogenous genes vary between 0.1% and 100%. Similar to our results, editing rates of 100% were seen in stably-transformed cassava (Odipio *et al.*, 2017), grapevine (Ren *et al.*, 2016), maize (Lee *et al.*, 2019), poplar (Zhou *et al.*, 2015; Wang *et al.*, 2017), rice (Shen *et al.*, 2017), tomato (Ueta *et al.*, 2017), and Wanjincheng orange (Peng *et al.*, 2017).

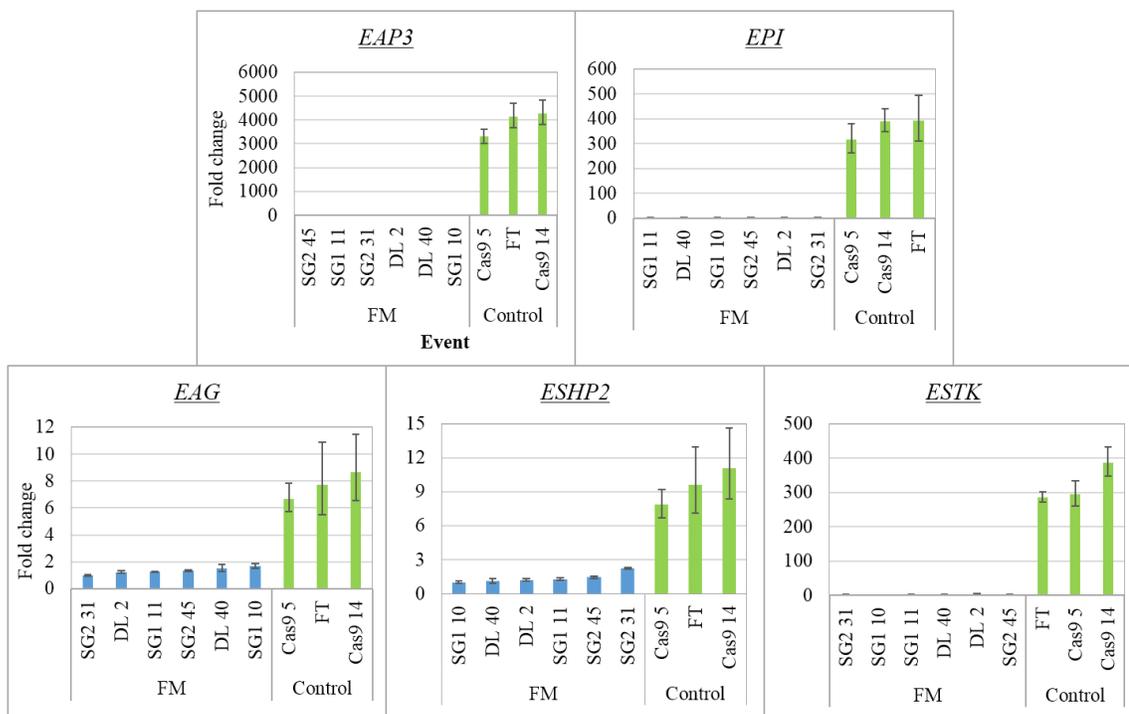


Fig. 4.9 Gene expression of organ identity genes downstream of *ELFY*. The average fold change in expression was calculated as a ratio to the lowest expressing event for each gene, which was set to 1. Error bars represent \pm SE of means. Gene expression was significantly different among FM and control groups for all the genes ($P < 0.02$ for all contrasts, two-tailed Student's *t*-test).

There were no consistent differences in any of the vegetative traits that suggested that the disruption of *ELFY* function had vegetative consequences. The only trait with any statistically significant differences among groups was leaf area in the non-flowering greenhouse plants, however, the FM group only differed from WT and not from the Cas9 transgenic (i.e., empty vector) or escape control groups. The WT group also tended to be larger in other traits, though not statistically significantly so. This suggests that the process of genetic modification and/or regeneration may have caused some impairment in vegetative development.

After a year and a half in the greenhouse, six predicted FM events did not transition to flowering, while the remaining 32 FM events produced sterile “flowers”. In general, if both alleles were mutated and at least one of the two mutated alleles had a frameshift mutation, the events would mostly have organless flowers (Table S5). Meanwhile if both alleles had mutations that left the C-terminals intact (e.g., in-frame inversions or in-frame

large deletions), the plants mostly had flowers with sterile reproductive organs that appeared early on in development (i.e., early organs in Supporting Information Table S4.4). Nonetheless, there was a category in between organless and early organs, “late organs”, where plants would have flowers that in their third to fifth layered pedicel would have sterile female reproductive organs (i.e., stigma-like and ovule-like but non-functional organs). In this “late organs” category, there were two events, 30-6 and 4-8, that were missing between one and three highly conserved amino acids. In 30-6, the *E. grandis* allele had a 6 bp deletion that removed a glutamic acid (E22; Supporting Information Fig. S4.9) and an alanine (A23; Supporting Information Fig. S4.9) and the *E. urophylla* allele had a 3 bp deletion that replaced the glutamic acid by an aspartic acid (E22D) and removed the same alanine (A23). In 4-8, the *E. grandis* allele had a 9 bp deletion that removed a phenylalanine, a glutamic acid, and an alanine (F21, E22, A23; Supporting Information Fig. S4.9), and the *E. urophylla* allele had a 1 bp insertion that modified the peptides completely and introduced a stop codon at the 91st AA position. The phenylalanine and alanine sites are highly conserved among all plant species and eudicots, respectively, thus they are likely essential to the interaction in *ELFY* dimers. These two events had a flowering phenotype similar to many events with frameshifts in both alleles that completely disturbed the peptide sequence.

We believe that part of the differences in floral phenotypes among our FM plants was due to partial *ELFY* function in the events with intact C-terminals compared to FM plants with completely disturbed *ELFY* alleles. All *LFY* homologs in the plant kingdom have two conserved domains; an N-terminal dimerization domain (Siriwardana & Lamb, 2012; Sayou *et al.*, 2016) and a C-terminal DNA binding domain (Hamès *et al.*, 2008). Siriwardana and Lamb (2012) found that *LFY* alleles with L/A or L69AL76A modifications in the N-terminal domain were enough to completely eliminate *LFY* function *in planta*. Their allele with only the L69A modification partially complemented the strong *lfy-6* mutant and had reduced activity compared to WT. Their alleles with *LFY*_{L/A} or *LFY*_{L69AL76A} modifications could not complement *lfy-6* and the plants produced sterile flowers with sepal-like and ovule-like organs after bolting. Last, their *LFY* allele that had the entire N-terminal removed could not complement *lfy-6* either and

the plants produced no reproductive organs. In similar experiments, Sayou and colleagues (2016) found that monomers of GbLFY_{T75E}, GbLFY_{R112E}, and GbLFY_{T75ER112E} (Gb, *Ginkgo biloba*) were able to partially complement the strong mutant *lfy-12* and to weakly bind to an AP1 DNA probe in EMSA (Sayou et al. 2016). However, they also learned that eliminating the entire N-terminal significantly disrupted LFY's DNA binding ability across the genome, in particular in sites of low-binding affinity (Sayou *et al.*, 2016). Thus, we hypothesize that removing highly conserved AAs in the N-terminal or removing the N-terminal domain completely, eliminated *ELFY*'s oligomerizing ability, thus rendering the flowers sterile. However, the remaining C-terminal protein may have been able to weakly bind some of *ELFY*'s DNA targets inducing the creation of some reproductive-like organs.

We examined gene expression upstream, near, and downstream of *ELFY* in the flowering pathway to help understand the developmental stage of the organless FM plants. Two patterns were seen in the gene expression analysis of floral genes. For *ELFY* and six genes upstream or at the same developmental stage as *ELFY* (i.e., *EFT*, *ESPL3*, *ESPL9*, *ECAL*, *EFUL1*, and *EFUL2*; Fig. 4.7 and Fig. 4.8), expression was significantly higher in the FM events than in the control events (*AtFT*-only and *AtFT* Cas9). We did not include the *ELFY* expression of FM event 30-11 in the plot (Fig. 4.7c) because its expression was 596X higher than that of the *AtFT* controls making the general pattern difficult to discern. *EFT* is a floral pathway integrator (FPI) gene similar to *ELFY*. *ECAL*, *EFUL1*, and *EFUL2* are floral meristem identity (FMI) genes just as *ELFY*. We selected *EFUL1* and *EFUL2* because there is no archetypical *APETALA1* (*API*) homolog in *Eucalyptus* (Vining *et al.*, 2015b). *API* and *FUL* are homologous genes created from a gene duplication predating the diversification of the eudicots. *CAL* is also a homolog of *API* and *FUL* that is believed to have arisen from *API* during a more recent duplication. *API* and *FUL* are not functionally equivalent. They can only partially rescue each other in *Arabidopsis* (McCarthy *et al.*, 2015). It is possible that one of the genes that has been identified as a *FUL* homolog (i.e., *EFUL1* or *EFUL2*) actually functions as an *API* homolog in *Eucalyptus*. However, we do not hypothesize which gene it could be, because their expression is similar and because they both have the *FUL*-like C-terminal

motif (i.e., LPAWML), which is missing in all the AP1 homologs (McCarthy *et al.*, 2015). Last, the *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL)* genes are essential for induction of flowering. *SPL3/4/5* are only essential to the transition to flowering when they assist the FT-FD complex in the activation of the FMI genes, *LFY*, *API*, and *FUL*, by directly binding to their promoter regions (Jung *et al.* 2016; Yamaguchi *et al.* 2009). Yamaguchi *et al.* (2014) hypothesized that *SPL9* recruits DELLA proteins to directly induce expression of *API* during transition from inflorescence meristem to flower meristem. During flowering *LFY* activates many floral organ identity (FOI) genes including *APETALA1 (API)*, which then itself induces more *LFY* expression, generating a feed-forward loop for controlling flowering (Gramzow & Theissen, 2010; Liu & Mara, 2010). With a non-functioning *ELFY*, the feed-forward loop cannot keep on cycling and increasing expression, causing flowering to be arrested in the inflorescence specification stage.

Additionally, for five genes directly or indirectly regulated by *ELFY* (i.e., *EAP3*, *EPI*, *EAG*, *ESHP2*, and *ESTK*; Fig. 4.9), expression was significantly lower in the FM events than in the controls. *EAP3*, *EPI*, *EAG*, *ESHP2*, and *ESTK* regulate expression of genes that make floral organs (reviewed in Pajoro *et al.* 2014). *EAP3* and *EPI* are B-class genes, *EAG* is the C-class gene, and *ESHP2* and *ESTK* are D-class genes of the ABCDE model of flower development. This model has been thoroughly studied in *Arabidopsis*, *Antirrhinum*, petunia, and tomato (reviewed in Causier, Schwarz-Sommer, and Davies 2010; Immink, Kaufmann, and Angenent 2010; Ó'Maoiléidigh, Graciet, and Wellmer 2014; Pajoro *et al.* 2014; Rijpkema *et al.* 2010). *ELFY* directly regulates expression of *EAP3*, *EPI*, and *EAG*, and indirectly of *ESHP2* and *ESTK*.

In *Arabidopsis*, *AG* regulates the formation of stamens (with the B-class genes, *AP3* and *PI*) and carpels, and its expression is essential for floral determinacy (Bowman *et al.*, 1989b; Yanofsky *et al.*, 1990; Mizukami *et al.*, 1996). Flowers become determinate when *AG* indirectly represses the stem cell maintenance gene *WUSCHEL (WUS)* (Sun *et al.*, 2009; Liu *et al.*, 2011). Our FM events had significantly lower expression of *EAG*, which may have been the reason for the repeated pedicel-like and bract-like structures, and thus the reduction in floral determinacy.

Because of the two gene expression patterns —where genes expressed upstream or at the same physiological level as *ELFY* had higher expression in FM events than in flowering control events, and genes expressed downstream or regulated directly by *ELFY* had lower expression in the FM events than in the flowering control events —one possible explanation is that the FM events were developmentally trapped in inflorescence development, but before flower development. The defective *ELFY* protein was stalling the process of flower development, causing our constantly-expressing *AtFT* FM events to not develop fertile flowers. Unfortunately, obtaining phenotypic data for *elfy* mutant flowers in a WT background was beyond the scope of our study as non-*AtFT* eucalypt trees typically do not flower for several years after regeneration.

CRISPR Cas9 nucleases appear to provide an efficient method for elimination of *ELFY* function, and thus a means for preventing both male and female sexual reproduction without adverse vegetative impacts. It is also expected to be highly stable over the long lifespans of trees in the field, especially when compared to previous methods for sterility induction such as the use of cytotoxins or gene suppression, whose efficacy can vary with environmental and developmental perturbations (Brunner *et al.*, 2007; Vining *et al.*, 2012). If sterility persists in the field under natural flowering, it should enable greater acceptance and regulatory approval of exotic or genetically engineered varieties, and thus speed the delivery of improved traits such as pest and disease resistance, modified wood properties, and biomass productivity (Chang *et al.* 2018).

Acknowledgements

We thank Dr. Jian-Kang Zhu at Purdue University and Dr. Yanfei Mao at Shanghai Center for Plant Stress Biology for providing us with the vectors AtU6-26SK and 35S-Cas9-SK. We thank many undergraduate students for their help, including Kathryn Dow, Analeslie Martinez, and Melissa Meyhoff for helping with media preparation, plant transformation, plant selection, and greenhouse operations. We thank FuturaGene for providing us with the SP7 hybrid *Eucalyptus urograndis* genotype. We also thank the members of the Tree Biosafety and Genomics Research Cooperative (TBGRC) at OSU, the United States Department of Agriculture (award # 2011-68005-30407, System For Advanced Biofuels Production From Woody Biomass In The Pacific Northwest), the USDA Biotechnology Risk Assessment (award # 2010-33522-21736), and the NSF I/UCRC Center for Advanced Forestry (award # 0736283). Last, we would like to thank Oregon State University's microCT facility and the Major Research Instrumentation Program of NSF's Earth Sciences (EAR) division (award # 1531316).

5 Flower to seed capsule transcriptome dynamics in *Eucalyptus grandis*

Contributions of authors

Estefania Elorriaga designed the study with the help of Amy L. Klocko and Steven H. Strauss. Estefania Elorriaga isolated the RNA, performed the bioinformatics analyses, and wrote the manuscript. Steven H. Strauss with help from Amy L. Klocko supervised the overall study.

Abstract

The genus *Eucalyptus* includes some of the most important and ecologically dominant forest trees on the planet because of their great diversity, adaptation to harsh environments, and fast growth under subtropical plantation conditions. In an effort to better understand the physiology of *Eucalyptus* flower and fruit development, we sequenced RNAs from seven tissues including flowers, capsules, pollen, and fully expanded leaves. We identified 11,438 genes that were differentially regulated when all the reproductive tissues were compared to leaves, 6,107 genes when the reproductive stages were compared to each other, and 3,483 genes when only the flower stages were compared. The genes in pollen, all the reproductive tissues, and the flower-only tissues represented 384, 612, and 394 gene ontology categories respectively. As expected genes known to be involved in flowering, seed development, lignin biosynthesis, pathogen defense, and nectar production were upregulated compared to leaves. We also found genes involved in nitrogen allocation and nitrogen starvation to be upregulated in mature pollen and cell morphogenesis and gravitropism genes to be upregulated in early late flower and developing capsules. Cell cycle genes went down as reproductive development proceeded. 7,848 (45%) of the differentially regulated genes had no functional annotation. Our results provide an atlas of reproductive development in *Eucalyptus* that will help inform both studies of eucalypt development and biotechnologies seeking to modify reproduction and other traits.

5.1 Introduction

The Myrtaceae family is composed of more than 5,500 species and 142 genera, found mostly in the tropics and Southern Hemisphere (Wilson, 2011). This family is taxonomically highly diverse, especially in their reproductive and seed dispersal strategies. For example, they have flowers with sepals and petals, flowers without any sepal or petals, or flowers with just sepals or just petals. Their flowers are pollinated by animals, including insects, birds, and small mammals. They have dry (i.e., nuts and capsules) and fleshy (i.e., drupes and berries) fruits dispersed by wind and animals (Wilson, 2011). It appears that dry fruits are more ancestral, and that the succulent pericarp evolved two or three times (Biffin *et al.*, 2010; Giaretta *et al.*, 2019).

The genus *Eucalyptus* is among the most researched genera in the Myrtaceae family. It has over 800 species, endemic to Australia, New Guinea, Indonesia, and the Philippines (Wilson, 2011). *Eucalyptus* is the most widely planted genus of broadleaf plants (CIRAD-FRA *et al.*, 2018), and *E. globulus* is the most widely cultivated species in the genus (Wilson, 2011). However, *E. grandis* is the only species with substantial genomic and transcriptomic data available (Myburg *et al.*, 2014; Vining *et al.*, 2015a).

Eucalyptus is fast-growing, drought tolerant, and able to grow on poor soils.

Commercially, it is valued for its fast-growth, phytotherapeutics (Goldbeck *et al.*, 2014; Luís *et al.*, 2016), and ecological support of other biota including bees, birds, and small mammals. Recently with the advent of modern breeding tools (i.e., genomic selection, genetic engineering and gene editing), *Eucalyptus* is also considered a great candidate for bioenergy (Taylor *et al.*, 2016).

Eucalyptus has umbels as inflorescences. An inflorescence is a compound structure with more than one flower. All the flowers in an umbel have pedicels (i.e., the stem attached directly to the flower) of the same length that originate at the same spot on the peduncle (i.e., stem that attaches the umbel to branch). The term was coined in the late 1500's because this inflorescence looks like an "umbrella" (i.e., "umbella" is Latin for parasol or sunshade). The umbels can be terminal or axillary, with anywhere between 1 and 30

flowers, but with mostly seven flowers (Wilson, 2011). The umbels are also enclosed by several bracts and bracteoles (Carr & Carr, 1959).

Flowers are ornate organs that contain reproductive tissues. Their ornateness is for attracting pollinators. *Eucalyptus* flowers are visited by local bees, honeybees, flies, beetles, birds, and sometimes even mammals (Armstrong, 1979; Ford *et al.*, 1979; Regal, 1982). However, many of these visitors are not effective at pollinating, but they are benefiting from the pollen or nectar. Most *Eucalyptus* species are effectively pollinated by insects, but some are pollinated by birds (Hingston *et al.*, 2004b,c,a; Griffin *et al.*, 2009).

The flowers in *Eucalyptus* are hermaphroditic and do not have separate sepals or petals. Instead, they have a fused calyx and fused corolla, resulting in two opercula, the calycine operculum and the corolline operculum. The corolline operculum remains right up to anthesis, whereas the calycine operculum is shed early in flower development. The stamens are found in several continuous whorls and are the most conspicuous feature of the flowers. Flowers in the *Eucalyptus* genus can have up to 300 stamens (Moncur & Boland, 1989). The stamens dehisce shortly after anthesis.

Woody capsules are an adaptation to the extreme environment in Australia including frequent fires and droughts (Gill *et al.*, 1992). During a period of five to six months, each pollinated flower metamorphoses into a wood capsule after which the locules open and the seeds are released. The changes include dehiscence of the stamens, senescence of the style and stigma, enlarging of the ovary, and desiccation and lignification of the capsule.

In this study, we sequenced the RNA from seven tissues including flowers, seed capsules, mature pollen, and fully-expanded leaves. We aim to supplement the knowledge gained from the transcriptome of developing flower buds in *Eucalyptus grandis* (*E. grandis*) (Vining *et al.*, 2015b) by completing the picture of *Eucalyptus*' flower and seed capsule development. In addition to reporting large numbers of differentially regulated genes, we tested several hypotheses related to gene expression changes during floral development. For example, given the changes that the flowers undergo to become woody capsules, we suspected that in addition to the expected embryo and seed development genes, the lignin

biosynthesis pathway will be upregulated during between early flower and late capsule stages. Based on Rutley & Twell (2015), we expected pollen to be enriched in membrane transport, vesicle trafficking, and signaling. Also, we expect to find genes involved in perianth senescence and flower longevity when we compare the bagged flowers to the unconfined flowers based on the literature (van Doorn, 1997; van Doorn & Woltering, 2008; Jibrán *et al.*, 2017). Given *Eucalyptus*' unique flower and seed capsules, we expect to find a number of genes that are not expressed in flowers or fruits in *Arabidopsis* or other model plants present. Lastly, this omic resource should inform and improve the annotated *Eucalyptus grandis* genome v2.

5.2 Materials and Methods

5.2.1 Plant material

Mature pollen, bagged flowers, unconfined flowers, seed capsules, and leaves were sampled from three *E. grandis* trees of clone T1099. The trees were in a single commercial plantation belonging to SAPPI in Howick, South Africa (S 29° 28' 47", E 30° 10' 52"). Starting the week of April 18, 2016 (i.e., week zero), the three trees were sampled on week zero, week one, week four, and week 12. The hypanthium of flowers was marked with silver nail polish if they were to be harvested the following week. Controlled pollination (CP) bags to allow normal air/moisture exchange, were placed over ripe unopened flowers (they were developmentally about a week from the time of abundant anthesis in the seed orchard planting). The CP bags were created in-house with bamboo fiber and were molded with metal wire frames. Bamboo was chosen because of its high breathability and antibacterial properties (Yao & Zhang, 2011; Imadi *et al.*, 2014). The flowers were harvested from the bags two weeks after bagging and were labeled as "late flower bagged" (LFB). Open pollinated flowers at anthesis were sampled on week zero and labeled "early flower" (EF). Open pollinated flowers were harvested a week after anthesis (i.e., week one) and labeled "late flower unconfined" (LFU). Developing seed capsules were harvested from flowers a month and three months after anthesis (i.e., on week four and week 12) and labeled "early capsule" (EC) and "late capsule" (LC) respectively. Fully opened leaves were sampled from the three trees at

week zero (labeled “L”). Mature pollen was harvested from flowers during the first three weeks (labeled “P”). We had three biological replicates for each tissue except for pollen, for which we only had two. All samples were immediately frozen with liquid nitrogen on site, stored at -80C between two and eight months, then shipped to Corvallis, OR with dry ice. Dry ice was added to the shipping container twice during the two-day transit.

5.2.2 RNA-Seq library preparation and sequencing

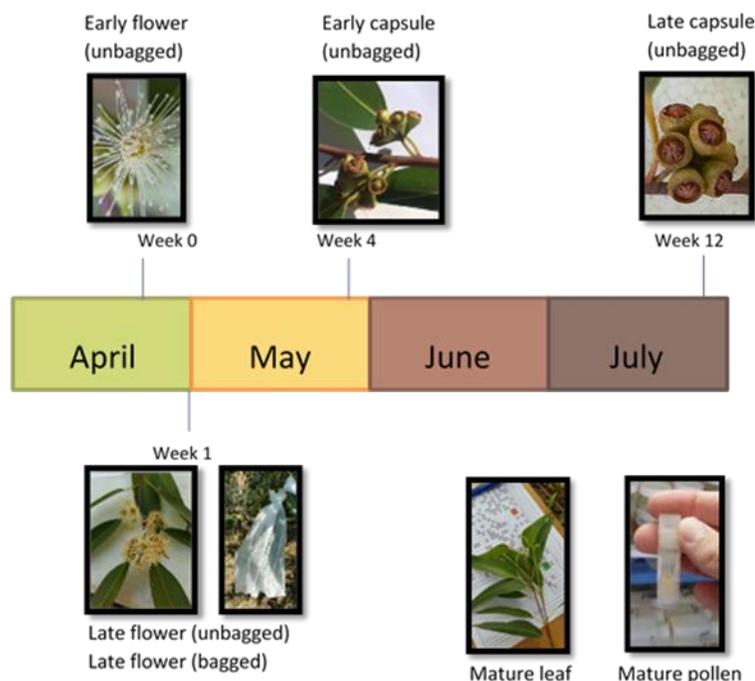


Fig. 5.1 Timeline of tissue sampling in the South African plantation. Trees begin flowering between February and March. Flower buds develop and begin opening in mid to late April. Seed capsules mature between May and October. Seed capsules shed seed between September and October.

About one gram of tissue was used for each RNA isolation. All tissues were first ground with a coffee grinder in the presence of dry ice. Ground tissue was inserted into 50 ml Falcon tubes with 15ml of RNA isolation buffer. Total RNA was extracted according to Howe et al. (2013) and treated with DNaseI (New England Biolabs) to remove any residual genomic DNA. The integrity of the DNase-treated RNAs was determined using the Agilent Bioanalyzer 2100 at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University (Corvallis, OR, USA). Library construction and

sequencing was done at the CGRB. For each sample, 1 μg of high integrity (i.e., with a RIN value of 7 or higher) total RNA was first subjected to poly(A) enrichment using the PrepX poly(A) mRNA isolation kit (WaferGen Bio-Systems Inc, Fremont, CA, USA). Then, cDNA libraries were created by a WaferGen robotic system for each poly(A)-enriched RNA sample using the PrepX RNA-seq Library Kit for Illumina (WaferGen Bio-Systems Inc). All the libraries were multiplex sequenced on 2.25 lanes of an Illumina HiSeq 3000 instrument as single-end 150 bp runs.

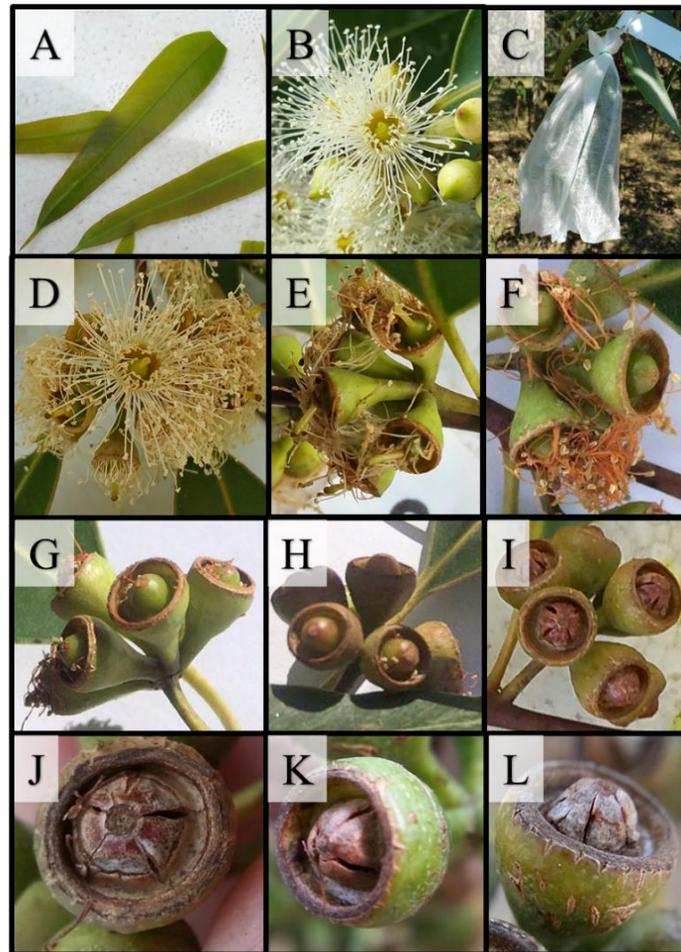


Fig. 5.2 *Eucalyptus grandis* mature leaves and flower to capsule sequence. (A) Mature leaves with petioles removed prior to RNA extraction. (B) Flower opened at anthesis (i.e., “EF”). (C) Flowers covered with a CP bag (i.e., “LFB”). (D) Open pollinated flowers one week (i.e., “LFU”) (E) two weeks, (F) three weeks, (G) four weeks (i.e., “EC”), (H) eight weeks, (I) twelve weeks (i.e., “LC”), (J) sixteen weeks, (K) twenty weeks, and (L) twenty four weeks after anthesis (WAA).

5.2.3 Evaluation of RNA-seq data

We mapped, assembled, and counted our transcripts using the HISAT2-StringTie-Ballgown pipeline (Frazee *et al.*, 2015; Pertea *et al.*, 2015, 2016; Kim *et al.*, 2019, p. 2; Kovaka *et al.*, 2019). We used the *E. grandis* genome v2.0 (Myburg *et al.*, 2014; downloaded from Phytozome v12.1) as the reference genome. The count tables generated by Ballgown (Frazee *et al.*, 2015) were imported into the RStudio environment (RStudio Team, 2015) where we used R version 3.6.1 (R Core Team, 2017) to analyze the data and the package ggplot2 (Wickham, 2009) to produce the figures. To assess overall similarity among our different tissues and samples, we created a Poisson distance plot (Fig. S1) and principal component analysis (PCA) plots (Fig. S2 and S3). We used a Poisson distance plot instead of a Euclidean distance plot because RNA-seq data generally follows a Poisson or negative binomial distribution (Witten, 2011). We calculated the Poisson dissimilarity matrix from the original counts data using the R package PoiClaClu (Witten, 2011). We executed PCA using the plotPCA function from the R package DESeq2 (Love *et al.*, 2014).

5.2.4 Identification of DEGs and GO categories enriched in each tissue compared to mature leaf

We used DESeq2 (Love *et al.*, 2014) to conduct negative binomial Wald tests for differential gene expression analysis between the flower/fruit libraries (i.e., EF, LFB, LFU, EC, LC, and P) versus the leaf library (i.e., L), and to examine gene expression changes during reproductive development also between LFU vs. EF, LFU vs. LFB, EC vs. LFU, and LC vs. EC. We used the lfcshrink function with the approximate posterior estimation for the generalized linear model option (apeglm) (Zhu *et al.*, 2019) to reduce the variance among genes with low expression or highly variable levels of expression among biological replicates. We filtered and created tables of differentially expressed genes (DEGs) using a p-value adjusted cutoff (i.e., P-values adjusted based on the Benjamin-Hochberg correction) of 0.05, and a log fold-change (LFC) cutoff of 1 or higher (i.e., 2X or higher) for upregulated genes and of -1 or lower for downregulated genes (i.e., -2X or lower). We identified genes unique to each tissue and genes common

to two or more tissues using the function `overLapper` from the `systemPipeR` R package (Backman & Girke, 2016). We also used this function to create two five-set Venn diagrams; one of upregulated genes (Fig. S4) and one of downregulated genes (Fig. S5). Homologs from *Arabidopsis* of the 20 most upregulated genes based on the log fold-change in each contrast were identified using `Phytomine` from `Phytozome v12` (Goodstein *et al.*, 2012). DEGs were subjected to GO enrichment analysis using the online GO database `AgriGOv2` (Tian *et al.*, 2017) and/or soft fuzzy c-means cluster analysis in R using `mfuzz` (Futschik & Carlisle, 2005; Kumar & E. Futschik, 2007). Further analyses (i.e., identification of protein domain enrichment, pathway enrichment, and orthologs in *Arabidopsis*) of specific GO enriched terms was implemented using `Phytomine` from `Phytozome v12` (Goodstein *et al.*, 2012).

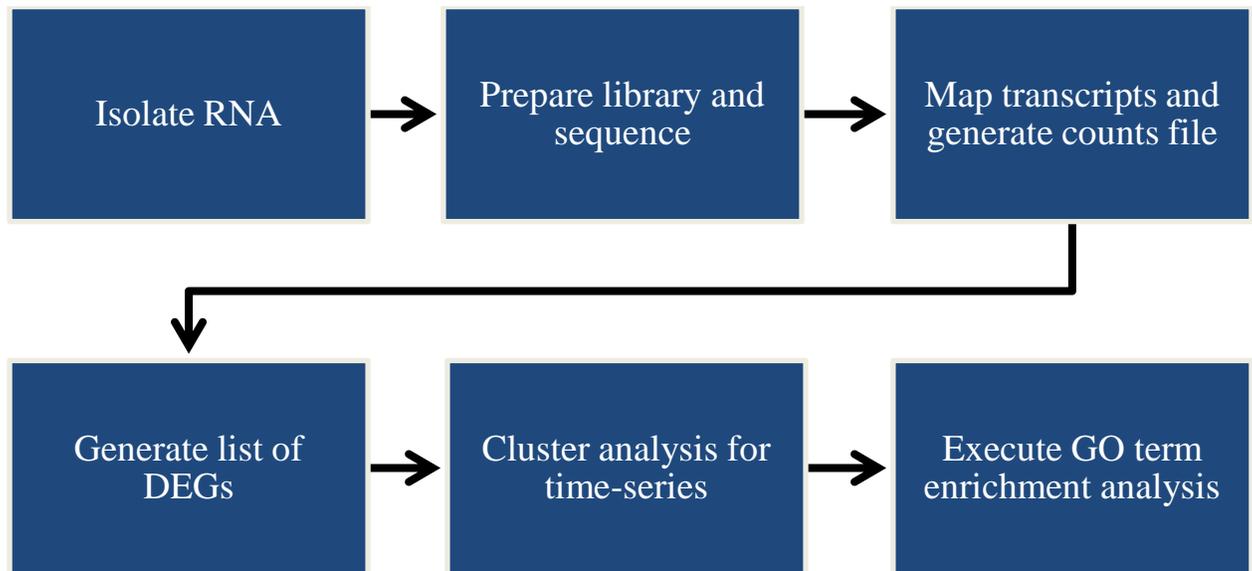


Fig. 5.3 Summary of methods for differential expression and gene ontology (GO) analysis. The transcripts were mapped and assembled to the *E. grandis* genome using `HISAT2` and `StringTie`. `Ballgown` was used to generate the counts file. Lists of differentially expressed genes (DEG) were generated in R using `DESeq2`. Fuzzy c-means clustering was done using the R program `mfuzz`. GO enrichment analyses were done with the online GO database `AgriGOv2`.

5.2.5 Expression of homologs in *Eucalyptus* to genes involved in lignin biosynthesis in *Arabidopsis*

We identified the genes and gene families characterized in the metabolism of phenylpropanoids including lignin in *Arabidopsis* (Fraser & Chapple, 2011). We created a heatmap in Microsoft Excell for all of the homologs in *Eucalyptus* (Table 5.5). Homologs of all the enzymes needed to generate lignin from chorismate all included in the heatmap

5.2.6 Gene expression validation using real-time quantitative PCR

We performed real-time quantitative PCR (qPCR) analysis to support the transcript abundance seen in our DESeq2-normalized count file. We synthesized cDNA from our DNase-treated RNAs with the Applied Biosystems™ High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor following the manufacturer's instructions (ThermoFisher Scientific). All reactions were executed by a StepOnePlus Real-Time PCR system (ThermoFisher Scientific). For each 20 µl reaction, we mixed 10 µl of PowerUp SYBR Green Master Mix (ThermoFisher Scientific), 1.0 µl (10 ng) of cDNA, 1.2 µl of forward and reverse primers, and 7.8 µl of water. The thermocycler conditions were: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 60 s. Gene amplification was followed by melt-curve analysis, which was done by increasing the temperature by 0.3°C s⁻¹ between 60 and 95°C. We selected the following seven genes, corresponding to the *Eucalyptus* orthologs of: *AGAMOUS* (*AG*), *CAULIFLOWER* (*CAL*), *FLOWERING LOCUS T* (*FT*), *FRUITFULL* (*FUL*, there are two in *Eucalyptus*), *SHATTERPROOF 2* (*SHP2*), and *SEEDSTICK* (*STK*). The delta-delta-Ct (ddCt) method was used to determine the relative gene expression of each gene. All reactions were done in duplicate. The *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GAPDH*) ortholog was selected as the housekeeping gene. Primer design, specificity, and efficiency were previously determined (Elorriaga et al. 2020, in review). The log₂ fold change (FC) determined in the qPCR experiments was compared side-by-side with the log₂ FC determined during differential expression of the RNA-Seq counts. The log₂ fold changes were also plotted against each other and the Pearson correlation coefficient was calculated.

5.3 Results

5.3.1 Overview of RNA-Seq data

We generated 20 RNA-Seq libraries corresponding to seven tissues, with all but pollen having three biological replicates. On average, we obtained between 41 and 49 million raw reads per tissue, of which, on average, 63 to 81% aligned one or more times to the *E. grandis* v2.0 genome. Out of the 36,349 protein coding loci in *E. grandis* v2.0 genome, we identified 27,450 unique transcripts among all the libraries. The different libraries all had more than 20,000 expressed gene models identified that had more than 10 counts in all replicates after DESeq2 normalization (EF:23,902, LFB:23,850, LFU:24,459, EC:24,734, LC:24,063, P:21,656, and L:23,731).

5.3.2 Distance between samples

The Poisson distance matrix (Fig. S5.1) illustrates that all replicates are most similar to their own tissues. The pollen libraries are the most distinct from the rest of the libraries, with the leaf libraries the second most distinct. The flower and seed capsule libraries are the most similar to each other with the LFB1 library clustering mostly with the EF1 library. The principal component analysis (PCA) resulted in similar clustering. The PCA plot with the first and second components, which explained 71 and 15% of the variance respectively, showed the three leaf libraries, the two pollen libraries, and the rest of the libraries clustered into three groups (Fig. S2). After removing the leaf and pollen libraries from the data, the replicates from each flowering tissue (i.e., EF, LFB, LFU, EC, and LC) clustered together except for the EF1 (i.e., “early flower” from tree one) which clustered with the LFB replicates.

5.3.3 Identification of differentially expressed genes (DEGs) and GO enriched categories

A total of 11,438 unique DEGs were identified from the contrasts of reproductive tissues to leaf (specifically: EF: 4,464, LFB: 4,165, LFU: 4,335, EC: 4,725, LC: 5,043, P: 7,958; all are from contrast vs. L). Contrasts were also made between EF and LFU (2,615 DEGs: 1,987 upregulated and 628 downregulated), LFB and LFU (2,136 DEGs: 1,930 upregulated and 446 downregulated), EC and LFU (1,902 DEGs: 1,015 upregulated and

887 downregulated), and LC and EC (3,870 DEGs: 1,313 upregulated and 2,557 downregulated) to identify genes active during floral senescence, carpel/capsule wall development, embryo/seed development, and capsule lignification. We, then, submitted the list of genes of P vs L to AgriGO for gene ontology (GO) enrichment analysis. The other DEGs were used in fuzzy c-means clustering first, to identify groups of genes among the reproductive tissues with similar expression patterns throughout development. Last, each list of genes corresponding to an expression pattern throughout development was subjected to GO enrichment analysis.

Table 5.1 Homologs in *Arabidopsis* of genes upregulated in flowers, capsules, and pollen when expression in compared to mature leaf.

Locus in <i>Eucalyptus</i>	Gene in <i>Arabidopsis</i>	Function	Tissue with overexpression
Eucgr.E00162	<i>ADF10</i>	Actin filament depolymerization factor	P
Eucgr.D01152	<i>CER1</i>	Aldehyde decarboxylase	LFU, EC, LC
Eucgr.K02547	<i>FUL</i>	MADS-box gene	All except P
Eucgr.L01734	<i>LAC4</i>	Laccase in lignin biosynthesis	LC
Eucgr.J02217	<i>MYB21</i>	Jasmonate response during stamen development	EF, LFB, LFU
Eucgr.D01819	<i>MYB103</i>	Regulation of lignin biosynthesis	LC
Eucgr.I01300	<i>PAP26</i>	Acclimation to Pi deprivation	P
Eucgr.E00014	<i>PPa1</i>	Pyrophosphatase	P
Eucgr.I02058	<i>SEP4</i>	MADS box transcription factor	All
Eucgr.D01671	<i>SND1</i>	Secondary wall biosynthesis	LC
Eucgr.F02981	<i>STK</i>	Seed development	LFU, EC, LC
Eucgr.H04154	<i>SWEET9</i>	Sucrose transporter for nectar secretion	EF, LFB, LFU

5.3.4 Orthologs in *Arabidopsis* of highest expressing genes in contrasts with leaf

We used Phytomine to identify homologs in *Arabidopsis* of the top 20 most upregulated genes in every tissue compared to leaf, and that were differentially regulated compared to leaf (Tables S5.1 through S5.6). Of each group of 20 genes, we identified well-characterized genes in *Arabidopsis* for further exploration (Table 5.1). We created

scatterplots of most of the genes in Table 5.1 that show the counts for each sample (Fig. 5.2, 5.3, and 5.4).

5.3.5 Clustering of DEGs based on expression

We filtered out 6,107 unique genes among all the upregulated (i.e., double the expression or higher) and downregulated (i.e., half the expression or lower) DEGs identified from the contrasts of flowers and developing capsules (excluding late flower bagged) (referred from here on as “seed capsule development cluster analysis”). After performing clustering analysis, the 6,107 genes were grouped into four clusters using the R program mfuzz (Futschik & Carlisle, 2005; Kumar & E. Futschik, 2007) (Fig. 5.5, 5.6, 5.7, and 5.8). Cluster one had 1,428 genes that were expressed higher in late capsule than in any other tissues. There were 977 genes in cluster two and they were expressed higher in early flower and decreased in expression along capsule development. Cluster three had 1,482 genes with highest expression in late flower unconfined. Cluster four had 2,220 genes with highest expression in early capsule.

We identified 3,482 unique genes among all the DEGs in early and late flowers (capsules excluded) when performing contrasts between each of them (referred from here on as “flowers-only cluster analysis”). The 3,482 genes were clustered into three groups (Fig. 5.9). Cluster one had 461 genes that were expressed higher in EF, lower in LFB, and lowest in LFU. Cluster two had 2,644 genes with high expression in LFU and low in EF and LFB. Cluster three had 377 genes with highest expression in LFB, lower in EF, and lowest in LFU.

5.3.6 Orthologs in *Arabidopsis* of highest pairwise differentially expressed genes in flowers and capsules

Among the hundreds to thousands of genes identified for each cluster (Fig. 5.5 through 5.8) in the seed capsule development cluster analysis, we selected the 20 with the highest membership score to each cluster (i.e., membership values of 0.999 or higher) and created tables with short descriptions (Tables S5.7, S5.8, S5.9, and S5.10). From those tables, we identified genes that also have been studied in depth in *Arabidopsis* (Table

5.2), excluding any from the table that did not have a clear *Arabidopsis* homolog or a clear function in *Arabidopsis*. In total, 17 of the 80 genes identified were selected for further investigation in our transcriptome.

Among the hundreds to thousands of genes identified for each cluster flowers-only cluster analysis (capsules excluded) (Fig. 5.9), we also generated lists of the top 20 genes with the highest membership scores (Tables S5.11, S5.12, S5.13, and S5.14). From those tables, we identified genes that also have been studied in depth in *Arabidopsis* (Table 5.3). In total, 11 of the 60 genes identified were selected for further investigation in our transcriptome.

5.3.7 GO enriched categories in flowers and capsules

We subjected each of the gene lists corresponding to the capsule development cluster analysis (excluding late flower bagged), and the flowers-only cluster analysis to gene ontology (GO) term enrichment analysis (using the online tool AgriGO version 2.0). We created a table for each GO term enrichment analysis result (Tables S5.14 through S5.20).

For cluster 1 in the capsule development cluster analysis (i.e., high expression in LC), we find the following among the GO significant terms (p-value based on chance): cellulose biosynthetic process, regulation of transcription, nicotamide nucleotide metabolic process, photosynthesis (light harvesting), glycerol ether metabolic process, cellular ion homeostasis, cellulose microfibril organization, and amino acid transmembrane transport. For cluster 2 (i.e., high expression in EF), we find: cell-redox process, terpenoid biosynthetic process, trehalose biosynthetic process, riboflavin biosynthetic process, sucros metabolic process, and L-phenylalanine metabolic process. For cluster 3 (i.e., high expression in LFU), we find: chitin catabolic process, proteolysis, cellular glucan metabolic process, DNA replication initiation, fatty acid biosynthetic process, sister chromatid segregation, defense response to fungus, and defense response to bacterium. For cluster 4 (i.e., high expression in EC), we find: microtubule-based movement, DNA-dependent DNA replication, signal transduction, cellulose biosynthetic process, cytokinesis, and potassium ion transport.

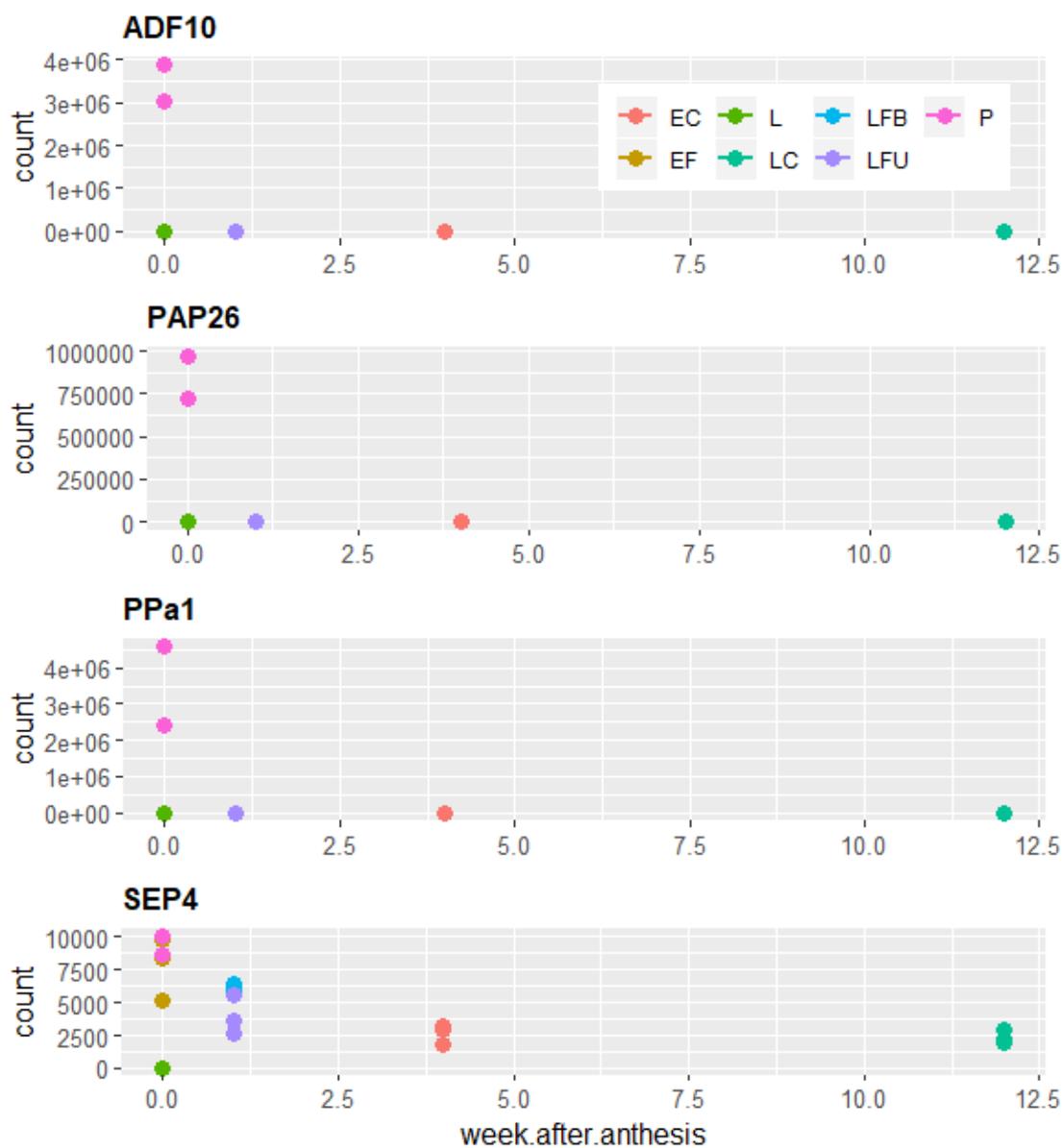


Fig. 5.4 Genes identified during differential expression analysis with high expression in pollen compared to leaf. *ADF10*, *PAP26*, and *PPa1* had high expression in P compared to L. *SEP4* had high expression in all tissues compared to L. The y-axis has the count values and x-axis corresponds to the development timeframe of our samples. The colors of the data points correspond to the different tissues. EC, early capsule; EF, early flower; L, mature leaf; LC, late capsule; LFB, late flower bagged; LFU, late flower unconfined; P, mature pollen.

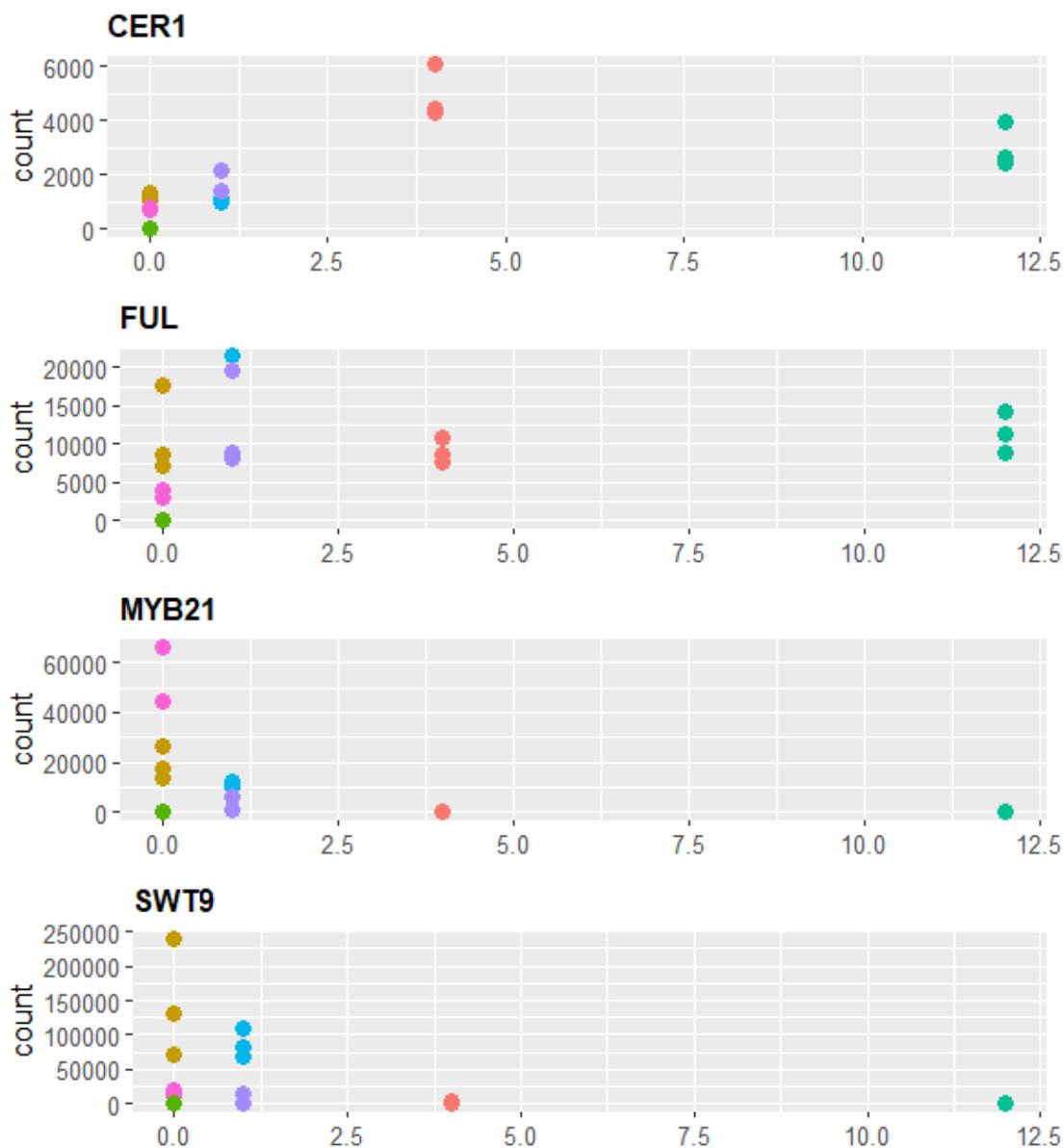


Fig. 5.5 Genes identified during differential expression analysis with high expression in flowers or early capsule compared to leaf. *CER1* had high expression in LFU, EC, and LC compared to L. *FUL* had high expression in EF, LFB, LFU, EC, and LC compared to L. *MYB21* and *SWT9* had high expression in EF, LFB, and LFU compared to L. The y-axis has the count values and x-axis corresponds to the development timeframe of our samples. The colors of the data points correspond to the different tissues. EC, early capsule; EF, early flower; L, mature leaf; LC, late capsule; LFB, late flower bagged; LFU, late flower unconfined; P, mature pollen.

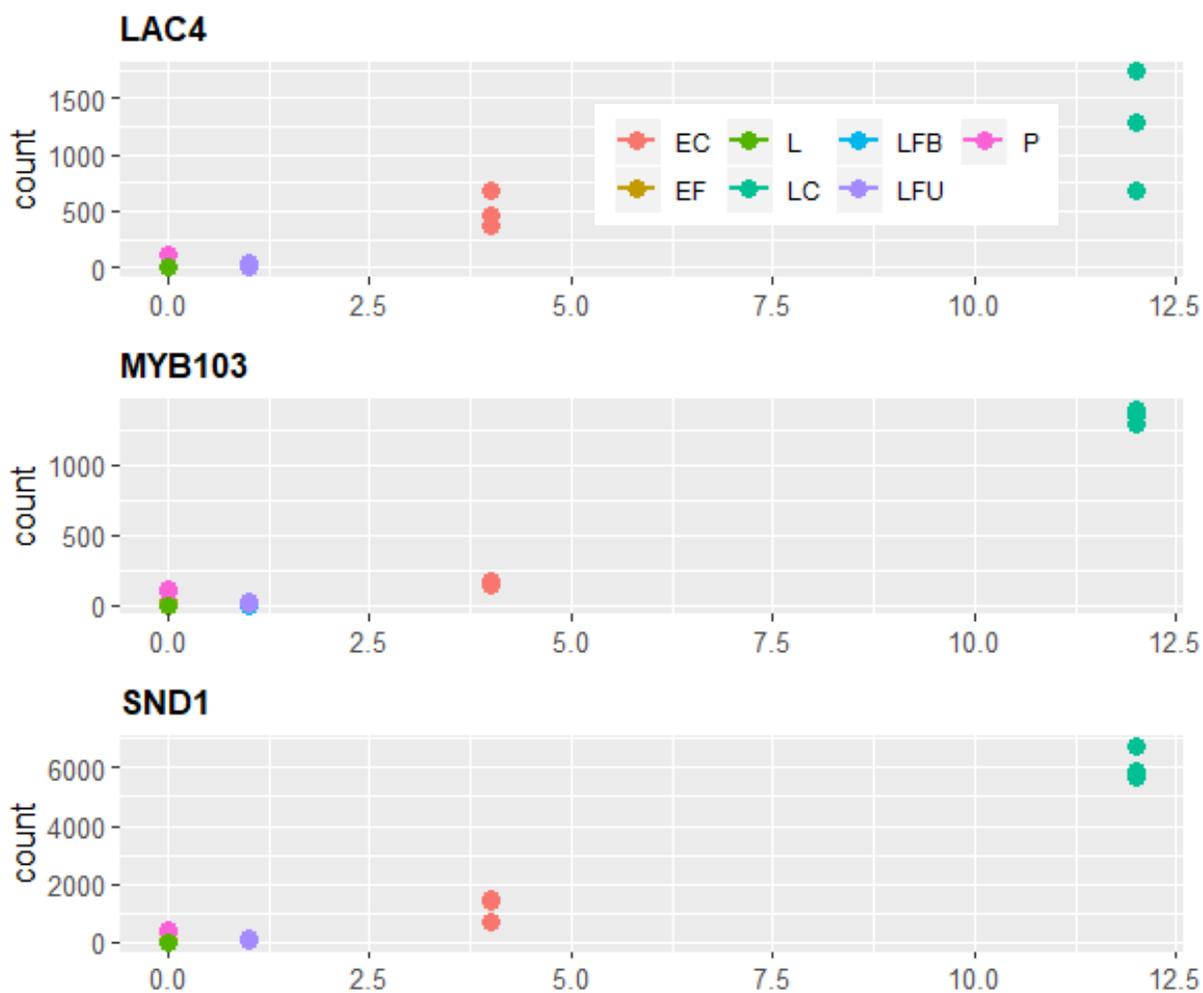


Fig. 5.6 Genes identified during differential expression analysis with high expression in late capsule compared to leaf. *LAC4*, *MYB103*, and *SND1* had high expression in LC compared to L. The y-axis has the count values and x-axis corresponds to the development timeframe of our samples. The colors of the data points correspond to the different tissues. EC, early capsule; EF, early flower; L, mature leaf; LC, late capsule; LFB, late flower bagged; LFU, late flower unconfined; P, mature pollen.

Table 5.2 Well-characterized homologs in *Arabidopsis* of upregulated genes in the seed capsule development cluster analysis.

Locus in <i>Eucalyptus</i>	Gene in <i>Arabidopsis</i>	Function	Cluster
Eucgr.C02284	<i>4CL1</i>	Phenylpropanoid pathway	1
Eucgr.F02557	<i>CSE</i>	Phenylpropanoid pathway	1
Eucgr.J01079	<i>PAL1</i>	Phenylpropanoid pathway	1
Eucgr.G03056	<i>PLC2</i>	Auxin-mediated reproductive development	1
Eucgr.H04617	<i>SEP3</i>	Flower development	1
Eucgr.C03853	<i>ALDH2C4</i>	Phenylpropanoid pathway	2
Eucgr.G02223	<i>CAD9</i>	Cinnamyl alcohol dehydrogenase	2
Eucgr.K00311	<i>LYS1</i>	Pathogen resistance	2
Eucgr.I00449	<i>UGT73B2</i>	Pathogen resistance	2
Eucgr.I02438	<i>ERDL6</i>	Vacuolar glucose exporter	3
Eucgr.A02311	<i>GPDHC1</i>	Glycerol metabolism	3
Eucgr.J01662	<i>RBOHD</i>	ROS production	3
Eucgr.I01041	<i>SNG1</i>	Sinapoylmalate synthesis	3
Eucgr.J00581	<i>IQD5</i>	Pavement cell morphogenesis	4
Eucgr.I00565	<i>RBK1</i>	Auxin-specific cell expansion	4
Eucgr.E00461	<i>SKU5</i>	Root size and gravitropism	4
Eucgr.I01402	<i>SLP2</i>	Germination	4

Table 5.3 Well-characterized homologs in *Arabidopsis* of upregulated genes the flowers-only cluster analysis.

Locus in <i>Eucalyptus</i>	Gene in <i>Arabidopsis</i>	Function	Cluster
Eucgr.H03170	<i>IAA7</i>	Stem gravitropic growth	1
Eucgr.G01774	<i>MYB4</i>	Regulation of flavonoid synthesis	1
Eucgr.K02977	<i>RAS1</i>	Salt tolerance and ABA sensitivity	1
Eucgr.I00659	<i>XTH6</i>	Cell wall modifications	1
Eucgr.B03746	<i>GPS1</i>	Stem gravitropic growth	2
Eucgr.I02677	<i>LHCB6</i>	PSII minor antenna complex	2
Eucgr.H04498	<i>LOX1</i>	Cell wall-mediated defense	2
Eucgr.C03822	<i>PERK1</i>	Cell wall protection	2
Eucgr.H02748	<i>TOPII</i>	Chromosome interlock resolution in meiosis	2
Eucgr.H04418	<i>HST</i>	Flowering inhibition	3

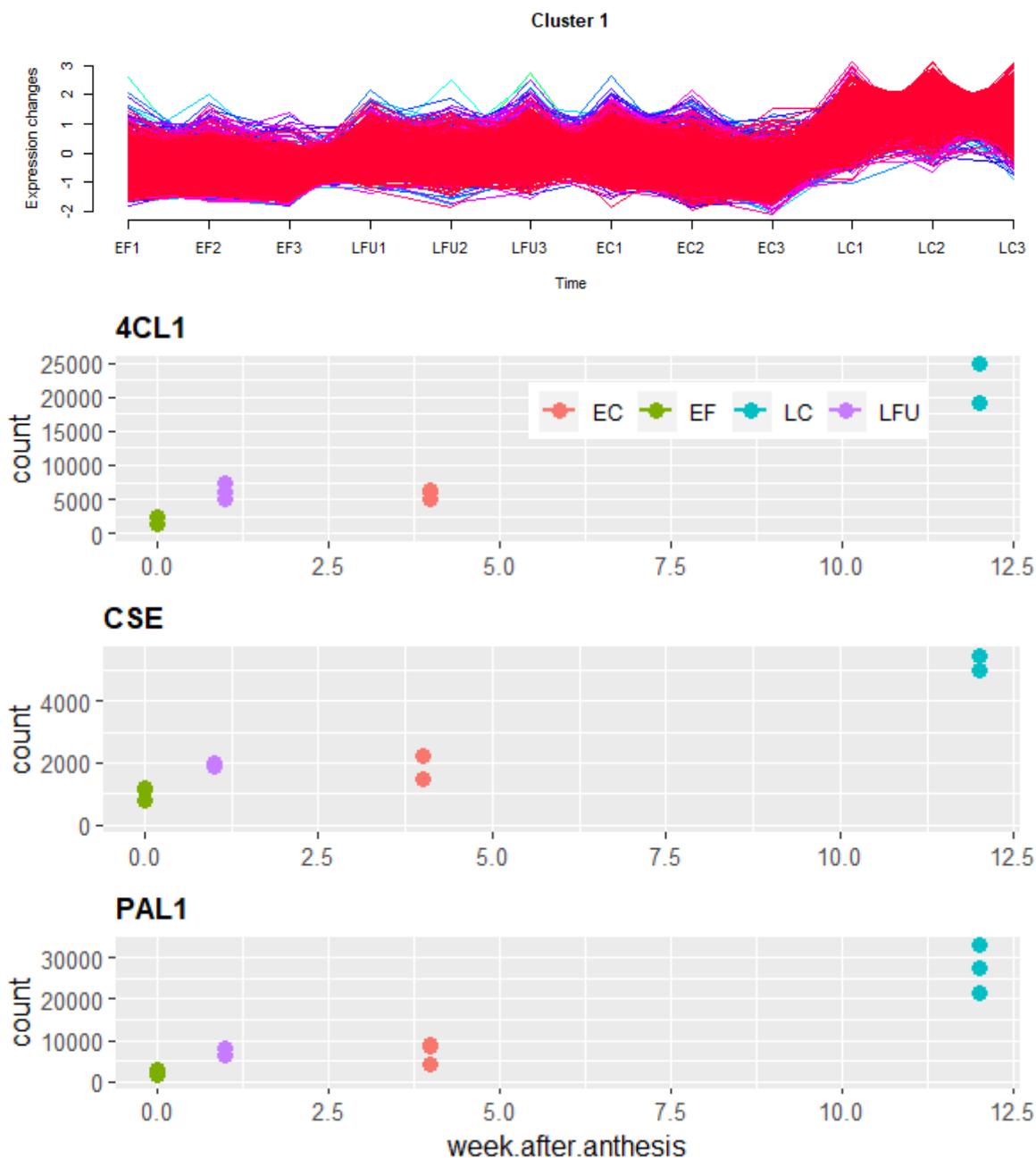


Fig. 5.7 Expression levels in cluster 1 of the capsule development cluster analysis. The genes identified as belonging to cluster one had low expression in EF, moderate expression in LFU and EC, and high expression in LC. *4CLI*, *CSE*, and *PALI* had high membership scores for cluster 1. EC, early capsule; EF, early flower; LC, late capsule; and LFU, late flower unconfined.

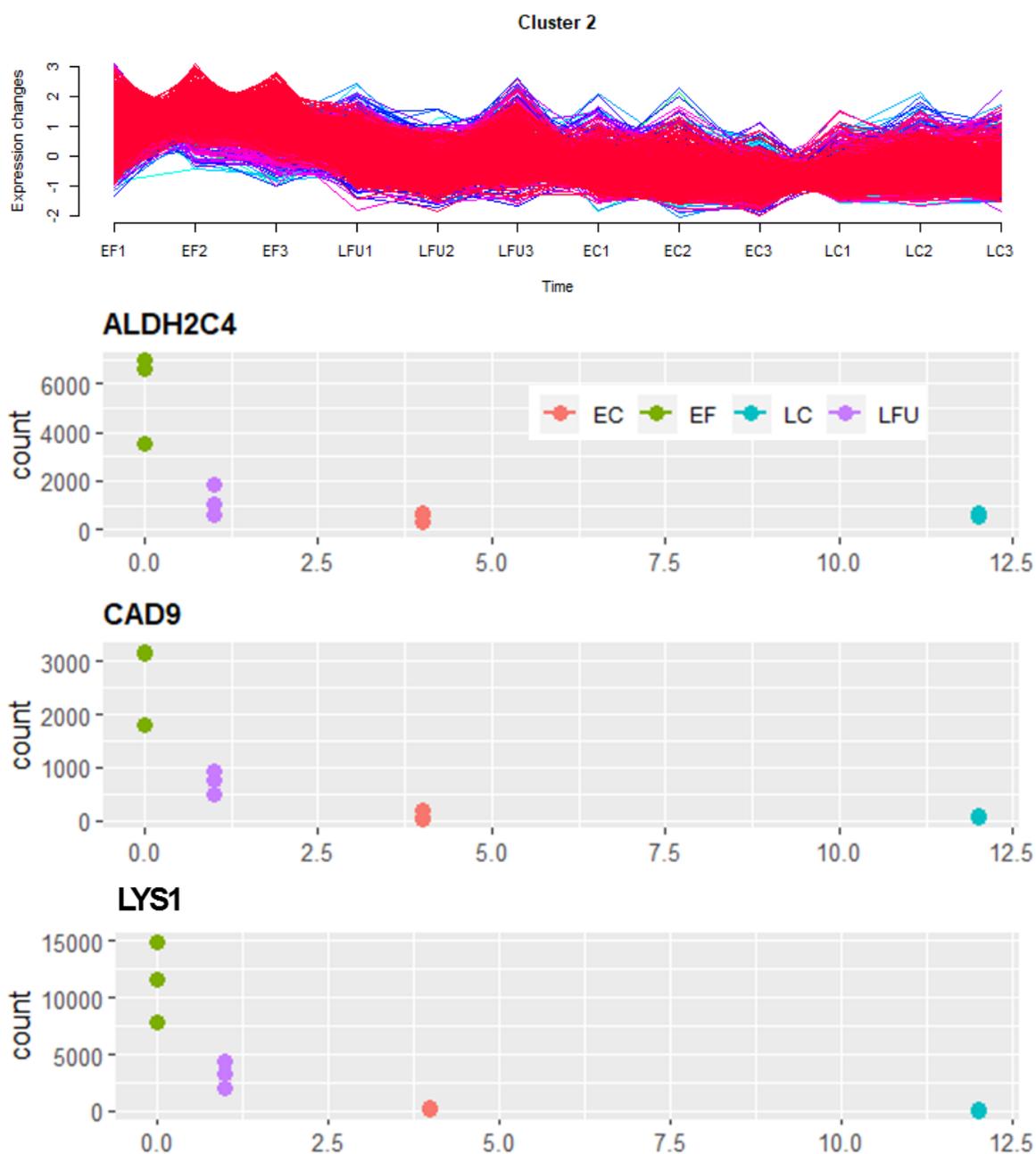


Fig. 5.8 Expression levels in cluster 2 of the capsule development cluster analysis. The genes identified as belonging to cluster 2 had high expression in EF, moderate expression in LFU, and low expression in EC and LC. *ALDH2C4*, *CAD9*, and *LYS1* had high membership scores for cluster 2. EC, early capsule; EF, early flower; LC, late capsule; and LFU, late flower unconfined.

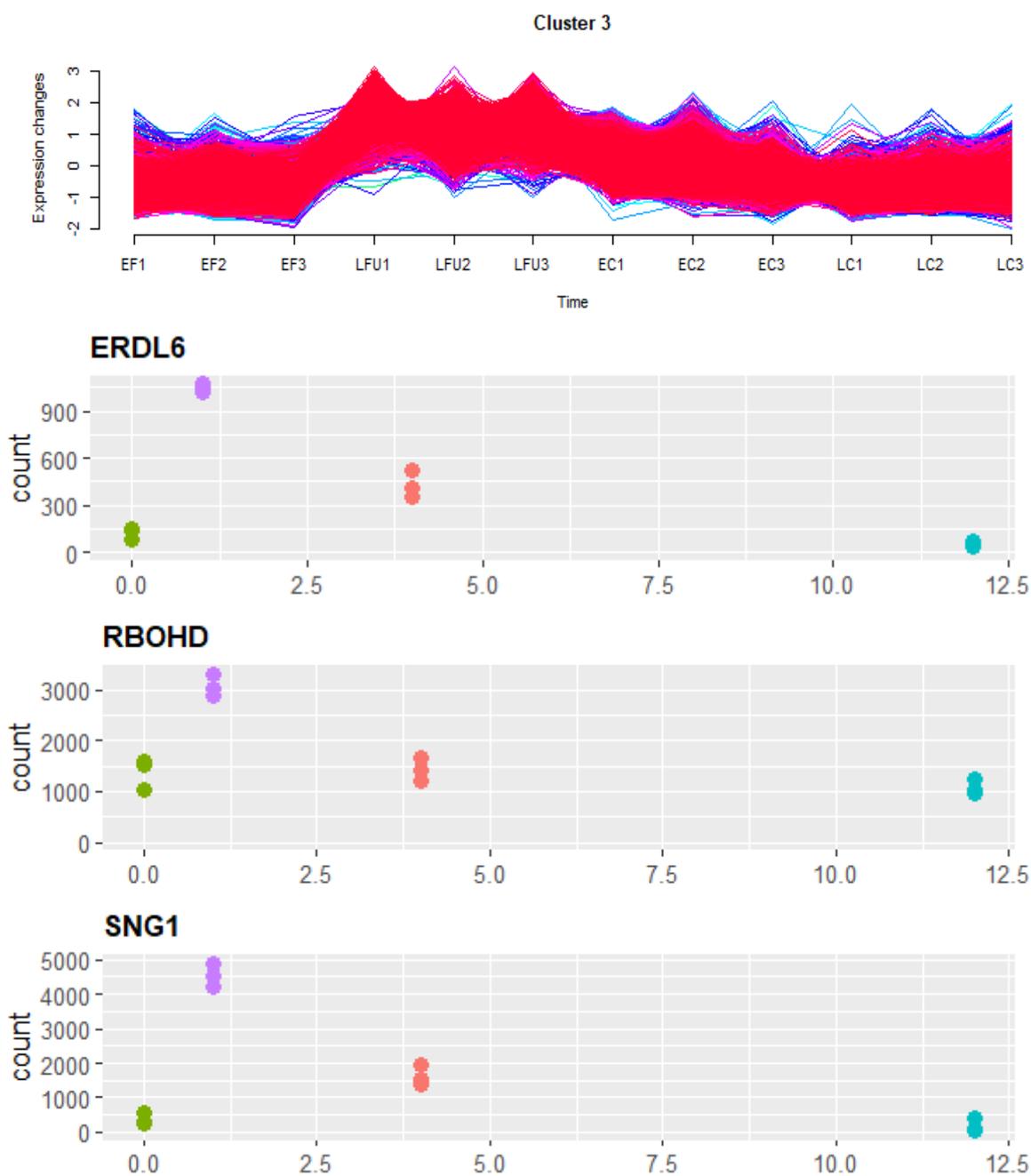


Fig. 5.9 Expression levels in cluster 3 of the capsule development cluster analysis. The genes identified as belonging to cluster 3 had low expression in EF, high expression in LFU, moderate expression in EC, and low expression in LC. *ERDL6*, *RBOHD*, and *SNG1* had high membership scores for cluster 3. EC, early capsule; EF, early flower; LC, late capsule; and LFU, late flower unconfined.

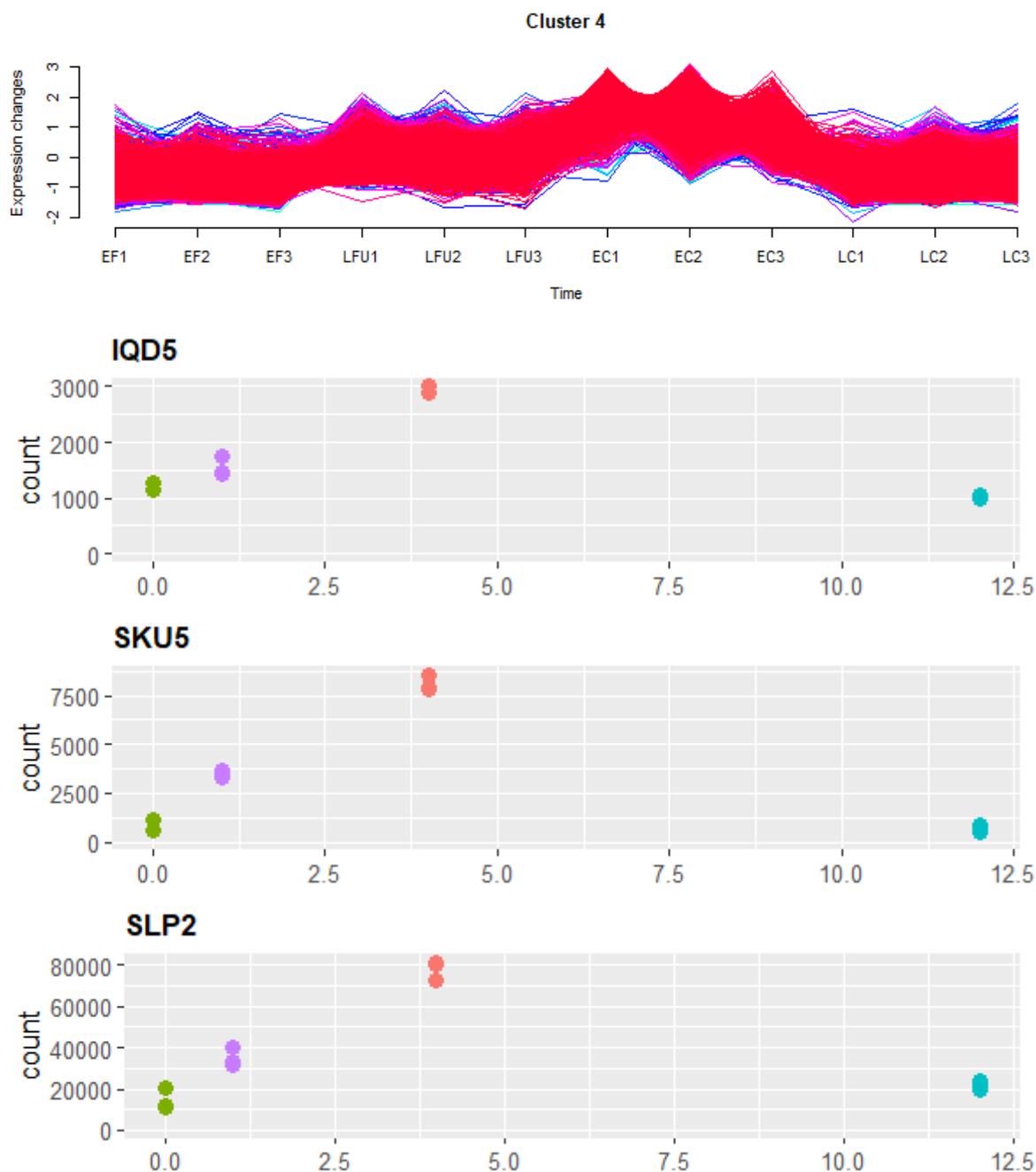


Fig. 5.10 Expression levels in cluster 4 of the capsule development cluster analysis. The genes identified as belonging to cluster 4 had low expression in EF, moderate expression in LFU, high expression in EC, and low expression in LC. *IQD5*, *SKU5*, and *SLP2* had high membership scores for cluster 4. EC, early capsule; EF, early flower; LC, late capsule; and LFU, late flower unconfined.

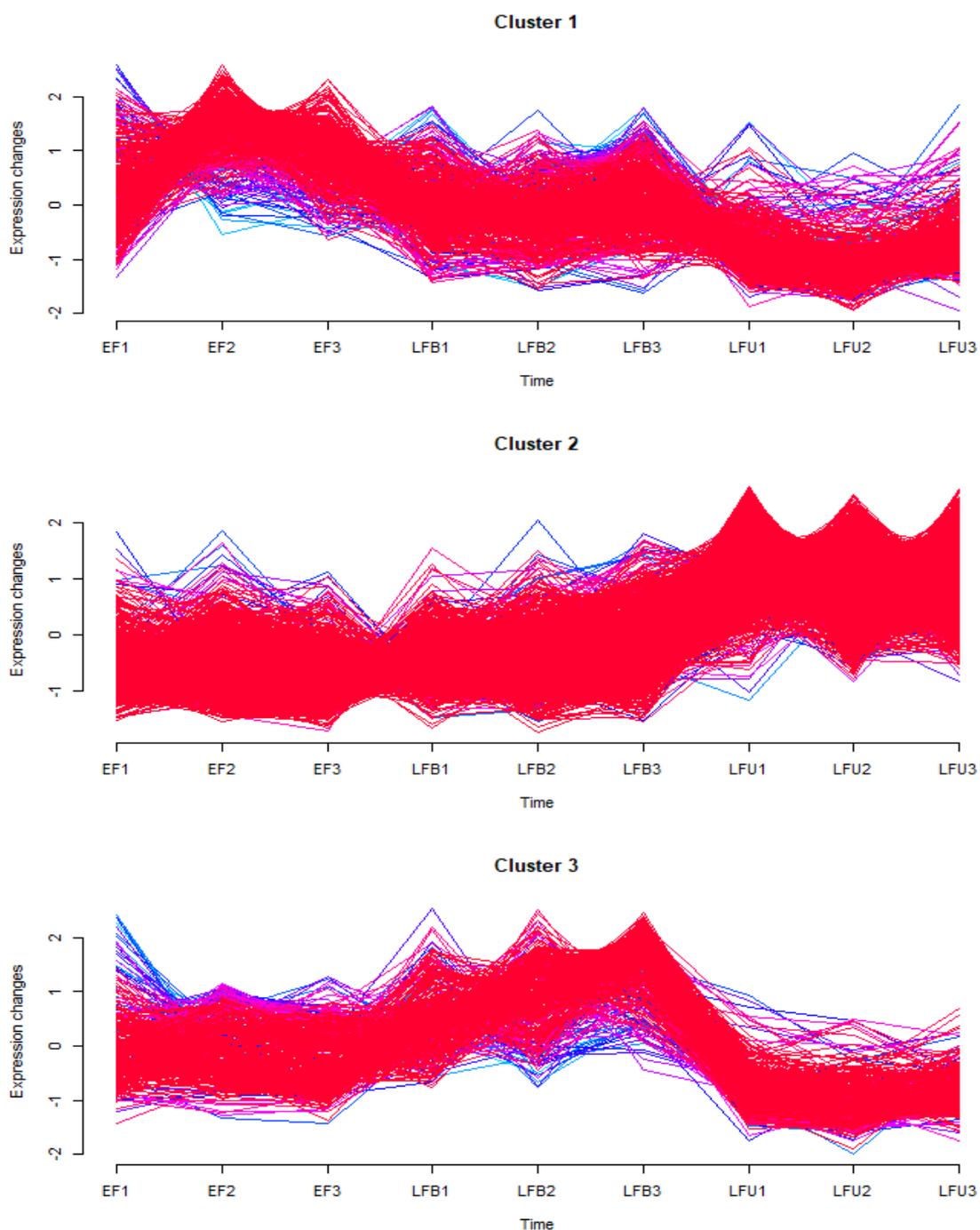


Fig. 5.11 Expression levels in the three clusters of the flowers-only cluster analysis. The genes in cluster 1 had high expression in EF, moderate expression in LFB, and high expression in LFU. The genes in cluster 2 had low expression in EF and LFB, and high expression in LFU. The genes in cluster 3 had moderate expression in EF, high expression in LFB, and low expression in LFU. EF, early flowerer; LFB, late flowerer bagged; and LFU, late flowerer unconfined.

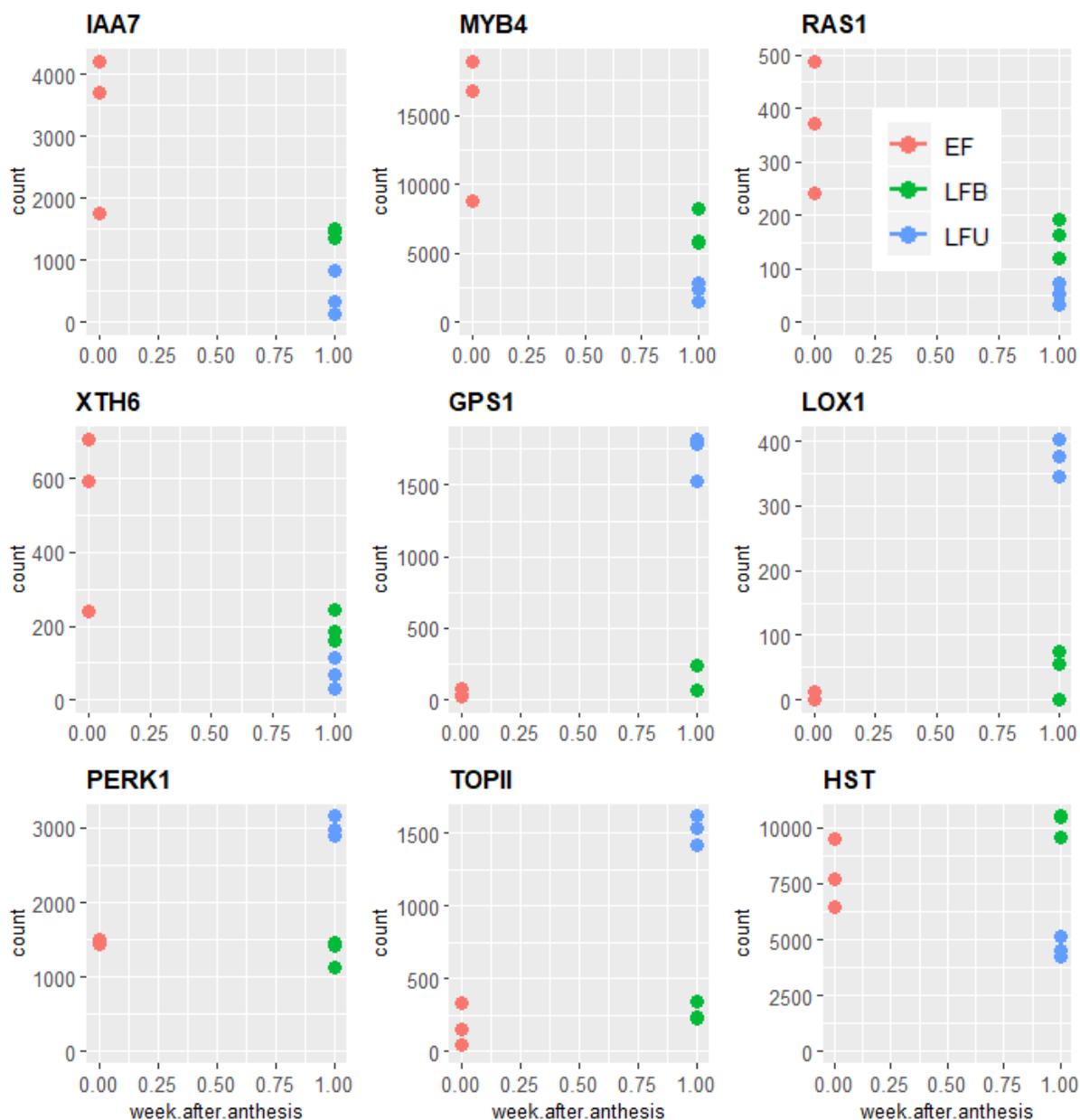


Fig. 5.12 Examples of the expression levels in the three clusters of the flowers-only cluster analysis. *IAA7*, *MYB4*, *RAS1*, and *XTH6* belong to cluster 1. *GPS1*, *LOX1*, *PERK1*, and *TOPII* belong to cluster 2. *HST* belongs to cluster 3. EF, early flower; LFB, late flower bagged; and LFU, late flower unconfined.

For cluster 1, in the flowers-only cluster analysis (i.e., high expression in EF), we find the following GO enriched terms: cell redox homeostasis, regulation of transcription

(DNA-templated), aromatic amino acid family metabolic process, terpenoid biosynthetic process, pyruvate metabolic process, small molecule catabolic process, pigment metabolic process, oxidation-reduction process, defense response, and response to stimulus. For cluster 2 (i.e., high expression in LFU), we find: oxidation-reduction process, microtubule-based movement, mitotic nuclear division, mannose metabolic process, cellulose biosynthetic process, DNA replication initiation, chitin catabolic process, response to water, response to oxidative stress, sodium ion transport, proteolysis, cytokinesis, and photosynthesis (light harvesting). For cluster 3 (i.e., high expression in LFB), we find: coenzyme biosynthetic process, purine ribonucleotide biosynthetic process, cellular amino acid metabolic process, RNA processing, ribosome biogenesis, response to auxin, transmembrane transport, and oxidation reduction process.

Table 5.4 Number of associated GO terms and annotated genes in AgriGO

Contrast/ cluster	Category	No of genes	No of genes annotated in AgriGO	No of significant GO terms
P vs L	N/A	7,958	4,194	384
1	Flower and seed development	1,428	764	142
2		977	609	161
3		1,482	892	209
4		2,220	1,243	100
1	Flower development (capsules excluded)	461	286	78
2		2,644	1,516	206
3		377	195	110

5.3.8 GO enriched categories in pollen

GO enrichment analysis associated 384 significant terms with the upregulated DEGs identified in the contrast of pollen vs leaf (Table S5.21). Among the GO significant terms based on p-value (based on chance), we find actin filament depolarization, autophagosome assembly, pyridine-containing compound biosynthetic process, chitin catabolic process, ubiquitin-dependent protein catabolic process, regulation of transcription from RNA polymerase II promoter, protein dephosphorylating, protein oligomerization, translational frameshifting, translational initiation, positive regulation of translational elongation, positive regulation of translational termination, ATP metabolic process, CTP biosynthetic process, GTP biosynthetic process, UTP

biosynthetic process, ATP hydrolysis coupled transmembrane transport, SRP-dependent cotranslational protein targeting to membrane, vacuolar transport, mitochondrial transport, vesicle mediated transport, iron-sulfur cluster assembly, phospholipid metabolic process, cell redox homeostasis, fatty acid biosynthetic process, water-soluble vitamin biosynthetic process, and steroid biosynthetic process.

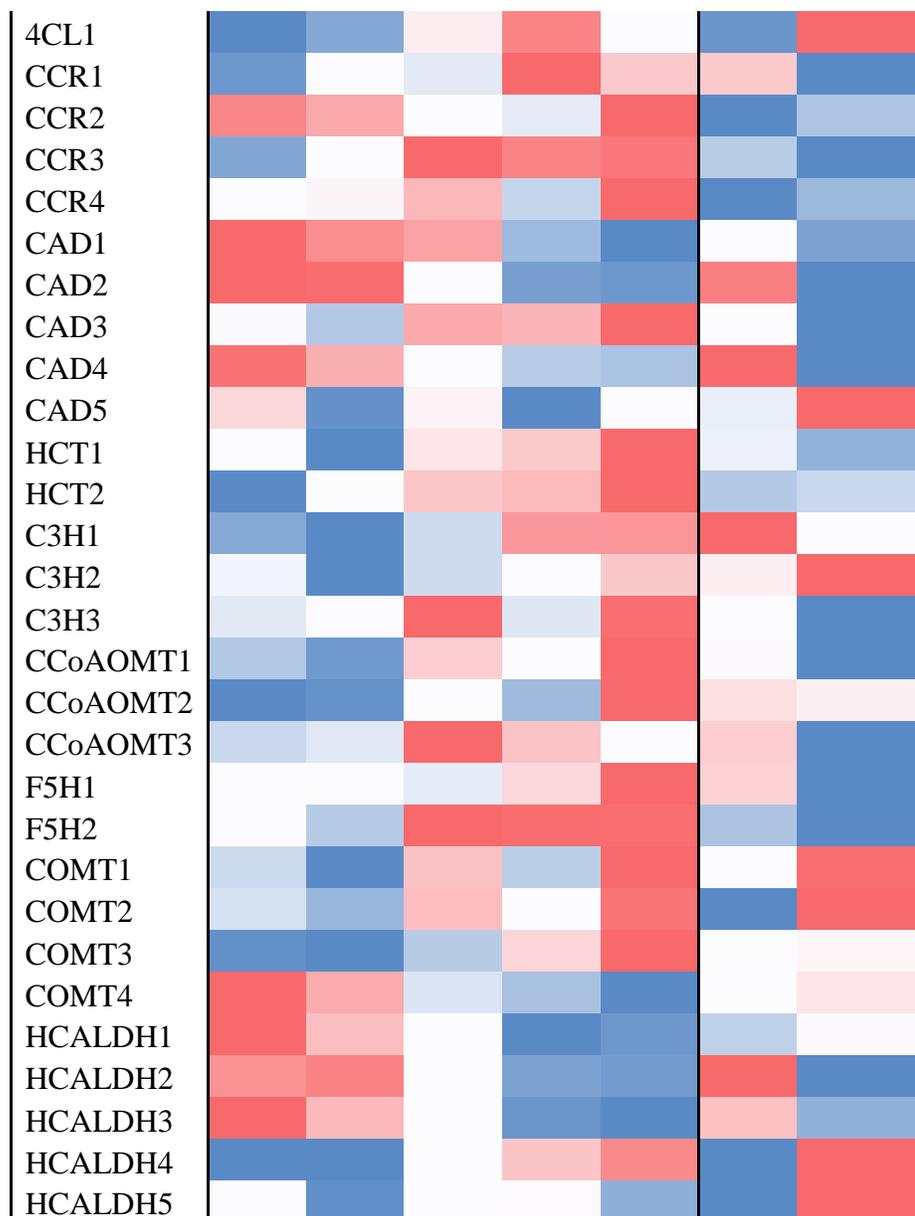
5.3.9 Expression of lignin-biosynthesis specific genes

The large majority of the *Eucalyptus*' homologs of the enzymes involved in phenylpropanoid biosynthesis have higher expression in the capsule stages (Table 5.5). However, several paralogs of prephenate dehydratase (PDT), arogenate dehydratase (ADT), cinnamyl alcohol dehydrogenase (CAD), and hydroxycinnamaldehyde dehydrogenase (HCALDH) have higher expression in the flower tissues compared to the capsules.

Table 5.5 Heatmap representing expression of lignin-biosynthesis specific homologs in *Eucalyptus*. CM, chorismate mutase; PDT, prephenate aminotransferase; PAT, prephenate aminotransferase; ADT, arogenate dehydratase ; PAL, phenylalanine ammonia-lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; HCT, hydroxycinnamoyl-coenzyme A shikimate:quinic acid hydroxycinnamoyl-transferase; C3'H, p-coumaroyl shikimate 3'-hydroxylase; CCoAOMT, caffeoyl CoA 3-O-methyltransferase; F5H, ferulate 5-hydroxylase; COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase; HCALDH, hydroxycinnamaldehyde dehydrogenase.

Gene	EF	LFB	LFU	EC	LC	P	L
CM1		Blue	White	Light Blue	Red	Red	Blue
CM2	Blue	Light Blue	White	Red	Blue	White	Red
PDT1	White	Red	Red	White	Red	Blue	Blue
PDT2	Red	Red	Light Red	Blue	Blue	White	White
PAT	Blue	Light Red	Red	White	Light Red	Blue	White
ADT1	Light Red	Light Blue	Light Red	White	Light Blue	Red	Blue
ADT2	Light Red	Red	White	Light Blue	Light Blue	Blue	Red
ADT3	White	Light Blue	Red	White	Red	Blue	Blue
PAL1	Blue	Blue	Light Red	Light Red	Red	White	Light Blue
PAL2	Blue	Light Blue	Red	Light Red	Red	White	Blue
C4H1	Blue	Blue	Light Red	White	Red	Light Red	Blue
C4H2	White	White	Red	White	Red	White	Blue

Table 5.5 Heatmap representing expression of lignin-biosynthesis specific homologs in *Eucalyptus* (continued)



5.3.10 Correlations between RNA-Seq and qRT-PCR expression levels

We executed quantitative real-time PCR (qPCR) on seven selected DEGs, including *EgAG*, *EgCAL*, *EgFT*, *EgFUL1.1*, *EgFUL1.2*, *EgGADPH*, *EgSHP2*, and *EgSTK*. *EgGAPDP* was the housekeeping reference gene. We checked gene expression in three of the seven tissues, including EF, EC, and LC. The expression levels, high or low, are

consistent between RNA-seq and qPCR (Fig. 5.5). The trends in gene expression are also consistent (Fig. 5.5). The Pearson correlation coefficient of the linear regression analysis was 0.91.

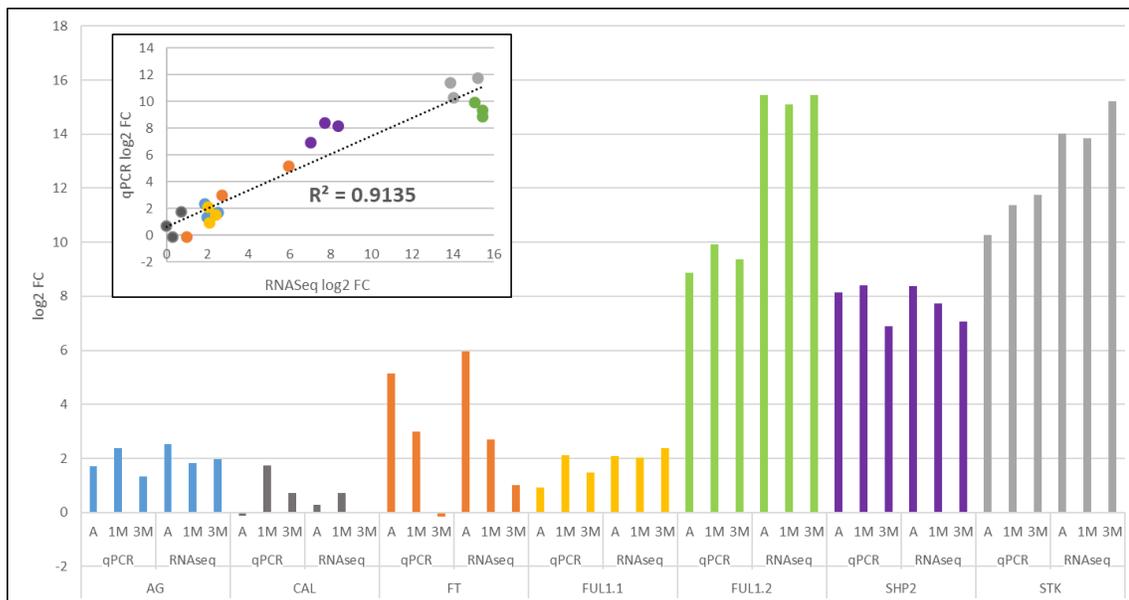


Fig. 5.13 Validation of RNA-seq expression with qPCR data. Expression levels in qPCR and RNA-Seq for seven DEG in early flower, early capsule, and late capsule. *AG*: *AGAMOUS*; *CAL*: *CAULIFLOWER*; *FT*: *FLOWERING LOCUS T*; *FUL1.1*: *FRUITFULL1.1*; *FUL1.2*: *FRUITFULL1.2*; *SHP2*: *SHATTERPROOF 2*; *STK*: *SEEDSTICK*. A: early flower; 1M: early capsule; 3M: late capsule.

5.4 Discussion

We have a wealth of information from *Arabidopsis*, which helps support investigations in other plants. However, genes can be conserved or co-opted to other functions, depending on the evolutionary pressures. Omic studies of non-model species provide insights into distinction, and also add to the wealth of gene function information available.

This study is the second study to publish a transcriptome on flowering tissues of *E. grandis*. In the first transcriptome, Vining *et al.* (2015) had early and late bud as their flowering tissues. In this RNA-seq database, we have flowers at anthesis, flowers one week later, developing seed capsules at one and three months after anthesis, and mature pollen. We also have fully-expanded leaves as comparator for vegetative vs. floral and fruit tissues. The *E. grandis* genome was published in 2014 and since it has had two

updates, and is now version 2.0 (Myburg *et al.*, 2014), which facilitated our transcriptomic analysis.

According to our gene-expression clustering analyses (Poisson distance plot and PCA), the transcriptomic profiles of our tissues can be divided into three groups: leaf, pollen, and flower/seed (Fig. S5.1 and S5.2). When we remove the leaf and pollen libraries, we find that the replicates (i.e., different trees) of each individual flower or capsule tissue type generally cluster together (Fig. S5.3). However, we do see that LFB1 clusters with EF1 in the Poisson distance plot and EF1 clusters with the LFB replicates in the PCA plot.

In differential expression analysis, we set the FDR cutoff to 0.05 and the LFC cutoff to 0.5 or higher for upregulated genes and -0.5 or lower for downregulated genes. We were able to identify 11,438 DEGs when contrasting all tissues against leaf. Among the most highly-expressed ones for every contrast, we identified several homologs in *Arabidopsis* that have been well characterized. We also identified DEGs among a simulated flower development series and implemented fuzzy c-means clustering analysis and gene ontology (GO) analysis. Below we comment on a number of identified genes and gene ontology classes that appeared to have biological interest and potential use for eucalypt biotechnology.

Flowering and fertility

A number of homologs to genes known to be required for carpel and ovule development were identified during our analyses. *FRUITFULL* (*FUL*) and *SEPALLATA* (*SEPI*) are both floral-specific MADS-box transcription factors. *FUL* and *APETALA1* (*API*) originated from a gene duplication that predated the diversification of the eudicots. In *Arabidopsis*, *API* and *FUL* have divergent functions since they can only partially rescue each other (McCarthy *et al.*, 2015). There are two *FUL* paralogs in *Eucalyptus* that have the *FUL*-like C-terminal motifs (i.e., LPAWML) (McCarthy *et al.*, 2015). This motif is missing from all recognized *API* homologs. In *Arabidopsis*, *FUL* has high expression in the axis of the inflorescence, the style and carpel of the mature flower, the siliques, and the pod of the first silique (Mandel & Yanofsky, 1995; Gu *et al.*, 1998). *FUL* is

important for valve cell elongation in the silique and cauline leaf development (Mandel & Yanofsky, 1995; Gu *et al.*, 1998). Similar to what we found in *Eucalyptus*, tomato has two *FUL* homologs, and both are redundantly involved in cell wall modification and fruit ripening (Bemer *et al.*, 2012).

SEP1 and *SEP3* are MADS-box transcription factors of class E in the ABCDE homeotic model (Krizek & Fletcher, 2005). There are four *SEP* genes in *Arabidopsis*, i.e. *SEP1*, *SEP2*, *SEP3*, and *SEP4*, all of which are expressed in all four floral whorls. Single and double *sep* mutants have weak mutant phenotypes. The *sep1 sep2 sep3* mutant phenotype has sepals in all four whorls (Pelaz *et al.*, 2000). In the *sep1 sep2 sep3 sep4* quadruple mutant, flowers are indeterminate and all whorls are made of vegetative leaves (Ditta *et al.*, 2004). We found that there were five *SEP*-like genes in *E. grandis* based on expression and sequence homology.

SEEDSTICK (STK) is a MADS-box transcription factor essential for carpel, ovule, and funiculus development (Pinyopich *et al.*, 2003; Favaro *et al.*, 2003). *STK* is also involved in seed abscission (Balanzà *et al.*, 2016) and seed development and metabolism (Mizzotti *et al.*, 2012, 2014; Ezquer *et al.*, 2016). We found that there was one *STK*-like homologous gene (89.6% sequence homology) in *Eucalyptus*, which was expressed, and at about the same level, in early flower, late flower (i.e., LFU), and early capsule. It had the highest expression in late capsule (about three times that of EF, LFB, and EC). It had no expression in leaf and the expression in late flower bagged and pollen was about half of that in early flower, late flower (i.e., LFU), and early capsule. The *Arabidopsis STK* gene has medium expression in the stigma and style, and high expression in the ovules and seeds. This is comparable to what we saw in our transcriptome, except for the expression in mature pollen. In *Arabidopsis*, there is no *STK* expression in pollen. However, *STK* expression was found in rice pollen (Zhang *et al.*, 2019).

ECERIFERUM1 (CER1) and *HASTY (HST)* were also among the genes identified in the flowering and seed development and the flower-only development clustering analyses respectively. They are not flower-identity genes but they are required for flower fertility. *CER1* is essential for the conversion of aldehydes to alkanes in the epicuticular wax

biosynthesis pathway (McNevin *et al.*, 1993; Hannoufa *et al.*, 1993). In *Arabidopsis*, *cer1* mutants have glossy stems and are male sterile under dry conditions (Aarts *et al.*, 1995). *HST* encodes an importin/exportin protein involved in miRNA transport. Mutations in *HST* lead to reduced fertility because of lower pollen number and reduced megaspore mother cells number.

We found that there were five *CER1* homologous genes of in *Eucalyptus* according to Phytozome. However, one of them had no expression in any tissue in our transcriptome (i.e., *Eucgr.D01120*). *Eucgr.D01144* had low expression in all tissues. *Eucgr.D01172* had low expression only in early and late capsule, and no expression in any of the other tissues. *Eucgr.D01149* had high expression in early flower and late flower bagged, and low expression in the rest of the reproductive tissues. The homolog we identified, *Eucgr.D01152*, had high expression in all the reproductive tissues with highest expression at late capsule followed by early capsule. *Arabidopsis CER1* has high expression in mature flower and lower expression in the rest of the reproductive tissues. Based on expression, it would seem that *Eucgr.D01149* is the true homolog of *CER1*. *Eucgr.D01152* might have diverged in function towards a more-seed or capsule specific role. All the *Eucalyptus* homologs mentioned had at least 75% nucleotide homology.

There is only one *HST* homolog in *Eucalyptus* (76% nucleotide homology). *HST* has high expression in flowers and other reproductive organs, but it is expressed in all tissues. The homolog we identified, *Eucgr.H04418*, had comparable expression: high in early flower, late flower bagged, and leaf, and medium expression in all the other tissues.

Cell shape modification

ACTIN DEPOLYMERIZING FACTOR 10 (ADF10), *ROP BINDING PROTEIN KINASES 1 (RBK1)*, *SKU5*, and *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 6 (XTH6)* are involved in different aspects of cell shape. *ADFs* are a highly conserved protein family involved in actin dynamics with the most number of isoforms in plants (Maciver & Hussey, 2002; Feng *et al.*, 2006; Bamburg & Bernstein, 2008). *ADF7* and *ADF10* are expressed exclusively in pollen and pollen tubes (Pina *et al.*, 2005; Ruzicka *et al.*, 2007; Bou Daher *et al.*, 2011) and they are required for proper pollen tube elongation

(Bou Daher *et al.*, 2011; Bou Daher & Geitmann, 2012; Zheng *et al.*, 2013). *ADF10* is specifically involved in actin filament reorganization and vesicle trafficking (Jiang *et al.*, 2017). Actin filament reorganization and vesicle trafficking are processes essential for proper pollen tube elongation. Actin filament depolarization and vesicle mediated transport were amongs the most enriched GO categories when we compared expression in pollen to leaf. We found that there were nine *ADF10-like* genes in *Eucalyptus*. However, there are ten *ADF* genes in *Arabidopsis*.

SKU5 encodes a glycosylphosphatidylinositol-anchored protein (GPI-AP). Single *sku5* mutants have shorter roots and the roots have a skew and don't go straight down (Sedbrook *et al.*, 2002; Swarbreck *et al.*, 2019). *XTH6* is involved in cell wall modification. Its expression was high in an *Arabidopsis* silique transcriptome 6 DPA (days post-anthesis) (Jaradat *et al.*, 2014). Single *rbk1* have shorter root cells (Enders *et al.*, 2017). Based on sequence homology, there are only one *RBK1*, one *SKU5*, and one *XTH6* homologous genes in *Eucalyptus*.

Defense

LIPOXYGENASE 1 (LOX1), *LYSOZYME 1 (LYS1)*, *PROLINE EXTENSION RECEPTOR 1 (PERK1)*, *RESPIRATORY BURST OXIDASE HOMOLOGUE D (RBOHD)*, *UDP-GLUCOSYLTRANSFERASE 73B2 (UGT73B2)* are involved in defense against pathogens. *LOX1* encodes a 9-lipoxygenase involved in the beginning of oxylipin biosynthesis pathway. Oxylipins are involved in cell wall-induced defense (Wasternack & Feussner, 2018). Single *lox1 Arabidopsis* mutants were more susceptible to pathogens (Hwang & Hwang, 2010), in part because they could not close their stomata (Montillet *et al.*, 2013). *LYS1* mutants in *Arabidopsis* are highly susceptible to bacterial infections (Liu *et al.*, 2014). Liu *et al.* (2014) propose that *LYS1* is involved in peptidoglycans breakdown of bacterial structures. In *Brassica napus*, *PERK1* encodes a plasma membrane-bound putative receptor protein kinase involved in cell wall wound-response (Silva & Goring, 2002). *PERK1* transcription is increased after wounding stimuli and pathogen infection (Silva & Goring, 2002). *RBOHD* is required for reactive oxygen species (ROS) production after pathogen attack and cell wall damage (CWD) (Liu & He,

2016). CWD-induced lignin deposition is reduced in *rbohd* mutants (Denness *et al.*, 2011). *UGT73B2* has not been well characterized. However, it is expressed in flowers and roots of *Arabidopsis* and its expression is double in powdery mildew resistant mutants compared to WT (Nishimura *et al.*, 2003). We found that there were one homologous gene to *PERK1*, six to *LOX1*, seven to *RBOHD*, eight to *LYS1*, and fifteen to *UGT73B2*.

Light and gravitropism

GRAVITY PERSISTENT SIGNAL 1 (GPS1), *INDOLE-3-ACETIC ACID 7 (IAA7)*, *LIGHT HARVESTING COMPLEX PHOTOSYSTEM II SUBUNIT 6 (LHCB6)* have light-dependent functions. *GPS1* and *IAA7* are important for gravitropic growth. *GPS1* encodes CYP705A22, a cytochrome P450 monooxygenase. *Arabidopsis gps1* and *iaa7* mutants have an altered stem curvature (Wilson *et al.*, 1990, Withers *et al.*, 2013) and *iaa7* mutants are also insensitive to auxin, ethylene, and salicylic acid and (Wilson *et al.*, 1990). Interestingly, a homolog in tomato, Sl-IAA17, is mostly transcriptionally active during fruit development. Single *si-iaa17* null mutants had larger fruits and higher ploidy in their pericarp cells compared to WT (Su *et al.*, 2014). *LHCB6* encodes a minor antenna complex of photosystem II (PSII) important for the formation of PSII-LCH supercomplexes (Dall'Osto *et al.*, 2020). Single *lhcb6* mutants had slower growth, reduced photosynthetic rate in light-limiting conditions, and delayed flowering (Kovács *et al.*, 2006). We found that there were two homologous genes to *LHCB6*, three to *GPS1*, and three to *IAA7* in *Eucalyptus*.

Lignification of flowers and cell wall modification

Eucalypt capsules get increasingly woody as they develop and mature; we therefore examined if genes related to cell wall maturation, such as those for lignification, were differentially regulated. The genes *ALDEHYDE DEHYDROGENASE 2C4 (ALDH2C4)*, *4-COUMARATE: COA LIGASE 1 (4CLI)*, *CAFFEOYL SHIKIMATE ESTERASE (CSE)*, *LACCASE4 (LAC4)*, *MYB DOMAIN PROTEIN 4 (MYB4)*, *MYB21*, *MYB103*, *PHENYLALANINE AMMONIA-LYASE 1 (PAL1)*, and *SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN 1 (SND1)* were strongly involved in capsule maturation based on their steadily increasing expression from early flower to late capsule.

ALDH2C4 encodes a sinapaldehyde dehydrogenase that catalyzes the oxidation of coniferylaldehyde and sinapaldehyde forming ferulic acid and sinapic acid, respectively (Nair *et al.*, 2004). *ALDH2C4* is involved in the phenylpropanoid pathway and *aldh2c4* single mutants in *Arabidopsis* have less cell wall-bound ferulic acid esters (Nair *et al.*, 2004). *4CL1*, *CSE*, and *PAL1* encode three enzymes involved in the lignin biosynthesis pathway (Raes *et al.*, 2003). There are four isoforms of *4CL* and *PAL* in *Arabidopsis* (Huang *et al.*, 2010). Triple *4CL* mutants, quadruple *PAL* mutants, and *CSE* mutants have reduced levels of lignin in *Arabidopsis*, *Populus*, and *Medicago truncatula* respectively (Huang *et al.*, 2010; Vanholme *et al.*, 2013; Li *et al.*, 2015; Ha *et al.*, 2016; Saleme *et al.*, 2017). We found that there were one homologous gene to *4CL1*, one to *CSE*, one to *MYB21*, one to *MYB103*, two to *SND1*, five to *CAD9*, six to *ALDH2C4*, two to *PAL1* in *Eucalyptus*.

LAC4 is one of 17 laccase genes in *Arabidopsis* (Turlapati *et al.*, 2011). *LAC4* is essential for lignin deposition in tracheary elements (TEs) (Schuetz *et al.*, 2014). *LAC4* is secreted in secondary cell walls where it remains during secondary cell wall development (Yi Chou *et al.*, 2018). *MYB4*, *MYB21*, and *MYB103* are members of the *R2R3-MYB* transcription factor family (Marocco *et al.*, 1989; Martin & Paz-Ares, 1997). *MYB* transcription factors are involved in biosynthesis of secondary metabolites, development, and stress response. Plants overexpressing *MYB4* have reduced levels of lignin in *Arabidopsis*, pine, maize, and switchgrass (Jin *et al.*, 2000; Patzlaff *et al.*, 2003; Fornalé *et al.*, 2010; Shen *et al.*, 2012). And *myb4 Arabidopsis* mutants have reduced levels of flavonols (Jin *et al.*, 2000; Fornalé *et al.*, 2014). In *Arabidopsis*, *MYB21* is expressed in flower buds and its ectopic expression leads to upregulation of *PHENYLALANINE AMMONIA LYASE (PAL)*, a gene in the lignin biosynthesis pathway (Shin *et al.*, 2002). *MYB103* belongs to the same TF family as *MYB21*. *SND1* is plant-specific NAC transcription factor that redundantly regulates secondary wall biosynthesis in fibers with *NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1 (NST1)* and *NST2* (Mitsuda *et al.*, 2005, 2007; Zhong *et al.*, 2006, 2007; Zhong & Ye, 2015). We found that there were eight *LAC4* homologous genes in *Eucalyptus*.

Nectar

Most eucalypts are predominantly insect or bird pollinated, for which nectar provides a significant attraction to pollinators. *E. grandis* has modified stomata instead of traditional nectaries in the hypanthium surface (Davis, 1969). These stomata look exactly like leaf stomata yet they remain open for most of the flower's existence, unless they become occluded, to potentially avoid entry of pathogens (Davis, 1997). *SWEET9* encodes a nectary-specific sucrose transporter (Lin *et al.*, 2014). *SWEET9* is highly conserved among angiosperms (Jeena *et al.*, 2019; Xie *et al.*, 2019; Jiang *et al.*, 2020), including in predominantly wind pollinated species like poplar (Lin *et al.*, 2014). Poplar extrafloral nectaries secrete nectar with the purpose of repelling herbivores from eating their flowers (Escalante-Pérez *et al.*, 2012). We found that there was only one gene homologous to *SWEET9* in *Eucalyptus*.

Phosphate

Phosphate metabolism is highly active during pollen metabolism (Rutley & Twell, 2015). We found that *ACTIN-DEPOLYMERIZING FACTOR 10 (ADF10)*, *PURPLE ACID PHOSPHATASE 26 (PAP26)*, and *PYROPHOSPHORYLASE 3 (PPa3)* were among the most highly expressed genes in pollen compared to leaf. *PAP26* is an acid phosphatase induced by phosphate (Pi) starvation (Veljanovski *et al.*, 2006; Tran *et al.*, 2010; Hurley *et al.*, 2010). In *Arabidopsis*, *PAP26* is required for Pi scavenging during leaf senescence and Pi-deprivation (Robinson *et al.*, 2012; Shane *et al.*, 2014). Moreover, there are several other *PAPs* that are expressed in pollen and flowers, including *PAP15* and *PAP23* (Zhu *et al.*, 2005; Kuang *et al.*, 2009). *PAP15* is not induced by Pi deficiency, so its proposed function is Pi mobilization in germinating pollen. We found that *Eucgr.I01300* is most homologous to *PAP26* and *PAP10* (also not pollen specific), however the expression pattern of *Eucgr.I01300* is so dramatic compared to the other seven homologs in *Eucalyptus*, that makes *Eucgr.I01300* appear to have the most similar function to *PAP15*.

PPa3 is an inorganic pyrophosphatase. PPases catalyze the generation of inorganic phosphate by cleavage of pyrophosphate (PPi). *PPa3* is expressed preferentially in

pollen, stamens, and flowers, with low expression in roots, and hardly any expression in other tissues (Navarro-De la Sancha *et al.*, 2007). We found that there were eight and seven homologous genes to *PPa3* and *PAP26* in *E. grandis* respectively.

SLP2 encodes a plant mitochondrial-specific phosphatase. Compared to WT NO seeds, *slp2* mutants exhibit fast germination and *35S::AtSLP2* seeds exhibit delayed germination (Uhrig *et al.*, 2017). We identified only one gene homologous to *SLP2* in *Eucalyptus*.

Genes that require further exploration

We found a number of differentially regulated genes whose functional homology to known genes was unclear. For example, often there were multiple eucalypt homologs to *GPDHC1* (2 homologs), *4-HYDROXY-3-METHYLBUT-2-ENYL DIPHOSPHATE REDUCTASE (HDR)* (2 homologs), *IQ67 DOMAIN5 (IQD5)* (2 homologs), *PHOSPHOLIPASE C2 (PLC2)* (6 homologs), *RESPONSE TO ABA AND SALT 1 (RAS1)* (3 homologs), and *TOPOISOMERASE II (TOPII)* (2 homologs) identified during our analyses but their specific function in flower or capsule development is not clear. Single *gpdhc1* mutants have defects in glycerol metabolism (Shen *et al.*, 2006). *GPDHC1* is expressed in several tissues but its highest expression is in flowers (Shen *et al.*, 2006). Its role in flowers is not yet well understood. In plants, *HDR* synthesizes isopentenyl diphosphate and dimethylallyl diphosphate in the last step of the methylerythritol phosphate pathway (MEP) (Hsieh & Hsieh, 2015). The MEP pathway occurs in the chloroplast and it is the main generator of cytokinins, monoterpenes, tocopherols, chlorophylls, carotenoids, gibberellins, phytoalexins, and others (Vranová *et al.*, 2013). *IQD5* is a microtubule-associated protein involved in microtubule organization. Single *iqd5* mutants have misshapen pavement cells (Liang *et al.*, 2018). *PLC2* mutants were sterile and had elevated levels of auxin in flowers (Li *et al.*, 2015). *RAS1* is involved in salt tolerance and *ABA* sensitivity in *Arabidopsis* (Ren *et al.*, 2010). *TOPII* is highly expressed in flowers of *Arabidopsis* when compared to leaves (Hu *et al.*, 2003). One of its functions is resolving interlocks between homologous chromosomes undergoing synapsis during meiosis (Martinez-Garcia *et al.*, 2018).

Thus, further protein domain and phylogenetic analyses are needed to identify the correct homologs in *Eucalyptus* for many of the genes identified during our differential-expression, expression pattern clustering, and gene ontology analyses. Co-expression analysis of the genes identified should help understand their function better in *Eucalyptus*.

Eucalyptus trees are praised for their rapid growth, drought tolerance and pathogen resistance, and their secondary metabolites. *Eucalyptus* flowers and seed capsules also are highly resistant to biotic attack and abiotic pressures. This transcriptome elucidates genes that play a functional role in producing these phenotypes. Further research is needed to fully understand the biological processes important to capsule development, maturation, and defense.

6 Conclusions

This dissertation documents the first time essential flowering genes have been mutagenized with CRISPR Cas nucleases in *Populus* and *Eucalyptus*. The broad goal was to study the feasibility of creating robust and complete genetic containment by destroying the gene function of these essential flowering genes. We had average mutation rates between 75% and 95% that translated to average loss-of-function rates of 60% to 90% in poplar and eucalypts respectively, with 0% of offtarget mutagenesis reported in several homologous loci. Most of our knockout mutants were completely sterile with no reproductive organs at all. However, emergence of underdeveloped reproductive-like organs occurred in some events of our *ELFY* eucalypt knockouts, which is of concern. Field studies of our knockout mutants, in non-early flowering clones, are essential to determine if the sterile phenotype is permanent and to fully understand the gross involvement of these genes in flowering on-set, reproductive organ development, growth, and leaf traits of forest trees.

Ideally, to create a robust genetic containment system, it would be best to target more than one essential flowering or reproductive gene. Targeting more than one gene would guarantee that in case of a second site mutation recovering the knocked-out gene function, viable gametes would not be produced. With CRISPR Cas nucleases, it is possible to target more than one gene with a single nuclease as long as the genes have homologous sections. Nonetheless, if there are no homologous sections, one could use a vector with multiple sgRNAs and have one to two sgRNAs target each gene of interest.

CRISPR Cas technologies have incredible potential in the area of genetic modification. Once sgRNA design is well understood, it is easy to imagine that mutagenesis rates of ~100% will be the norm. Reverse genetics with CRISPR will likely become as common and routine as PCR in molecular labs of all kingdoms. Sequence data and tissue culture systems will likely become the most difficult hurdles for generating knockouts in less common plants species.

To implement CRISPR Cas techniques, the plants of choice must have enough sequence data to design highly specific and active sgRNAs. The plants must also have tissue

culture, transformation, and regeneration systems in place, or at the least, an *in-planta* transformation system. Many species have systems but usually its just one or a few varieties per species (we see this in both *Populus* and *Eucalyptus*). While other varieties, and many times the varieties used commercially, are recalcitrant to tissue culture, transformation, and/or regeneration techniques. Understanding what makes some varieties of plants recalcitrant to these techniques is a highly active area of research at the moment. More and more plants are having their genomes and transcriptomes sequenced and a lot of related species already have data.

With all the information becoming easily accessible and the tools for implementing CRISPR mutagenesis becoming easier to implement and more inexpensive, the goal of a completely sterile forest tree is just around the corner. Our lab is currently testing other genes that might render our trees sterile, so it is not hard to decide to target two or moth of those genes simultaneously to get a safeguarded sterile clone.

On the other hand, genetic containment technologies have never been deregulated for commercial use. All the backlask originated from Monsanto possibly using GeneSafe Technologies (i.e., GURT with patent) has had a dramatic effect on the public trust on genetic containment and on genetic engineering in general. Brasil, which grows thousands of acres of hybrid eucalyptus for pulp, paper, and biocellulos, has a law that forbids the commercial use of any V-GURTs and the Forest Stewardship Council (FSC) does not yet endorse paper produced from genetically modified trees

Forest trees plantations not only provide source material for pulp, paper, fuel, and biocellulose, but they also support other biota that are beneficial including food and shelter for wildlife, cultural enrichment, carbon sequestration, etc. Thus, completely removing flowers via genetic means reduces the food sources for birds, insects, and small mammals. This is why targeting genes more expecific to embryo viability might be less detrimental to those benefits from the reproducing tissues of trees. This might be even more relevant now that our climate is changing and the extreme weather patterns will likely affect the physiology of plants significantly. Trees might have lower flower and fruit set, not only because of the warmer weather, but because their phenological

calendars will change. We might have to use CRISPR Cas techniques to modify specific genes and help crops adapt to their newer environments.

Choosing what gene to target involves not only genomic data and homology analysis to model plant species (i.e., *Arabidopsis* or *antirrhinum*), but also expression data. Given significant chromosomal events i.e. chromosomal duplication and/or functional divergence, expression information is necessary to determine which gene is the true homolog and will likely lead to a similar phenotype seen in *Arabidopsis* or *antirrhinum*. Our seed capsule and pollen transcriptome will help define genes in the *Eucalyptus grandis* genome and will also add more understanding to the evolutionary history of the Myrtaceae.

Bibliography

- Aarts MG, Keijzer CJ, Stiekema WJ, Pereira A. 1995.** Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *The Plant Cell* **7**: 2115–2127.
- Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DBT, Kellner MJ, Regev A, et al. 2017.** RNA targeting with CRISPR–Cas13. *Nature* **550**: 280–284.
- Ahearn KP, Johnson HA, Weigel D, Wagner DR. 2001.** NFL1, a Nicotiana tabacum LEAFY-Like Gene, Controls Meristem Initiation and Floral Structure. *Plant and Cell Physiology* **42**: 1130–1139.
- Andrés F, Coupland G. 2012.** The genetic basis of flowering responses to seasonal cues. *Nature Reviews Genetics* **13**: 627–639.
- Angenent GC, Colombo L. 1996.** Molecular control of ovule development. *Trends in Plant Science* **1**: 228–232.
- Anonymous. 2015.** Brazil approves transgenic eucalyptus. *Nature Biotechnology* **33**: 577–577.
- Armstrong JA. 1979.** Biotic pollination mechanisms in the Australian flora — a review. *New Zealand Journal of Botany* **17**: 467–508.
- Ault K, Viswanath V, Jayawickrama J, Ma C, Eaton J, Meilan R, Beauchamp G, Hohenschuh W, Murthy G, H. Strauss S. 2016.** Improved growth and weed control of glyphosate-tolerant poplars. *New Forests* **47**: 653–667.
- Avery OT, MacLeod CM, McCarty M. 1944.** Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *Journal of Experimental Medicine* **79**: 137–158.
- Backman TWH, Girke T. 2016.** systemPipeR: NGS workflow and report generation environment. *BMC Bioinformatics* **17**: 388.
- Bae S, Park J, Kim J-S. 2014.** Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**: 1473–1475.
- Balanzà V, Roig-Villanova I, Marzo MD, Masiero S, Colombo L. 2016.** Seed abscission and fruit dehiscence required for seed dispersal rely on similar genetic networks. *Development* **143**: 3372–3381.
- Bamburg JR, Bernstein BW. 2008.** ADF/cofilin. *Current biology: CB* **18**: R273-275.

Baranski R, Klimek-Chodacka M, Lukasiewicz A. 2019. Approved genetically modified (GM) horticultural plants: A 25-year perspective. *Folia Horticulturae; Kraków* **31**: 3–49.

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V. 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* **9**: 39.

Belhaj K, Chaparro-Garcia A, Kamoun S, Patron NJ, Nekrasov V. 2015. Editing plant genomes with CRISPR/Cas9. *Current Opinion in Biotechnology* **32**: 76–84.

Bemer M, Karlova R, Ballester AR, Tikunov YM, Bovy AG, Wolters-Arts M, Rossetto P de B, Angenent GC, de Maagd RA. 2012. The Tomato FRUITFULL Homologs TDR4/FUL1 and MBP7/FUL2 Regulate Ethylene-Independent Aspects of Fruit Ripening[W]. *The Plant Cell* **24**: 4437–4451.

Bertani G. 1951. STUDIES ON LYSOGENESIS I. *Journal of Bacteriology* **62**: 293–300.

Bewg WP, Ci D, Tsai C-J. 2018a. Genome Editing in Trees: From Multiple Repair Pathways to Long-Term Stability. *Frontiers in Plant Science* **9**.

Bewg WP, Ci D, Tsai C-J. 2018b. Genome Editing in Trees: From Multiple Repair Pathways to Long-Term Stability. *Frontiers in Plant Science* **9**.

Biffin E, Lucas EJ, Craven LA, Ribeiro da Costa I, Harrington MG, Crisp MD. 2010. Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. *Annals of Botany* **106**: 79–93.

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O. 2006. CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. *Science* **312**: 1040–1043.

Borner R, Kampmann G, Chandler J, Gleißner R, Wisman E, Apel K, Melzer S. 2000. A MADS domain gene involved in the transition to flowering in Arabidopsis. *The Plant Journal* **24**: 591–599.

Bortesi L, Zhu C, Zischewski J, Perez L, Bassié L, Nadi R, Forni G, Lade SB, Soto E, Jin X, et al. 2016. Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnology Journal* **14**: 2203–2216.

Bou Daher F, Geitmann A. 2012. Actin depolymerizing factors ADF7 and ADF10 play distinct roles during pollen development and pollen tube growth. *Plant Signaling & Behavior* **7**: 879–881.

Bou Daher F, van Oostende C, Geitmann A. 2011. Spatial and Temporal Expression of Actin Depolymerizing Factors ADF7 and ADF10 during Male Gametophyte Development in *Arabidopsis thaliana*. *Plant and Cell Physiology* **52**: 1177–1192.

Bouché F, Lobet G, Tocquin P, Périlleux C. 2016. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Research* **44**: D1167–D1171.

Bowman JL, Alvarez J, Weigel D, Meyerowitz EM, Smyth DR. 1993. Control of flower development in *Arabidopsis thaliana* by APETALA1 and interacting genes. *Development* **119**: 721–743.

Bowman JL, Smyth DR, Meyerowitz EM. 1989a. Genes Directing Flower Development in *Arabidopsis*. *The Plant Cell Online* **1**: 37–52.

Bowman JL, Smyth DR, Meyerowitz EM. 1989b. Genes directing flower development in *Arabidopsis*. *The Plant Cell* **1**: 37–52.

Bowman JL, Smyth DR, Meyerowitz EM. 1991. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1–20.

Braun AC. 1947. Thermal Studies on the Factors Responsible for Tumor Initiation in Crown Gall. *American Journal of Botany* **34**: 234–240.

Brinster RL, Chen HY, Trumbauer M, Senear AW, Warren R, Palmiter RD. 1981. Somatic Expression of Herpes Thymidine Kinase in Mice following Injection of a Fusion Gene into Eggs. *Cell* **27**: 223–231.

Brisson N, Paszkowski J, Penswick JR, Gronenborn B, Potrykus I, Hohn T. 1984. Expression of a bacterial gene in plants by using a viral vector. *Nature* **310**: 511–514.

Brockerhoff EG, Jactel H, Parrotta JA, Quine CP, Sayer J. 2008. Plantation forests and biodiversity: oxymoron or opportunity? *Biodiversity and Conservation* **17**: 925–951.

Brown H. 1906. On the culture of the excised embryos of barley on nutrient solutions containing nitrogen in different forms. *Trans. Guinness Res. Lab* **1**: 288–299.

Brunner AM, Li J, DiFazio SP, Shevchenko O, Montgomery BE, Mohamed R, Wei H, Ma C, Elias AA, VanWormer K, et al. 2007. Genetic containment of forest plantations. *Tree Genetics & Genomes* **3**: 75–100.

Brunner AM, Rottmann WH, Sheppard LA, Krutovskii K, DiFazio SP, Leonardi S, Strauss SH. 2000. Structure and expression of duplicate AGAMOUS orthologues in poplar. *Plant Molecular Biology* **44**: 619–634.

Butler NM, Atkins PA, Voytas DF, Douches DS. 2015. Generation and Inheritance of Targeted Mutations in Potato (*Solanum tuberosum* L.) Using the CRISPR/Cas System. *PLoS ONE* **10**.

Carr DJ, Carr SGM. 1959. Floral Morphology and the Taxonomy of Eucalyptus. *Nature* **184**: 1549–1552.

Carrière Y, Dutilleul P, Ellers-Kirk C, Pedersen B, Haller S, Antilla L, Dennehy TJ, Tabashnik BE. 2004. SOURCES, SINKS, AND THE ZONE OF INFLUENCE OF REFUGES FOR MANAGING INSECT RESISTANCE TO Bt CROPS. *Ecological Applications* **14**: 1615–1623.

Causier B, Schwarz-Sommer Z, Davies B. 2010. Floral organ identity: 20 years of ABCs. *Seminars in Cell & Developmental Biology* **21**: 73–79.

Cha RS, Zarbl H, Keohavong P, Thilly WG. 1992. Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. *Genome Research* **2**: 14–20.

Chauhan R, Veale A, Cathleen M, Strauss S, Myburg A. 2014. Genetic Transformation of Eucalyptus—Challenges and Future Prospects. In: Ramawat K, Mérillon J-M, Ahuja M, eds. *Tree Biotechnology*. CRC Press.

Chen K, Gao C. 2014. Targeted genome modification technologies and their applications in crop improvements. *Plant Cell Reports* **33**: 575–583.

Chen Z, Rao P, Yang X, Su X, Zhao T, Gao K, Yang X, An X. 2018a. A Global View of Transcriptome Dynamics During Male Floral Bud Development in *Populus tomentosa*. *Scientific Reports* **8**.

Chen R, Xu Q, Liu Y, Zhang J, Ren D, Wang G, Liu Y. 2018b. Generation of Transgene-Free Maize Male Sterile Lines Using the CRISPR/Cas9 System. *Frontiers in Plant Science* **9**.

Chen F, Zhang X, Liu X, Zhang L. 2017. Evolutionary Analysis of MIKCC-Type MADS-Box Genes in Gymnosperms and Angiosperms. *Frontiers in Plant Science* **8**: 895.

Chilton M-D, Saiki RK, Yadav N, Gordon MP, Quetier F. 1980. T-DNA from *Agrobacterium* Ti Plasmid is in the Nuclear DNA Fraction of Crown Gall Tumor Cells. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 4060–4064.

Cho H-J, Kim S, Kim M, Kim B-D. 2001. Production of Transgenic Male Sterile Tobacco Plants with the cDNA Encoding a Ribosome Inactivating Protein in *Dianthus sinensis* L. *Molecules and Cells* **11**: 326–333.

- Chrimes D, Rogers HJ, Francis D, Jones HD, Ainsworth C. 2005.** Expression of fission yeast *cdc25* driven by the wheat ADP-glucose pyrophosphorylase large subunit promoter reduces pollen viability and prevents transmission of the transgene in wheat. *New Phytologist* **166**: 185–192.
- CIRAD-FRA, IUFRO-AUT, MUSE-FRA. 2018.** *Eucalyptus 2018: Managing Eucalyptus plantation under global changes. Abstracts book.* Montpellier, France: CIRAD.
- Coen ES, Meyerowitz EM. 1991.** The war of the whorls: genetic interactions controlling flower development. *Nature* **353**: 31–37.
- Coen ES, Romero JoséM, Doyle S, Elliott R, Murphy G, Carpenter R. 1990a.** *floricaula*: A homeotic gene required for flower development in *antirrhinum majus*. *Cell* **63**: 1311–1322.
- Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R. 1990b.** *floricaula*: A homeotic gene required for flower development in *antirrhinum majus*. *Cell* **63**: 1311–1322.
- Cohen SN, Chang ACY, Boyer HW, Helling RB. 1973.** Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences* **70**: 3240–3244.
- Cox DBT, Gootenberg JS, Abudayyeh OO, Franklin B, Kellner MJ, Joung J, Zhang F. 2017.** RNA editing with CRISPR-Cas13. *Science* **358**: 1019–1027.
- Criswell JG, Shibles RM. 1971.** Physiological Basis for Genotypic Variation in Net Photosynthesis of Oat Leaves 1. *Crop Science* **11**: 550–553.
- Crous J, Sale G, Naidoo T. 2019.** The influence of species, tree improvement and cultural practices on rotation-end fibre production of Eucalyptus pulpwood plantations in South Africa. *Southern Forests: a Journal of Forest Science* **81**: 307–317.
- Crowley TM, Muralitharan MS, Stevenson TW. 2003.** Isolating conifer DNA: A superior polysaccharide elimination method. *Plant Molecular Biology Reporter* **21**: 97–97.
- Dall'Osto L, Cazzaniga S, Zappone D, Bassi R. 2020.** Monomeric light harvesting complexes enhance excitation energy transfer from LHCII to PSII and control their lateral spacing in thylakoids. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1861**: 148035.
- Danilevskaya ON, Meng X, Ananiev EV. 2010.** Concerted Modification of Flowering Time and Inflorescence Architecture by Ectopic Expression of TFL1-Like Genes in Maize. *Plant Physiology* **153**: 238–251.

- Davis GL. 1969.** Floral morphology and the development of the gametophytes in *Eucalyptus stellulata* Sieb. *Australian Journal of Botany* **17**: 177–190.
- Davis AR. 1997.** Influence of floral visitation on nectar-sugar composition and nectary surface changes in *Eucalyptus*. *Apidologie* **28**: 27–42.
- De Block M, Herrera-Estrella L, Van Montagu M, Schell J, Zambryski P. 1984.** Expression of foreign genes in regenerated plants and in their progeny. *The EMBO Journal* **3**: 1681–1689.
- De Greve H, Decraemer H, Seurinck J, Van Montagu M, Schell J. 1981.** The functional organization of the octopine *Agrobacterium tumefaciens* plasmid pTiB6S3. *Plasmid* **6**: 235–248.
- Demirci Y, Zhang B, Unver T. 2017.** CRISPR/Cas9: An RNA-guided highly precise synthetic tool for plant genome editing. *Journal of Cellular Physiology*.
- Denness L, McKenna JF, Segonzac C, Wormit A, Madhou P, Bennett M, Mansfield J, Zipfel C, Hamann T. 2011.** Cell Wall Damage-Induced Lignin Biosynthesis Is Regulated by a Reactive Oxygen Species- and Jasmonic Acid-Dependent Process in *Arabidopsis*1[C][W][OA]. *Plant Physiology* **156**: 1364–1374.
- Dhaese P, De Greve H, Decraemer H, Schell J, Van Montagu M. 1979.** Rapid mapping of transposon insertion and deletion mutations in the large Ti-plasmids of *Agrobacterium tumefaciens*. *Nucleic Acids Research* **7**: 1837–1849.
- Diggle PK, Di Stilio VS, Gschwend AR, Golenberg EM, Moore RC, Russell JRW, Sinclair JP. 2011.** Multiple developmental processes underlie sex differentiation in angiosperms. *Trends in Genetics* **27**: 368–376.
- Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. 2004.** The SEP4 Gene of *Arabidopsis thaliana* Functions in Floral Organ and Meristem Identity. *Current Biology* **14**: 1935–1940.
- Dodet M, Collet C. 2012.** When should exotic forest plantation tree species be considered as an invasive threat and how should we treat them? *Biological Invasions* **14**: 1765–1778.
- Donaldson JE, Hui C, Richardson DM, Robertson MP, Webber BL, Wilson JRU. 2014.** Invasion trajectory of alien trees: the role of introduction pathway and planting history. *Global Change Biology* **20**: 1527–1537.
- van Doorn WG. 1997.** Effects of pollination on floral attraction and longevity. *Journal of Experimental Botany* **48**: 1615–1622.
- van Doorn WG, Woltering EJ. 2008.** Physiology and molecular biology of petal senescence. *Journal of Experimental Botany* **59**: 453–480.

- Dornelas MC, Amaral WAN do, Rodriguez APM. 2004.** EgLFY, the Eucalyptus grandis homolog of the Arabidopsis gene LEAFY is expressed in reproductive and vegetative tissues. *Brazilian Journal of Plant Physiology* **16**: 105–114.
- Dornhoff GM, Shibles RM. 1970.** Varietal Differences in Net Photosynthesis of Soybean Leaves 1. *Crop Science* **10**: 42–45.
- Duke SO, Powles SB. 2008.** Glyphosate: a once-in-a-century herbicide. *Pest Management Science* **64**: 319–325.
- Elorriaga E, Klocko AL, Ma C, Strauss SH. 2018.** Variation in Mutation Spectra Among CRISPR/Cas9 Mutagenized Poplars. *Frontiers in Plant Science* **9**.
- Enders TA, Frick EM, Strader LC. 2017.** An Arabidopsis kinase cascade influences auxin-responsive cell expansion. *The Plant Journal: For Cell and Molecular Biology* **92**: 68–81.
- Endo T, Shimada T, Fujii H, Kobayashi Y, Araki T, Omura M. 2005.** Ectopic Expression of an FT Homolog from Citrus Confers an Early Flowering Phenotype on Trifoliolate Orange (*Poncirus trifoliata* L. Raf.). *Transgenic Research* **14**: 703–712.
- Escalante-Pérez M, Jaborsky M, Lautner S, Fromm J, Müller T, Dittrich M, Kunert M, Boland W, Hedrich R, Ache P. 2012.** Poplar Extrafloral Nectaries: Two Types, Two Strategies of Indirect Defenses against Herbivores. *Plant Physiology* **159**: 1176–1191.
- Etchells JP, Mishra LS, Kumar M, Campbell L, Turner SR. 2015.** Wood Formation in Trees Is Increased by Manipulating PXY-Regulated Cell Division. *Current biology: CB* **25**: 1050–1055.
- Ezquer I, Mizzotti C, Nguema-Ona E, Gotté M, Beauzamy L, Viana VE, Dubrulle N, Oliveira AC de, Caporali E, Koroney A-S, et al. 2016.** The Developmental Regulator SEEDSTICK Controls Structural and Mechanical Properties of the Arabidopsis Seed Coat. *The Plant Cell* **28**: 2478–2492.
- Fan D, Liu T, Li C, Jiao B, Li S, Hou Y, Luo K. 2015.** Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in Populus in the First Generation. *Scientific Reports* **5**: 12217.
- FAO. 2010.** *Forests and genetically modified trees*. Rome, Italy: FAO.
- FAO, WFP, IFAD. 2012.** *The state of food insecurity in the world 2012: Economic growth is necessary but not sufficient to accelerate reduction of hunger and malnutrition*. Rome, Italy: FAO.
- Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L. 2003.** MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *The Plant Cell* **15**: 2603–2611.

- Feng Y, Liu Q, Xue Q. 2006.** Comparative study of rice and Arabidopsis actin-depolymerizing factors gene families. *Journal of Plant Physiology* **163**: 69–79.
- Feng Z, Mao Y, Xu N, Zhang B, Wei P, Yang D-L, Wang Z, Zhang Z, Zheng R, Yang L, et al. 2014.** Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences* **111**: 4632–4637.
- Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, Cao F, Zhu S, Zhang F, Mao Y, et al. 2013.** Efficient genome editing in plants using a CRISPR/Cas system. *Cell Research* **23**: 1229–1232.
- Ferrandiz C, Gu Q, Martienssen R, Yanofsky MF. 2000.** Redundant regulation of meristem identity and plant architecture by FRUITFULL, APETALA1 and CAULIFLOWER. *Development* **127**: 725–734.
- Fleming D, Musser F, Reisig D, Greene J, Taylor S, Parajulee M, Lorenz G, Catchot A, Gore J, Kerns D, et al. 2018.** Effects of transgenic *Bacillus thuringiensis* cotton on insecticide use, heliothine counts, plant damage, and cotton yield: A meta-analysis, 1996-2015. *PLoS ONE* **13**.
- Fonfara I, Richter H, Bratovič M, Le Rhun A, Charpentier E. 2016.** The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**: 517–521.
- Ford HA, Paton DC, Forde N. 1979.** Birds as pollinators of Australian plants. *New Zealand Journal of Botany* **17**: 509–519.
- Fornalé S, Lopez E, Salazar-Henao JE, Fernández-Nohales P, Rigau J, Caparros-Ruiz D. 2014.** AtMYB7, a New Player in the Regulation of UV-Sunscreens in Arabidopsis thaliana. *Plant and Cell Physiology* **55**: 507–516.
- Fornalé S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres J-L, Rovira P, Puigdomènech P, et al. 2010.** ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *The Plant Journal* **64**: 633–644.
- Fraser CM, Chapple C. 2011.** The Phenylpropanoid Pathway in Arabidopsis. *The Arabidopsis Book / American Society of Plant Biologists* **9**.
- Frazer AC, Perteza G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. 2015.** Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature biotechnology* **33**: 243–246.
- Friedrichs S, Takasu Y, Kearns P, Dagallier B, Oshima R, Schofield J, Moreddu C. 2019.** An overview of regulatory approaches to genome editing in agriculture. *Biotechnology Research and Innovation* **3**: 208–220.

Fromm M, Taylor LP, Walbot V. 1985. Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proceedings of the National Academy of Sciences of the United States of America* **82**: 5824–5828.

Fromm ME, Taylor LP, Walbot V. 1986. Stable transformation of maize after gene transfer by electroporation. *Nature* **319**: 791–793.

Futschik ME, Carlisle B. 2005. Noise-robust soft clustering of gene expression time-course data. *Journal of Bioinformatics and Computational Biology* **3**: 965–988.

Gamfeldt L, Snäll T, Bagchi R, Jonsson M, Gustafsson L, Kjellander P, Ruiz-Jaen MC, Fröberg M, Stendahl J, Philipson CD, et al. 2013. Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications* **4**: 1340.

Gao B, Chen M, Li X, Zhang J. 2019a. Ancient duplications and grass-specific transposition influenced the evolution of LEAFY transcription factor genes. *Communications Biology* **2**.

Gao B, Chen M, Li X, Zhang J. 2019b. Ancient duplications and grass-specific transposition influenced the evolution of LEAFY transcription factor genes. *Communications Biology* **2**: 237.

Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ. 2016. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Research* **26**: 901–913.

Gautheret R-J. 1934. *Culture du tissu cambial*.

Gautheret R. 1939. Sur la possibilité de réaliser la culture indéfinie des tissus de tubercules de carotte. *CR Hebd. Seances Acad. Sc* **208**: 118–120.

Ghogare R, Williamson-Benavides B, Ramírez-Torres F, Dhingra A. 2019. CRISPR-associated nucleases: the Dawn of a new age of efficient crop improvement. *Transgenic Research*.

Giaretta A, Vasconcelos TNC, Mazine FF, Faria JEQ, Flores R, Holst B, Sano PT, Lucas E. 2019. Calyx (con)fusion in a hyper-diverse genus: Parallel evolution of unusual flower patterns in *Eugenia* (Myrtaceae). *Molecular Phylogenetics and Evolution* **139**: 106553.

Gill AM, Brooker MIH, Moore PHR. 1992. Seed Weights and Numbers as a Function of Fruit Size and Subgenus in Some *Eucalyptus* Species From South-Western Australia. *Australian Journal of Botany* **40**: 103–111.

Goetz M, Godt DE, Guivarc’h A, Kahmann U, Chriqui D, Roitsch T. 2001. Induction of male sterility in plants by metabolic engineering of the carbohydrate supply.

Proceedings of the National Academy of Sciences of the United States of America **98**: 6522–6527.

Goldbeck JC, do Nascimento JE, Jacob RG, Fiorentini ÂM, da Silva WP. 2014. Bioactivity of essential oils from *Eucalyptus globulus* and *Eucalyptus urograndis* against planktonic cells and biofilms of *Streptococcus mutans*. *Industrial Crops and Products* **60**: 304–309.

Gonçalves JL de M, Alvares CA, Higa AR, Silva LD, Alfenas AC, Stahl J, Ferraz SF de B, Lima W de P, Brancalion PHS, Hubner A, et al. 2013. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. *Forest Ecology and Management* **301**: 6–27.

Goslin K, Zheng B, Serrano-Mislata A, Rae L, Ryan PT, Kwaśniewska K, Thomson B, Ó'Maoiléidigh DS, Madueño F, Wellmer F, et al. 2017. Transcription Factor Interplay between LEAFY and APETALA1/CAULIFLOWER during Floral Initiation. *Plant Physiology* **174**: 1097–1109.

Gramzow L, Theissen G. 2010. A hitchhiker's guide to the MADS world of plants. *Genome Biology* **11**.

Griffin AR, Hingston AB, Ohmart CP. 2009. Pollinators of *Eucalyptus regnans* (Myrtaceae), the world's tallest flowering plant species. *Australian Journal of Botany* **57**: 18–25.

Gu Q, Ferrandiz C, Yanofsky MF, Martienssen R. 1998. The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* **125**: 1509–1517.

Guerineau F, Sorensen A-M, Fenby N, Scott RJ. 2003. Temperature sensitive diphtheria toxin confers conditional male-sterility in *Arabidopsis thaliana*. *Plant Biotechnology Journal* **1**: 33–42.

Ha CM, Escamilla-Trevino L, Yarce JCS, Kim H, Ralph J, Chen F, Dixon RA. 2016. An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in *Medicago truncatula*. *The Plant Journal* **86**: 363–375.

Haberlandt G. 1902. Culturversuche mit isolierten pflanzenzellen, sitz-ungsab. *Akad. D. wissensch. Mathematisch-naturwissenschaftlicher c169*.

Hallerman E, Grabau E. 2016. Crop biotechnology: a pivotal moment for global acceptance. *Food and Energy Security* **5**: 3–17.

Hamès C, Ptchelkine D, Grimm C, Thevenon E, Moyroud E, Gérard F, Martiel J-L, Benlloch R, Parcy F, Müller CW. 2008. Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *The EMBO Journal* **27**: 2628–2637.

- Hammer RE, Pursel VG, Rexroad CE, Wall RJ, Bolt DJ, Ebert KM, Palmiter RD, Brinster RL. 1985.** Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* **315**: 680–683.
- Hannig EE. 1904a.** *Ueber die cultur von cruciferen-embryonen ausserhalb des embryosacks.*
- Hannig EE. 1904b.** *Zur Physiologie pflanzlicher Embryonen.* A. Felix.
- Hannoufa A, McNevin J, Lemieux B. 1993.** Epicuticular waxes of eceriferum mutants of *Arabidopsis thaliana*. *Phytochemistry* **33**: 851–855.
- Harfouche A, Meilan R, Altman A. 2011.** Tree genetic engineering and applications to sustainable forestry and biomass production. *Trends in Biotechnology* **29**: 9–17.
- Heap I, Duke SO. 2018.** Overview of glyphosate-resistant weeds worldwide. *Pest Management Science* **74**: 1040–1049.
- Hernalsteens JP, De Greve H, Van Montagu M, Schell J. 1978.** Mutagenesis by insertion of the drug resistance transposon Tn7 applied to the Ti plasmid of *Agrobacterium tumefaciens*. *Plasmid* **1**: 218–225.
- Hingston AB, McQuillan PB, Potts BM. 2004a.** Pollinators in seed orchards of *Eucalyptus nitens* (Myrtaceae). *Australian Journal of Botany* **52**: 209–222.
- Hingston AB, Potts BM, McQuillan PB. 2004b.** The swift parrot *Lathamus discolor* (Psittacidae), social bees (Apidae), and native insects as pollinators of *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **52**: 371–379.
- Hingston AB, Potts BM, McQuillan PB. 2004c.** Pollination services provided by various size classes of flower visitors to *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **52**: 353–369.
- Hofer J, Turner L, Hellens R, Ambrose M, Matthews P, Michael A, Ellis N. 1997.** UNIFOLIATA regulates leaf and flower morphogenesis in pea. *Current Biology* **7**: 581–587.
- Höfig KP, Möller R, Donaldson L, Putterill J, Walter C. 2006.** Towards male sterility in *Pinus radiata*—a stilbene synthase approach to genetically engineer nuclear male sterility. *Plant Biotechnology Journal* **4**: 333–343.
- Holsters M, Silva B, Van Vliet F, Genetello C, De Block M, Dhaese P, Depicker A, Inzé D, Engler G, Villarreal R, et al. 1980.** The functional organization of the nopaline *A. tumefaciens* plasmid pTiC58. *Plasmid* **3**: 212–230.

- Howe GT, Yu J, Knaus B, Cronn R, Kolpak S, Dolan P, Lorenz WW, Dean JF. 2013a.** A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* **14**.
- Howe GT, Yu J, Knaus B, Cronn R, Kolpak S, Dolan P, Lorenz WW, Dean JF. 2013b.** A SNP resource for Douglas-fir: de novo transcriptome assembly and SNP detection and validation. *BMC Genomics* **14**: 137.
- Hsieh W-Y, Hsieh M-H. 2015.** The amino-terminal conserved domain of 4-hydroxy-3-methylbut-2-enyl diphosphate reductase is critical for its function in oxygen-evolving photosynthetic organisms. *Plant Signaling & Behavior* **10**: e988072.
- Hsu C-Y, Adams JP, Kim H, No K, Ma C, Strauss SH, Drnevich J, Vandervelde L, Ellis JD, Rice BM, et al. 2011.** FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 10756–10761.
- Hu W, Wang Y, Bowers C, Ma H. 2003.** Isolation, sequence analysis, and expression studies of florally expressed cDNAs in Arabidopsis. *Plant Molecular Biology* **53**: 545–563.
- Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou Y-H, Yu J-Q, Chen Z. 2010.** Functional Analysis of the Arabidopsis PAL Gene Family in Plant Growth, Development, and Response to Environmental Stress. *Plant Physiology* **153**: 1526–1538.
- Hurley BA, Tran HT, Marty NJ, Park J, Snedden WA, Mullen RT, Plaxton WC. 2010.** The dual-targeted purple acid phosphatase isozyme AtPAP26 is essential for efficient acclimation of Arabidopsis to nutritional phosphate deprivation. *Plant Physiology* **153**: 1112–1122.
- Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh J-RJ, Joung JK. 2013.** Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* **31**: 227–229.
- Hwang IS, Hwang BK. 2010.** The Pepper 9-Lipoxygenase Gene CaLOX1 Functions in Defense and Cell Death Responses to Microbial Pathogens. *Plant Physiology* **152**: 948–967.
- Imadi SR, Mahmood I, Kazi AG. 2014.** Bamboo Fiber Processing, Properties, and Applications. In: Hakeem KR, Jawaid M, Rashid U, eds. Biomass and Bioenergy: Processing and Properties. Cham: Springer International Publishing, 27–46.
- Immink RGH, Kaufmann K, Angenent GC. 2010.** The ‘ABC’ of MADS domain protein behaviour and interactions. *Seminars in Cell & Developmental Biology* **21**: 87–93.
- Inc G. 2007.** Nutrition and Food. *Gallup.com*.

ISAAA. 2017. Global Status of Commercialized Biotech/GM Crops in 2017: Biotech crop adoption surges as economic benefits accumulate in 22 years.

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA. 2015. Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnology* **15**: 16.

Jaeger KE, Pullen N, Lamzin S, Morris RJ, Wigge PA. 2013. Interlocking Feedback Loops Govern the Dynamic Behavior of the Floral Transition in Arabidopsis. *The Plant Cell* **25**: 820–833.

Jain SM, Gupta PK, Newton RJ. 2013. *Somatic Embryogenesis in Woody Plants*. Springer Science & Business Media.

James C. 2015. *20th anniversary (1996 to 2015) of the global commercialization of biotech crops and biotech crop highlights in 2015*. International Service for the Acquisition of Agri-Biotech Applications.

James C, Krattiger AF. 1996. *Global Review of the Field Testing and Commercialization of Transgenic Plants 1986 to 1995: The First Decade of Crop Biotechnology*. International Service for the Acquisition of Agri-biotech Applications (ISAAA).

Jaradat MR, Ruegger M, Bowling A, Butler H, Cutler AJ. 2014. A comprehensive transcriptome analysis of silique development and dehiscence in Arabidopsis and Brassica integrating genotypic, interspecies and developmental comparisons. *GM Crops & Food* **5**: 302–320.

Jeena GS, Kumar S, Shukla RK. 2019. Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Molecular Biology* **100**: 351–365.

Jiang S, Balan B, Assis R de AB, Sagawa CHD, Wan X, Han S, Wang L, Zhang L, Zaini PA, Walawage SL, et al. 2020. Genome-Wide Profiling and Phylogenetic Analysis of the SWEET Sugar Transporter Gene Family in Walnut and Their Lack of Responsiveness to *Xanthomonas arboricola* pv. *juglandis* Infection. *International Journal of Molecular Sciences* **21**: 1251.

Jiang Y, Wang J, Xie Y, Chen N, Huang S. 2017. ADF10 shapes the overall organization of apical actin filaments by promoting their turnover and ordering in pollen tubes. *Journal of Cell Science* **130**: 3988–4001.

Jibran R, Tahir J, Cooney J, Hunter DA, Dijkwel PP. 2017. Arabidopsis AGAMOUS Regulates Sepal Senescence by Driving Jasmonate Production. *Frontiers in Plant Science* **8**: 2101.

Jin H, Cominelli E, Bailey P, Parr A, Mehrrens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000. Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO journal* **19**: 6150–6161.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**: 816–821.

Johnston SA, Riedy M, De Vit MJ, Sanford JC, McElligott S, Williams RS. 1991. Biolistic transformation of animal tissue. *In Vitro Cellular & Developmental Biology - Plant* **27**: 11–14.

Jung J-H, Lee H-J, Ryu JY, Park C-M. 2016. SPL3/4/5 Integrate Developmental Aging and Photoperiodic Signals into the FT-FD Module in Arabidopsis Flowering. *Molecular Plant* **9**: 1647–1659.

Keb-Llanes M, González G, Chi-Manzanero B, Infante D. 2002. A rapid and simple method for small-scale DNA extraction in Agavaceae and other tropical plants. *Plant Molecular Biology Reporter* **20**: 299–299.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**: 907–915.

Klein TM, Harper EC, Svab Z, Sanford JC, Fromm ME, Maliga P. 1988. Stable genetic transformation of intact *Nicotiana* cells by the particle bombardment process. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 8502–8505.

Klein TM, Wolf ED, Wu R, Sanford JC. 1987. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* **327**: 70–73.

Klocko AL, Brunner AM, Huang J, Meilan R, Lu H, Ma C, Morel A, Zhao D, Ault K, Dow M, et al. 2016a. Genetic containment of forest trees by RNAi suppression of *LEAFY*.

Klocko AL, Brunner AM, Huang J, Meilan R, Lu H, Ma C, Morel A, Zhao D, Ault K, Dow M, et al. 2016b. Containment of transgenic trees by suppression of *LEAFY*. *Nature Biotechnology* **34**: 918–922.

Klocko AL, Brunner AM, Huang J, Meilan R, Lu H, Ma C, Morel A, Zhao D, Ault K, Dow M, et al. 2016c. Containment of transgenic trees by suppression of *LEAFY*. *Nature Biotechnology*.

Klocko AL, Ma C, Robertson S, Esfandiari E, Nilsson O, Strauss SH. 2016d. FT overexpression induces precocious flowering and normal reproductive development in *Eucalyptus*. *Plant Biotechnology Journal* **14**: 808–819.

Klümper W, Matin Q. 2014. A Meta-Analysis of the Impacts of Genetically Modified Crops. *PLOS ONE* **9**: e111629.

- Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T. 1999.** A Pair of Related Genes with Antagonistic Roles in Mediating Flowering Signals. *Science* **286**: 1960–1962.
- Kotoda N, Wada M. 2005.** MdTFL1, a TFL1-like gene of apple, retards the transition from the vegetative to reproductive phase in transgenic Arabidopsis. *Plant Science* **168**: 95–104.
- Kotte W. 1922.** Kulturversuche mit isolierten Wurzelspitzen. *Beitr. Allg. Bot* **2**: 413–434.
- Kovács L, Damkjær J, Kereiche S, Illoia C, Ruban AV, Boekema EJ, Jansson S, Horton P. 2006.** Lack of the Light-Harvesting Complex CP24 Affects the Structure and Function of the Grana Membranes of Higher Plant Chloroplasts. *The Plant Cell* **18**: 3106–3120.
- Kovaka S, Zimin AV, Perteu GM, Razaghi R, Salzberg SL, Perteu M. 2019.** Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**: 278.
- Kramer MG, Redenbaugh K. 1994.** Commercialization of a tomato with an antisense polygalacturonase gene: The FLAVR SAVR™ tomato story. *Euphytica* **79**: 293–297.
- Krishnaraj S, Vasil IK. 1995.** Somatic Embryogenesis in Herbaceous Monocots. In: Thorpe TA, ed. *Current Plant Science and Biotechnology in Agriculture. In Vitro Embryogenesis in Plants*. Dordrecht: Springer Netherlands, 417–470.
- Krizek BA, Fletcher JC. 2005.** Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics* **6**: 688–698.
- Kuang R, Chan K-H, Yeung E, Lim BL. 2009.** Molecular and Biochemical Characterization of AtPAP15, a Purple Acid Phosphatase with Phytase Activity, in Arabidopsis. *Plant Physiology* **151**: 199–209.
- Kumar L, E. Futschik M. 2007.** Mfuzz: A software package for soft clustering of microarray data. *Bioinformatics* **2**: 5–7.
- Kumar K, Gambhir G, Dass A, Tripathi AK, Singh A, Jha AK, Yadava P, Choudhary M, Rakshit S. 2020.** Genetically modified crops: current status and future prospects. *Planta* **251**: 91.
- Kuvshinov V, Koivu K, Kanerva A, Pehu E. 2001.** Molecular control of transgene escape from genetically modified plants. *Plant Science* **160**: 517–522.
- Laibach F. 1925.** *Das Taubwerden von Bastardsamen und die künstliche Aufzucht früh absterbender Bastardembryonen.*
- Lakshmanan P, Taji A. 2000.** Somatic Embryogenesis in Leguminous Plants. *Plant Biology* **2**: 136–148.

Lawrence WA, Davies DR. 1985. A method for the microinjection and culture of protoplasts at very low densities. *Plant Cell Reports* **4**: 33–35.

Lawrenson T, Shorinola O, Stacey N, Li C, Østergaard L, Patron N, Uauy C, Harwood W. 2015. Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. *Genome Biology* **16**: 258.

Lee R, Baldwin S, Kenel F, McCallum J, Macknight R. 2013. FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nature Communications* **4**.

Lee Y-H, Chung K-H, Kim H-U, Jin Y-M, Kim H-I, Park B-S. 2003. Induction of male sterile cabbage using a tapetum-specific promoter from Brassica campestris L. ssp. pekinensis. *Plant Cell Reports* **22**: 268–273.

Lee K, Zhang Y, Kleinstiver BP, Guo JA, Aryee MJ, Miller J, Malzahn A, Zarecor S, Lawrence-Dill CJ, Joung JK, et al. 2019. Activities and specificities of CRISPR/Cas9 and Cas12a nucleases for targeted mutagenesis in maize. *Plant Biotechnology Journal* **17**: 362–372.

Leijten W, Koes R, Roobeek I, Frugis G. 2018. Translating Flowering Time from Arabidopsis thaliana to Brassicaceae and Asteraceae Crop Species. *Plants* **7**: 111.

Lemmers M, De Beuckeleer M, Holsters M, Zambryski P, Depicker A, Hernalsteens JP, Van Montagu M, Schell J. 1980. Internal organization, boundaries and integration of Ti-plasmid DNA in nopaline grown gall tumours. *Journal of Molecular Biology* **144**: 353–376.

Li Y, Kim JI, Pysh L, Chapple C. 2015. Four Isoforms of Arabidopsis 4-Coumarate:CoA Ligase Have Overlapping yet Distinct Roles in Phenylpropanoid Metabolism1[OPEN]. *Plant Physiology* **169**: 2409–2421.

Li J-F, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J. 2013. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology* **31**: 688–691.

Liang G, Zhang H, Lou D, Yu D. 2016. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific Reports* **6**: 21451.

Liang H, Zhang Y, Martinez P, Rasmussen CG, Xu T, Yang Z. 2018. The Microtubule-Associated Protein IQ67 DOMAIN5 Modulates Microtubule Dynamics and Pavement Cell Shape1. *Plant Physiology* **177**: 1555–1568.

Lifschitz E, Eshed Y. 2006. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *Journal of Experimental Botany* **57**: 3405–3414.

- Liljegren SJ, Gustafson-Brown C, Pinyopich A, Ditta GS, Yanofsky MF. 1999.** Interactions among APETALA1, LEAFY, and TERMINAL FLOWER1 specify meristem fate. *The Plant Cell* **11**: 1007–1018.
- Limasset P, Cornuet P. 1949.** RECHERCHE DU VIRUS DE LA MOSAIQUE DU TABAC (MARMOR-TABACI HOLMES) DANS LES MERISTEMES DES PLANTES INFECTEES. *COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES* **228**: 1971–1972.
- Lin IW, Sosso D, Chen L-Q, Gase K, Kim S-G, Kessler D, Klinkenberg PM, Gorder MK, Hou B-H, Qu X-Q, et al. 2014.** Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9. *Nature* **508**: 546–549.
- Liu X, Grabherr HM, Willmann R, Kolb D, Brunner F, Bertsche U, Kühner D, Franz-Wachtel M, Amin B, Felix G, et al. 2014.** Host-induced bacterial cell wall decomposition mediates pattern-triggered immunity in Arabidopsis. *eLife* **3**.
- Liu Y, He C. 2016.** Regulation of plant reactive oxygen species (ROS) in stress responses: learning from AtRBOHD. *Plant Cell Reports* **35**: 995–1007.
- Liu X, Kim YJ, Müller R, Yumul RE, Liu C, Pan Y, Cao X, Goodrich J, Chen X. 2011.** AGAMOUS Terminates Floral Stem Cell Maintenance in Arabidopsis by Directly Repressing WUSCHEL through Recruitment of Polycomb Group Proteins. *The Plant Cell* **23**: 3654–3670.
- Liu L, Li X, Ma J, Li Z, You L, Wang J, Wang M, Zhang X, Wang Y. 2017a.** The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**: 714-726.e10.
- Liu L, Li X, Wang J, Wang M, Chen P, Yin M, Li J, Sheng G, Wang Y. 2017b.** Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**: 121-134.e12.
- Liu Z, Mara C. 2010.** Regulatory mechanisms for floral homeotic gene expression. *Seminars in Cell & Developmental Biology* **21**: 80–86.
- Liu W, Xie X, Ma X, Li J, Chen J, Liu Y-G. 2015.** DSDecode: A Web-Based Tool for Decoding of Sequencing Chromatograms for Genotyping of Targeted Mutations. *Molecular Plant* **8**: 1431–1433.
- Lopez-Obando M, Hoffmann B, Géry C, Guyon-Debast A, Téoulé E, Rameau C, Bonhomme S, Nogué F. 2016.** Simple and Efficient Targeting of Multiple Genes Through CRISPR-Cas9 in *Physcomitrella patens*. *G3: Genes/Genomes/Genetics* **6**: 3647–3653.
- Love MI, Huber W, Anders S. 2014.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

- Lu H, Klocko AL, Brunner AM, Ma C, magnuson AC, Strauss SH. 2018a.** RNAi Suppression of Agamous-like Genes Causes Field Sterility in Populus.
- Lu H, Klocko AL, Brunner AM, Magnuson AC, Ma C, Strauss SH. 2018b.** Cross-Suppression of AG and AG-like 11 Genes gives Sterility in Field Grown Poplar.
- Lu H, Klocko AL, Dow M, Ma C, Amarasinghe V, Strauss SH. 2016.** Low frequency of zinc-finger nuclease-induced mutagenesis in Populus. *Molecular Breeding* **36**: 121.
- Luís Â, Duarte A, Gominho J, Domingues F, Duarte AP. 2016.** Chemical composition, antioxidant, antibacterial and anti-quorum sensing activities of Eucalyptus globulus and Eucalyptus radiata essential oils. *Industrial Crops and Products* **79**: 274–282.
- Luo H, Kausch AP, Hu Q, Nelson K, Wipff JK, Fricker CCR, Owen TP, Moreno MA, Lee J-Y, Hodges TK. 2005.** Controlling Transgene Escape in GM Creeping Bentgrass. *Molecular Breeding* **16**: 185–188.
- Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, et al. 2015.** A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Molecular Plant* **8**: 1274–1284.
- Ma X, Zhu Q, Chen Y, Liu Y-G. 2016.** CRISPR/Cas9 Platforms for Genome Editing in Plants: Developments and Applications. *Molecular Plant* **9**: 961–974.
- Maciver SK, Hussey PJ. 2002.** The ADF/cofilin family: actin-remodeling proteins. *Genome Biology* **3**: reviews3007.
- Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. 2019.** The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **47**: W636–W641.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013.** RNA-Guided Human Genome Engineering via Cas9. *Science* **339**: 823–826.
- Mamun ANK. 2007.** Reversible male sterility in transgenic tobacco carrying a dominant-negative mutated glutamine synthetase gene under the control of microspore-specific promoter. *IJEB Vol.45(12) [December 2007]*.
- Mandel MA, Yanofsky MF. 1995.** The Arabidopsis AGL8 MADS box gene is expressed in inflorescence meristems and is negatively regulated by APETALA1. *The Plant Cell* **7**: 1763–1771.
- Manghwar H, Lindsey K, Zhang X, Jin S. 2019.** CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing. *Trends in Plant Science* **24**: 1102–1125.

- Mariani C, Beuckeleer MD, Truettner J, Leemans J, Goldberg RB. 1990.** Induction of male sterility in plants by a chimaeric ribonuclease gene. *Nature* **347**: 737–741.
- Marocco A, Wissenbach M, Becker D, Paz-Ares J, Saedler H, Salamini F, Rohde W. 1989.** Multiple genes are transcribed in *Hordeum vulgare* and *Zea mays* that carry the DNA binding domain of the myb oncoproteins. *Molecular and General Genetics MGG* **216**: 183–187.
- Martin C, Paz-Ares J. 1997.** MYB transcription factors in plants. *Trends in Genetics* **13**: 67–73.
- Martinez-Garcia M, Schubert V, Osman K, Darbyshire A, Sanchez-Moran E, Franklin FCH. 2018.** TOP2 and chromosome movement help remove interlocks between entangled chromosomes during meiosis. *The Journal of Cell Biology* **217**: 4070–4079.
- McCabe DE, Swain WF, Martinell BJ, Christou P. 1988.** Stable Transformation of Soybean (*Glycine Max*) by Particle Acceleration. *Bio/Technology* **6**: 923–926.
- McCarthy EW, Mohamed A, Litt A. 2015.** Functional Divergence of APETALA1 and FRUITFULL is due to Changes in both Regulation and Coding Sequence. *Frontiers in Plant Science* **6**.
- McNevin JP, Woodward W, Hannoufa A, Feldmann KA, Lemieux B. 1993.** Isolation and characterization of eceriferum (cer) mutants induced by T-DNA insertions in *Arabidopsis thaliana*. *Genome* **36**: 610–618.
- Mellerowicz EJ, Horgan K, Walden A, Coker A, Walter C. 1998.** PRFLL– a *Pinus radiata* homologue of FLORICAULA and LEAFY is expressed in buds containing vegetative shoot and undifferentiated male cone primordia. *Planta* **206**: 619–629.
- Mendoza L, Thieffry D, Alvarez-Buylla ER. 1999.** Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics* **15**: 593–606.
- Miller CO, Skoog F, Okumura FS, Von Saltza MH, Strong F. 1955a.** Structure and synthesis of kinetin. *Journal of the American Chemical Society* **77**: 2662–2663.
- Miller CO, Skoog F, Von Saltza MH, Strong F. 1955b.** Kinetin, a cell division factor from deoxyribonucleic acid. *Journal of the American Chemical Society* **77**: 1392–1392.
- Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M. 2007.** NAC Transcription Factors, NST1 and NST3, Are Key Regulators of the Formation of Secondary Walls in Woody Tissues of *Arabidopsis*. *The Plant Cell* **19**: 270–280.

- Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. 2005.** The NAC Transcription Factors NST1 and NST2 of Arabidopsis Regulate Secondary Wall Thickenings and Are Required for Anther Dehiscence. *The Plant Cell* **17**: 2993–3006.
- Mitsunobu H, Teramoto J, Nishida K, Kondo A. 2017.** Beyond Native Cas9: Manipulating Genomic Information and Function. *Trends in Biotechnology* **35**: 983–996.
- Mizukami Y, Huang H, Tudor M, Hu Y, Ma H. 1996.** Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. *The Plant Cell* **8**: 831–845.
- Mizzotti C, Ezquer I, Paolo D, Rueda-Romero P, Guerra RF, Battaglia R, Rogachev I, Aharoni A, Kater MM, Caporali E, et al. 2014.** SEEDSTICK is a Master Regulator of Development and Metabolism in the Arabidopsis Seed Coat. *PLoS Genetics* **10**.
- Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L. 2012.** The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. *The Plant Journal* **70**: 409–420.
- Molinero-Rosales N, Jamilena M, Zurita S, Gómez P, Capel J, Lozano R. 1999.** FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and floral meristem identity. *The Plant Journal: For Cell and Molecular Biology* **20**: 685–693.
- Molliard M. 1921.** Sur le développement des plantules fragmentées. *CR Soc Biol (Paris)* **84**: 770–772.
- Monceau H-LDD. 1758.** *La Physique des arbres, où il est traité de l'anatomie des plantes et de l'économie végétale... par M. Duhamel du Monceau,...* chez H.-L. Guérin.
- Moncur MW, Boland DJ. 1989.** Floral Morphology of *Eucalyptus melliodora* A. Cunn. ex Schau. And Comparisons With Other Eucalypt Species. *Australian Journal of Botany* **37**: 125–135.
- Montenegro M. 2016.** CRISPR is coming to agriculture — with big implications for food, farmers, consumers and nature. *Ensis*.
- Montillet J-L, Leonhardt N, Mondy S, Tranchimand S, Rumeau D, Boudsocq M, Garcia AV, Douki T, Bigeard J, Laurière C, et al. 2013.** An Abscisic Acid-Independent Oxylipin Pathway Controls Stomatal Closure and Immune Defense in Arabidopsis. *PLoS Biology* **11**.
- MOREL GM. 1960.** Producing virus-free Cymbidium. *American Orchid Society Bulletin*.

- Mou Z, Wang X, Fu Z, Dai Y, Han C, Ouyang J, Bao F, Hu Y, Li J. 2002.** Silencing of Phosphoethanolamine N-Methyltransferase Results in Temperature-Sensitive Male Sterility and Salt Hypersensitivity in Arabidopsis. *The Plant Cell* **14**: 2031–2043.
- Mouradov A, Glassick T, Hamdorf B, Murphy L, Fowler B, Marla S, Teasdale RD. 1998.** NEEDLY, a *Pinus radiata* ortholog of FLORICAULA/LEAFY genes, expressed in both reproductive and vegetative meristems. *Proceedings of the National Academy of Sciences* **95**: 6537–6542.
- Moyroud E, Kusters E, Monniaux M, Koes R, Parcy F. 2010.** LEAFY blossoms. *Trends in Plant Science* **15**: 346–352.
- Moyroud E, Tichtinsky G, Parcy F. 2009.** The LEAFY Floral Regulators in Angiosperms: Conserved Proteins with Diverse Roles. *Journal of Plant Biology* **52**: 177–185.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986.** Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51 Pt 1**: 263–273.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, et al. 2014.** The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.
- Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C. 2004.** The Arabidopsis thaliana REDUCED EPIDERMAL FLUORESCENCE1 Gene Encodes an Aldehyde Dehydrogenase Involved in Ferulic Acid and Sinapic Acid Biosynthesis. *The Plant Cell* **16**: 544–554.
- Nakhooda M, Jain SM. 2016.** A review of eucalyptus propagation and conservation. *Propagation of ornamental plants : an international journal*. **16**: 101–119.
- National Academies of Sciences E, Studies D on E and L, Resources B on A and N, Prospects C on GECPE and F. 2016.** *Regulation of Current and Future Genetically Engineered Crops*. National Academies Press (US).
- Navarro-De la Sancha E, Coello-Coutiño MP, Valencia-Turcotte LG, Hernández-Domínguez EE, Trejo-Yepes G, Rodríguez-Sotres R. 2007.** Characterization of two soluble inorganic pyrophosphatases from *Arabidopsis thaliana*. *Plant Science* **172**: 796–807.
- Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S. 2013.** Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nature Biotechnology* **31**: 691–693.

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research* **17**: 2503–2516.

Nicolia A, Manzo A, Veronesi F, Rosellini D. 2014. An overview of the last 10 years of genetically engineered crop safety research. *Critical Reviews in Biotechnology* **34**: 77–88.

Niiler E. 1999. Terminator technology temporarily terminated. *Nature Biotechnology* **17**: 1054–1054.

Nishimura MT, Stein M, Hou B-H, Vogel JP, Edwards H, Somerville SC. 2003. Loss of a callose synthase results in salicylic acid-dependent disease resistance. *Science (New York, N.Y.)* **301**: 969–972.

Nobécourt P. 1939. Sur la pérennité et l'augmentation de volume des cultures de tissus végétaux. *CR Seances Soc. Biol. Ses Fil* **130**: 1270–1271.

Odipto J, Alicai T, Ingelbrecht I, Nusinow DA, Bart R, Taylor NJ. 2017. Efficient CRISPR/Cas9 Genome Editing of Phytoene desaturase in Cassava. *Frontiers in Plant Science* **8**.

Okada A, Arndell T, Borisjuk N, Sharma N, Watson-Haigh NS, Tucker EJ, Baumann U, Langridge P, Whitford R. 2019. CRISPR/Cas9-mediated knockout of Ms1 enables the rapid generation of male-sterile hexaploid wheat lines for use in hybrid seed production. *Plant Biotechnology Journal* **17**: 1905–1913.

Oliver MJ, Hake K. 2012. Seed-Based Gene Containment Strategies. In: *Plant Gene Containment*. John Wiley & Sons, Ltd, 113–124.

Oliver MJ, Quisenberry JE, Trolinder NLG, Keim DL. 1998. Control of plant gene expression.

Oliver MJ, Quisenberry JE, Trolinder NLG, Keim DL. 1999. Control of plant gene expression.

Ó'Maoiléidigh DS, Graciet E, Wellmer F. 2014. Gene networks controlling *Arabidopsis thaliana* flower development. *New Phytologist* **201**: 16–30.

Pajoro A, Biewers S, Dougali E, Leal Valentim F, Mendes MA, Porri A, Coupland G, Van de Peer Y, van Dijk ADJ, Colombo L, et al. 2014. The (r)evolution of gene regulatory networks controlling *Arabidopsis* plant reproduction: a two-decade history. *Journal of Experimental Botany* **65**: 4731–4745.

Parcy F. 2005. Flowering: a time for integration. *The International Journal of Developmental Biology* **49**: 585–593.

- Parcy F, Bomblies K, Weigel D. 2002.** Interaction of LEAFY, AGAMOUS and TERMINAL FLOWER1 in maintaining floral meristem identity in Arabidopsis. *Development* **129**: 2519–2527.
- Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, Bevan MW, Mansfield S, Whetten RW, Sederoff RR, et al. 2003.** Characterisation of a pine MYB that regulates lignification. *The Plant Journal* **36**: 743–754.
- Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000.** B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* **405**: 200–203.
- Peng A, Chen S, Lei T, Xu L, He Y, Wu L, Yao L, Zou X. 2017.** Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnology Journal* **15**: 1509–1519.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016.** Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**: 1650–1667.
- Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015.** StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**: 290–295.
- Peterson BA, Haak DC, Nishimura MT, Teixeira PJPL, James SR, Dangel JL, Nimchuk ZL. 2016.** Genome-Wide Assessment of Efficiency and Specificity in CRISPR/Cas9 Mediated Multiple Site Targeting in Arabidopsis. *PLOS ONE* **11**: e0162169.
- Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014.** The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**: 1246752.
- Pina C, Pinto F, Feijó JA, Becker JD. 2005.** Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiology* **138**: 744–756.
- Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF. 2003.** Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**: 85–88.
- Plackett AR, Conway SJ, Hewett Hazelton KD, Rabbinowitsch EH, Langdale JA, Di Stilio VS. 2018.** LEAFY maintains apical stem cell activity during shoot development in the fern *Ceratopteris richardii* (CS Hardtke and S McCormick, Eds.). *eLife* **7**: e39625.
- Porth I. 2014.** Current status of the development of genetically modified (GM) forest trees world-wide: a comparison with the development of other GM plants in agriculture.

CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources **9**.

Quétier F. 2016. The CRISPR-Cas9 technology: Closer to the ultimate toolkit for targeted genome editing. *Plant Science* **242**: 65–76.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W. 2003. Genome-Wide Characterization of the Lignification Toolbox in Arabidopsis. *Plant Physiology* **133**: 1051–1071.

Raghavan V. 1986. *Embryogenesis in Angiosperms: A Developmental and Experimental Study*. CUP Archive.

Ratcliffe OJ, Bradley DJ, Coen ES. 1999. Separation of shoot and floral identity in Arabidopsis. *Development* **126**: 1109–1120.

Regal PJ. 1982. Pollination by Wind and Animals: Ecology of Geographic Patterns. *Annual Review of Ecology and Systematics* **13**: 497–524.

Reinhardt D, Kuhlemeier C. 2002. Plant architecture. *EMBO Reports* **3**: 846.

Ren C, Liu X, Zhang Z, Wang Y, Duan W, Li S, Liang Z. 2016. CRISPR/Cas9-mediated efficient targeted mutagenesis in Chardonnay (*Vitis vinifera* L.). *Scientific Reports* **6**.

Ren Z, Zheng Z, Chinnusamy V, Zhu J, Cui X, Iida K, Zhu J-K. 2010. RAS1, a quantitative trait locus for salt tolerance and ABA sensitivity in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 5669–5674.

Richardson DM. 1998. Forestry Trees as Invasive Aliens. *Conservation Biology* **12**: 18–26.

Rijkema AS, Vandenbussche M, Koes R, Heijmans K, Gerats T. 2010. Variations on a theme: Changes in the floral ABCs in angiosperms. *Seminars in Cell & Developmental Biology* **21**: 100–107.

Robbins WJ. 1922. Cultivation of Excised Root Tips and Stem Tips Under Sterile Conditions. *Botanical Gazette* **73**: 376–390.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* **42**: W320–W324.

Robinson WD, Park J, Tran HT, Del Vecchio HA, Ying S, Zins JL, Patel K, McKnight TD, Plaxton WC. 2012. The secreted purple acid phosphatase isozymes

AtPAP12 and AtPAP26 play a pivotal role in extracellular phosphate-scavenging by *Arabidopsis thaliana*. *Journal of Experimental Botany* **63**: 6531–6542.

Rottmann WH, Meilan R, Sheppard LA, Brunner AM, Skinner JS, Ma C, Cheng S, Jouanin L, Pilate G, Strauss SH. 2000. Diverse effects of overexpression of LEAFY and PTLF, a poplar (*Populus*) homolog of LEAFY/FLORICAULA, in transgenic poplar and *Arabidopsis*. *The Plant Journal: For Cell and Molecular Biology* **22**: 235–245.

RStudio Team. 2015. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc.

Rutley N, Twell D. 2015. A decade of pollen transcriptomics. *Plant Reproduction* **28**: 73–89.

Ruzicka DR, Kandasamy MK, McKinney EC, Burgos-Rivera B, Meagher RB. 2007. The ancient subclasses of *Arabidopsis* ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. *The Plant Journal* **52**: 460–472.

Saleme M de LS, Cesarino I, Vargas L, Kim H, Vanholme R, Goeminne G, Van Acker R, Fonseca FC de A, Pallidis A, Voorend W, et al. 2017. Silencing CAFFEOYL SHIKIMATE ESTERASE Affects Lignification and Improves Saccharification in Poplar. *Plant Physiology* **175**: 1040–1057.

Samanta MK, Dey A, Gayen S. 2016. CRISPR/Cas9: an advanced tool for editing plant genomes. *Transgenic Research* **25**: 561–573.

Sander JD, Maeder ML, Reyon D, Voytas DF, Joung JK, Dobbs D. 2010. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Research* **38**: W462–W468.

Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D. 2007. Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Research* **35**: W599–W605.

Sang Y, Millwood RJ, Neal Stewart C. 2013. Gene use restriction technologies for transgenic plant bioconfinement. *Plant Biotechnology Journal* **11**: 649–658.

Sauer NJ, Narváez-Vásquez J, Mozoruk J, Miller RB, Warburg ZJ, Woodward MJ, Mihiret YA, Lincoln TA, Segami RE, Sanders SL, et al. 2016. Oligonucleotide-Mediated Genome Editing Provides Precision and Function to Engineered Nucleases and Antibiotics in Plants. *Plant Physiology* **170**: 1917–1928.

Sayou C, Nanao MH, Jamin M, Posé D, Thévenon E, Grégoire L, Tichtinsky G, Denay G, Ott F, Peirats Llobet M, et al. 2016. A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nature Communications* **7**.

Schilperoort RA, Veldstra H, Warnaar SO, Mulder G, Cohen JA. 1967. Formation of complexes between DNA isolated from tobacco crown gall tumours and RNA complementary to *Agrobacterium tumefaciens* DNA. *Biochimica Et Biophysica Acta* **145**: 523–525.

Schiml S, Puchta H. 2016. Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas. *Plant Methods* **12**: 8.

Schindele P, Wolter F, Puchta H. 2019. Transforming plant biology and breeding with CRISPR/Cas9, Cas12 and Cas13. *FEBS Letters*: 1954–1967.

Schleiden MJ. 1838. *Beiträge zur phytogenesis*.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Schuetz M, Benske A, Smith RA, Watanabe Y, Tobimatsu Y, Ralph J, Demura T, Ellis B, Samuels AL. 2014. Laccases Direct Lignification in the Discrete Secondary Cell Wall Domains of Protoxylem1[W][OPEN]. *Plant Physiology* **166**: 798–807.

Schultz EA, Haughn GW. 1991. LEAFY, a Homeotic Gene That Regulates Inflorescence Development in Arabidopsis. *The Plant Cell Online* **3**: 771–781.

Schwann T. 1839. Mikroskopische Untersuchungen über die Übereinstimmung in der Struktur und dem Wachsthum der Thiere und Pflanzen.

Scortecci KC, Michaels SD, Amasino RM. 2001. Identification of a MADS-box gene, FLOWERING LOCUS M, that represses flowering. *The Plant Journal* **26**: 229–236.

Sedbrook JC, Carroll KL, Hung KF, Masson PH, Somerville CR. 2002. The Arabidopsis SKU5 gene encodes an extracellular glycosyl phosphatidylinositol-anchored glycoprotein involved in directional root growth. *The Plant Cell* **14**: 1635–1648.

Serrano-Mislata A, Fernández-Nohales P, Doménech MJ, Hanzawa Y, Bradley D, Madueño F. 2016. Separate elements of the TERMINAL FLOWER 1 cis-regulatory region integrate pathways to control flowering time and shoot meristem identity. *Development* **143**: 3315–3327.

Serrano-Mislata A, Goslin K, Zheng B, Rae L, Wellmer F, Graciet E, Madueño F. 2017. Regulatory interplay between LEAFY, APETALA1/CAULIFLOWER and TERMINAL FLOWER1: New insights into an old relationship. *Plant Signaling & Behavior* **12**.

Shan Q, Wang Y, Li J, Zhang Y, Chen K, Liang Z, Zhang K, Liu J, Xi JJ, Qiu J-L, et al. 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature Biotechnology* **31**: 686–688.

- Shane MW, Stigter K, Fedosejevs ET, Plaxton WC. 2014.** Senescence-inducible cell wall and intracellular purple acid phosphatases: implications for phosphorus remobilization in *Hakea prostrata* (Proteaceae) and *Arabidopsis thaliana* (Brassicaceae). *Journal of Experimental Botany* **65**: 6097–6106.
- Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DGJ, Wang H, Jackson L, Tang Y, Stewart CN, et al. 2012.** Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* **193**: 121–136.
- Shen L, Hua Y, Fu Y, Li J, Liu Q, Jiao X, Xin G, Wang J, Wang X, Yan C, et al. 2017.** Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. *Science China Life Sciences* **60**: 506–515.
- Shen W, Wei Y, Dauk M, Tan Y, Taylor DC, Selvaraj G, Zou J. 2006.** Involvement of a Glycerol-3-Phosphate Dehydrogenase in Modulating the NADH/NAD⁺ Ratio Provides Evidence of a Mitochondrial Glycerol-3-Phosphate Shuttle in *Arabidopsis*. *The Plant Cell* **18**: 422–441.
- Shim JS, Imaizumi T. 2015.** Circadian Clock and Photoperiodic Response in *Arabidopsis*: From Seasonal Flowering to Redox Homeostasis. *Biochemistry* **54**: 157–170.
- Shin B, Choi G, Yi H, Yang S, Cho I, Kim J, Lee S, Paek N-C, Kim J-H, Song P-S, et al. 2002.** AtMYB21, a gene encoding a flower-specific transcription factor, is regulated by COP1. *The Plant Journal* **30**: 23–32.
- Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, et al. 2017.** Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology* **15**: 169–182.
- Silva NF, Goring DR. 2002.** The proline-rich, extensin-like receptor kinase-1 (PERK1) gene is rapidly induced by wounding. *Plant Molecular Biology* **50**: 667–685.
- Silva CS, Puranik S, Round A, Brennich M, Jourdain A, Parcy F, Hugouvieux V, Zubieta C. 2016.** Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower. *Frontiers in Plant Science* **6**.
- Siriwardana NS, Lamb RS. 2012.** A conserved domain in the N-terminus is important for LEAFY dimerization and function in *Arabidopsis thaliana*. *The Plant Journal* **71**: 736–749.
- Skinner JS, Meilan R, Ma C, Strauss SH. 2003.** The *Populus* PTD promoter imparts floral-predominant expression and enables high levels of floral-organ ablation in *Populus*, *Nicotiana* and *Arabidopsis*. *Molecular Breeding* **12**: 119–132.

- Skoog F, Miller C. 1957.** Chemical regulation of growth and organ formation in plant tissues cultured. In: *Vitro Symp Soc Exp Biol*.
- Smaczniak C, Immink RGH, Angenent GC, Kaufmann K. 2012.** Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. *Development* **139**: 3081–3098.
- Smith EF. 1894.** Field Notes, 1892. *The Journal of Mycology* **7**: 373–377.
- Smith FD, Harpending PR, Sanford JC. 1992.** Biolistic transformation of prokaryotes: factors that affect biolistic transformation of very small cells. *Journal of General Microbiology* **138**: 239–248.
- Smith EF, Townsend CO. 1907.** A Plant-Tumor of Bacterial Origin. *Science* **25**: 671–673.
- Song G, Jia M, Chen K, Kong X, Khattak B, Xie C, Li A, Mao L. 2016.** CRISPR/Cas9: A powerful tool for crop genome editing. *The Crop Journal* **4**: 75–82.
- Sternberg SH, LaFrance B, Kaplan M, Doudna JA. 2015.** Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**: 110–113.
- Strauss SH, Jones KN, Lu H, Petit JD, Klocko AL, Betts MG, Brosi BJ, Fletcher RJ, Needham MD. 2017.** Reproductive modification in forest plantations: impacts on biodiversity and society. *New Phytologist* **213**: 1000–1021.
- Strong Support for Labeling Modified Foods - The New York Times.**
- Su L, Bassa C, Audran C, Mila I, Cheniclet C, Chevalier C, Bouzayen M, Roustan J-P, Chervin C. 2014.** The Auxin SI-IAA17 Transcriptional Repressor Controls Fruit Size Via the Regulation of Endoreduplication-Related Cell Expansion. *Plant and Cell Physiology* **55**: 1969–1976.
- Sun B, Xu Y, Ng K-H, Ito T. 2009.** A timing mechanism for stem cell maintenance and differentiation in the Arabidopsis floral meristem. *Genes & Development* **23**: 1791–1804.
- Swarbreck SM, Guerringue Y, Matthus E, Jamieson FJC, Davies JM. 2019.** Impairment in karrikin but not strigolactone sensing enhances root skewing in Arabidopsis thaliana. *The Plant Journal* **98**: 607–621.
- Swarts DC, van der Oost J, Jinek M. 2017.** Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular Cell* **66**: 221-233.e4.
- Tabashnik BE, Brévault T, Carrière Y. 2013.** Insect resistance to Bt crops: lessons from the first billion acres. *Nature Biotechnology* **31**: 510–521.

- Tabashnik BE, Carrière Y. 2017.** Surge in insect resistance to transgenic crops and prospects for sustainability. *Nature Biotechnology* **35**: 926–935.
- Takebe I, Labib G, Melchers G. 1971.** Regeneration of whole plants from isolated mesophyll protoplasts of tobacco | SpringerLink. *Naturwissenschaften* **58**: 318–320.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**: 2725–2729.
- Tanahashi T, Sumikawa N, Kato M, Hasebe M. 2005.** Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* **132**: 1727–1736.
- Taylor G, Allwright MR, Smith HK, Polle A, Wildhagen H, Hertzberg M, Bhalerao R, Keurentjes JJB, Scalabrin S, Scaglione D, et al. 2016.** Bioenergy Trees: Genetic and Genomic Strategies to Improve Yield. In: Barth S, Murphy-Bokern D, Kalinina O, Taylor G, Jones M, eds. *Perennial Biomass Crops for a Resource-Constrained World*. Cham: Springer International Publishing, 167–190.
- Thangavel G, Nayar S. 2018.** A Survey of MIKC Type MADS-Box Genes in Non-seed Plants: Algae, Bryophytes, Lycophytes and Ferns. *Frontiers in Plant Science* **9**.
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Münster T, Winter KU, Saedler H. 2000.** A short history of MADS-box genes in plants. *Plant Molecular Biology* **42**: 115–149.
- Theißen G, Melzer R, Rümpler F. 2016.** MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development* **143**: 3259–3271.
- Theißen G, Saedler H. 2001.** Floral quartets. *Nature* **409**: 469.
- Toffaletti DL, Rude TH, Johnston SA, Durack DT, Perfect JR. 1993.** Gene transfer in *Cryptococcus neoformans* by use of biolistic delivery of DNA. *Journal of Bacteriology* **175**: 1405–1411.
- Tran HT, Qian W, Hurley BA, She Y-M, Wang D, Plaxton WC. 2010.** Biochemical and molecular characterization of AtPAP12 and AtPAP26: the predominant purple acid phosphatase isozymes secreted by phosphate-starved *Arabidopsis thaliana*. *Plant, Cell & Environment* **33**: 1789–1803.
- Tukey H. 1934.** Growth of the embryo, seed, and pericarp of the sour cherry (*Prunus cerasus*) in relation to season of fruit ripening. In: Proc. Amer. Soc. Hort. Sci. 125–144.

- Turlapati PV, Kim K-W, Davin LB, Lewis NG. 2011.** The laccase multigene family in *Arabidopsis thaliana*: towards addressing the mystery of their gene function(s). *Planta* **233**: 439–470.
- Turnbull JW. 2000.** Economic and social importance of eucalypts. : 1–9.
- Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, et al. 2006.** The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.
- Ueta R, Abe C, Watanabe T, Sugano SS, Ishihara R, Ezura H, Osakabe Y, Osakabe K. 2017.** Rapid breeding of parthenocarpic tomato plants using CRISPR/Cas9. *Scientific Reports* **7**.
- Uhrig RG, Labandera A-M, Tang L-Y, Sieben NA, Goudreault M, Yeung E, Gingras A-C, Samuel MA, Moorhead GBG. 2017.** Activation of Mitochondrial Protein Phosphatase SLP2 by MIA40 Regulates Seed Germination1[OPEN]. *Plant Physiology* **173**: 956–969.
- UniProt: a worldwide hub of protein knowledge. 2019.** *Nucleic Acids Research* **47**: D506–D515.
- Upadhyay SK, Kumar J, Alok A, Tuli R. 2013.** RNA-Guided Genome Editing for Target Gene Mutations in Wheat. *G3: Genes/Genomes/Genetics* **3**: 2233–2238.
- Vaeck M, Reynaerts A, Höfte H, Jansens S, De Beuckeleer M, Dean C, Zabeau M, Montagu MV, Leemans J. 1987.** Transgenic plants protected from insect attack. *Nature* **328**: 33–37.
- Van Acker RC, Szumgalski AR, Friesen LF. 2007.** The potential benefits, risks and costs of genetic use restriction technologies. *Canadian Journal of Plant Science* **87**: 753–762.
- Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, et al. 2013.** Caffeoyle Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in *Arabidopsis*. *Science* **341**: 1103–1106.
- van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, Reece-Hoyes JS, Nye C, Gradia S, Vidal B, et al. 2016.** DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Molecular Cell* **63**: 633–646.
- Vázquez-Lobo A, Carlsbecker A, Vergara-Silva F, Alvarez-Buylla ER, Piñero D, Engström P. 2007.** Characterization of the expression patterns of LEAFY/FLORICAULA and NEEDLY orthologs in female and male cones of the conifer genera *Picea*, *Podocarpus*, and *Taxus*: implications for current evo-devo hypotheses for gymnosperms. *Evolution & Development* **9**: 446–459.

Veljanovski V, Vanderbeld B, Knowles VL, Snedden WA, Plaxton WC. 2006. Biochemical and molecular characterization of AtPAP26, a vacuolar purple acid phosphatase up-regulated in phosphate-deprived Arabidopsis suspension cells and seedlings. *Plant Physiology* **142**: 1282–1293.

Venables WN, Ripley BD. 2002. *Modern Applied Statistics with S*. New York: Springer.

Vining KJ, Contreras RN, Ranik M, Strauss SH. 2012. Genetic Methods for Mitigating Invasiveness of Woody Ornamental Plants: Research Needs and Opportunities. *HortScience* **47**: 1210–1216.

Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M, et al. 2015a. The floral transcriptome of *Eucalyptus grandis*. *New Phytologist* **206**: 1406–1422.

Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M, et al. 2015b. The floral transcriptome of *Eucalyptus grandis*. *New Phytologist* **206**: 1406–1422.

Voytas DF. 2013. Plant Genome Engineering with Sequence-Specific Nucleases. *Annual Review of Plant Biology* **64**: 327–350.

Vranová E, Coman D, Gruissem W. 2013. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology* **64**: 665–700.

Waltz E. 2016. Gene-edited CRISPR mushroom escapes US regulation. *Nature News* **532**: 293.

Waltz E. 2018. With a free pass, CRISPR-edited plants reach market in record time. *Nature Biotechnology* **36**: 6–7.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu J-L. 2014. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology* **32**: 947–951.

Wang L, Ran L, Hou Y, Tian Q, Li C, Liu R, Fan D, Luo K. 2017. The transcription factor MYB115 contributes to the regulation of proanthocyanidin biosynthesis and enhances fungal resistance in poplar. *New Phytologist* **215**: 351–367.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R. 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**: 910–918.

Wasternack C, Feussner I. 2018. The Oxylipin Pathways: Biochemistry and Function. *Annual Review of Plant Biology* **69**: 363–386.

- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM. 1992.** LEAFY controls floral meristem identity in Arabidopsis. *Cell* **69**: 843–859.
- Weigel D, Glazebrook J. 2006.** Transformation of agrobacterium using the freeze-thaw method. *CSH protocols* **2006**.
- Weinthal D, Tovkach A, Zeevi V, Tzfira T. 2010.** Genome editing in plant cells by zinc finger nucleases. *Trends in Plant Science* **15**: 308–321.
- Wellmer F, Riechmann JL. 2010.** Gene networks controlling the initiation of flower development. *Trends in Genetics* **26**: 519–527.
- Went F. 1926.** On growth-accelerating substances in the coleoptile of *Avena sativa*. In: Proc Kon Akad Wetensch Amsterdam. 10–19.
- White PR. 1934a.** *Multiplication of the viruses of tobacco and Aucuba mosaics in growing excised tomato root tips.*
- White PR. 1934b.** POTENTIALLY UNLIMITED GROWTH OF EXCISED TOMATO ROOT TIPS IN A LIQUID MEDIUM. *Plant Physiology* **9**: 585–600.
- White PR. 1937a.** VITAMIN B1 IN THE NUTRITION OF EXCISED TOMATO ROOTS. *Plant Physiology* **12**: 803–811.
- White PR. 1937b.** SEPARATION FROM YEAST OF MATERIALS ESSENTIAL FOR GROWTH OF EXCISED TOMATO ROOTS. *Plant Physiology* **12**: 777–791.
- White PR. 1939.** Potentially Unlimited Growth of Excised Plant Callus in an Artificial Nutrient. *American Journal of Botany* **26**: 59–64.
- Wickham H. 2009.** *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
- Willmitzer L, Beuckeleer MD, Lemmers M, Montagu MV, Schell J. 1980.** DNA from Ti plasmid present in nucleus and absent from plastids of crown gall plant cells. *Nature* **287**: 359–361.
- Wils CR, Kaufmann K. 2017.** Gene-regulatory networks controlling inflorescence and flower development in *Arabidopsis thaliana*. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**: 95–105.
- Wilson PG. 2011.** Myrtaceae. In: Kubitzki K, ed. *The Families and Genera of Vascular Plants. Flowering Plants. Eudicots: Sapindales, Cucurbitales, Myrtaceae*. Berlin, Heidelberg: Springer, 212–271.

- Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM. 2009.** Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution* **24**: 136–144.
- Wilson AK, Pickett FB, Turner JC, Estelle M. 1990.** A dominant mutation in *Arabidopsis* confers resistance to auxin, ethylene and abscisic acid. *Molecular and General Genetics MGG* **222**: 377–383.
- Withers JC, Shipp MJ, Rupasinghe SG, Sukumar P, Schuler MA, Muday GK, Wyatt SE. 2013.** GRAVITY PERSISTENT SIGNAL 1 (GPS1) Reveals Novel Cytochrome P450s Involved in Gravitropism. *American Journal of Botany* **100**: 183–193.
- Witten DM. 2011.** Classification and clustering of sequencing data using a Poisson model. *The Annals of Applied Statistics* **5**: 2493–2518.
- Wolt JD, Wang K, Sashital D, Lawrence-Dill CJ. 2016.** Achieving Plant CRISPR Targeting that Limits Off-Target Effects. *The Plant Genome* **9**: 1–8.
- Xiao Y, Wu K. 2019.** Recent progress on the interaction between insects and *Bacillus thuringiensis* crops. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**: 20180316.
- Xie H, Wang D, Qin Y, Ma A, Fu J, Qin Y, Hu G, Zhao J. 2019.** Genome-wide identification and expression analysis of SWEET gene family in *Litchi chinensis* reveal the involvement of LcSWEET2a/3b in early seed development. *BMC Plant Biology* **19**: 499.
- Xie K, Yang Y. 2013.** RNA-Guided Genome Editing in Plants Using a CRISPR–Cas System. *Molecular Plant* **6**: 1975–1983.
- Xu J, Hua K, Lang Z. 2019.** Genome editing for horticultural crop improvement. *Horticulture Research* **6**.
- Xu R-F, Li H, Qin R-Y, Li J, Qiu C-H, Yang Y-C, Ma H, Li L, Wei P-C, Yang J-B. 2015.** Generation of inheritable and “transgene clean” targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Scientific Reports* **5**: 11491.
- Xue L-J, Alabady MS, Mohebbi M, Tsai C-J. 2015.** Exploiting genome variation to improve next-generation sequencing data analysis and genome editing efficiency in *Populus tremula* × *alba* 717-1B4. *Tree Genetics & Genomes* **11**: 82.
- Yamagishi N, Sasaki S, Yamagata K, Komori S, Nagase M, Wada M, Yamamoto T, Yoshikawa N. 2011.** Promotion of flowering and reduction of a generation time in apple seedlings by ectopical expression of the *Arabidopsis thaliana* FT gene using the Apple latent spherical virus vector. *Plant Molecular Biology* **75**: 193–204.

Yamaguchi N, Winter CM, Wu M-F, Kanno Y, Yamaguchi A, Seo M, Wagner D. 2014. Gibberellin Acts Positively Then Negatively to Control Onset of Flower Formation in Arabidopsis. *Science* **344**: 638–641.

Yamaguchi A, Wu M-F, Yang L, Wu G, Poethig RS, Wagner D. 2009. The microRNA regulated SBP-box transcription factor SPL3 is a direct upstream activator of LEAFY, FRUITFULL, and APETALA1. *Developmental cell* **17**: 268–278.

Yang T, Du M, Guo Y, Liu X. 2017. Two LEAFY homologs ILFY1 and ILFY2 control reproductive and vegetative developments in Isoetes L. *Scientific Reports* **7**.

Yang L, Güell M, Niu D, George H, Lesha E, Grishin D, Aach J, Shrock E, Xu W, Poci J, et al. 2015. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science* **350**: 1101–1104.

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. 1990. The protein encoded by the Arabidopsis homeotic gene *agamous* resembles transcription factors. *Nature* **346**: 35–39.

Yao W, Zhang W. 2011. Research on Manufacturing Technology and Application of Natural Bamboo Fibre. In: 2011 Fourth International Conference on Intelligent Computation Technology and Automation. 143–148.

Yaseen M, Ahmad T, Sablok G, Standardi A, Hafiz IA. 2013. Review: role of carbon sources for in vitro plant growth and development. *Molecular Biology Reports; Dordrecht* **40**: 2837–49.

Ye X, Al-Babili S, Klöti A, Zhang J, Lucca P, Beyer P, Potrykus I. 2000. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science (New York, N.Y.)* **287**: 303–305.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**.

Yi Chou E, Schuetz M, Hoffmann N, Watanabe Y, Sibout R, Samuels AL. 2018. Distribution, mobility, and anchoring of lignin-related oxidative enzymes in Arabidopsis secondary cell walls. *Journal of Experimental Botany* **69**: 1849–1859.

Yui R, Iketani S, Mikami T, Kubo T. 2003. Antisense inhibition of mitochondrial pyruvate dehydrogenase E1 α subunit in anther tapetum causes male sterility. *The Plant Journal* **34**: 57–66.

Zaenen I, van Larebeke N, Teuchy H, van Montagu M, Schell J. 1974. Supercoiled circular DNA in crown-gall inducing Agrobacterium strains. *Journal of Molecular Biology* **86**: 109–127.

- Zambryski P, Depicker A, Kruger K, Goodman HM. 1982.** Tumor induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. *Journal of Molecular and Applied Genetics* **1**: 361–370.
- Zambryski P, Holsters M, Kruger K, Depicker A, Schell J, Van Montagu M, Goodman HM. 1980.** Tumor DNA Structure in Plant Cells Transformed by *A. tumefaciens*. *Science* **209**: 1385–1391.
- Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. 2015.** Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**: 759–771.
- Zhang C, Liu C, Weng J, Cheng B, Liu F, Li X, Xie C. 2017.** Creation of targeted inversion mutations in plants using an RNA-guided endonuclease. *The Crop Journal* **5**: 83–88.
- Zhang F, Liu X, Zhang A, Jiang Z, Chen L, Zhang X. 2019.** Genome-wide dynamic network analysis reveals a critical transition state of flower development in *Arabidopsis*. *BMC Plant Biology* **19**: 11.
- Zhang Y, Shewry PR, Jones H, Barcelo P, Lazzeri PA, Halford NG. 2001.** Expression of antisense SnRK1 protein kinase sequence causes abnormal pollen development and male sterility in transgenic barley. *The Plant Journal* **28**: 431–441.
- Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, et al. 2014.** The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnology Journal* **12**: 797–807.
- Zheng Y, Xie Y, Jiang Y, Qu X, Huang S. 2013.** *Arabidopsis* ACTIN-DEPOLYMERIZING FACTOR7 Severs Actin Filaments and Regulates Actin Cable Turnover to Promote Normal Pollen Tube Growth. *The Plant Cell* **25**: 3405–3423.
- Zhong R, Demura T, Ye Z-H. 2006.** SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of *Arabidopsis*. *The Plant Cell* **18**: 3158–3170.
- Zhong R, Richardson EA, Ye Z-H. 2007.** Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of *Arabidopsis*. *Planta* **225**: 1603–1611.
- Zhong R, Ye Z-H. 2015.** The *Arabidopsis* NAC transcription factor NST2 functions together with SND1 and NST1 to regulate secondary wall biosynthesis in fibers of inflorescence stems. *Plant Signaling & Behavior* **10**: e989746.
- Zhou X, Jacobs TB, Xue L-J, Harding SA, Tsai C-J. 2015.** Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate:CoA ligase specificity and redundancy. *New Phytologist* **208**: 298–301.

- Zhou H, Liu B, Weeks DP, Spalding MH, Yang B. 2014a.** Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research* **42**: 10903–10914.
- Zhou H, Liu B, Weeks DP, Spalding MH, Yang B. 2014b.** Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Research* **42**: 10903–10914.
- Zhou S, Runge T, Karlen SD, Ralph J, Gonzales-Vigil E, Mansfield SD. 2017.** Chemical Pulping Advantages of Zip-lignin Hybrid Poplar. *ChemSusChem* **10**: 3565–3573.
- Zhu A, Ibrahim JG, Love MI. 2019.** Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**: 2084–2092.
- Zhu H, Qian W, Lu X, Li D, Liu X, Liu K, Wang D. 2005.** Expression patterns of purple acid phosphatase genes in Arabidopsis organs and functional analysis of AtPAP23 predominantly transcribed in flower. *Plant Molecular Biology* **59**: 581–594.
- Zhu J, Song N, Sun S, Yang W, Zhao H, Song W, Lai J. 2016.** Efficiency and Inheritance of Targeted Mutagenesis in Maize Using CRISPR-Cas9. *Journal of Genetics and Genomics = Yi Chuan Xue Bao* **43**: 25–36.
- Zianis D, Muukkonen P, Mäkipää R, Mencuccini M. 2005.** Biomass and stem volume equations for tree species in Europe. *Silva Fennica Monographs* **4**.
- Zou T, He Z, Qu L, Liu M, Zeng J, Liang Y, Wang T, Chen D, Xiao Q, Zhu J, et al. 2017.** Knockout of OsACOS12 caused male sterility in rice. *Molecular Breeding* **37**: 126.

APPENDICES

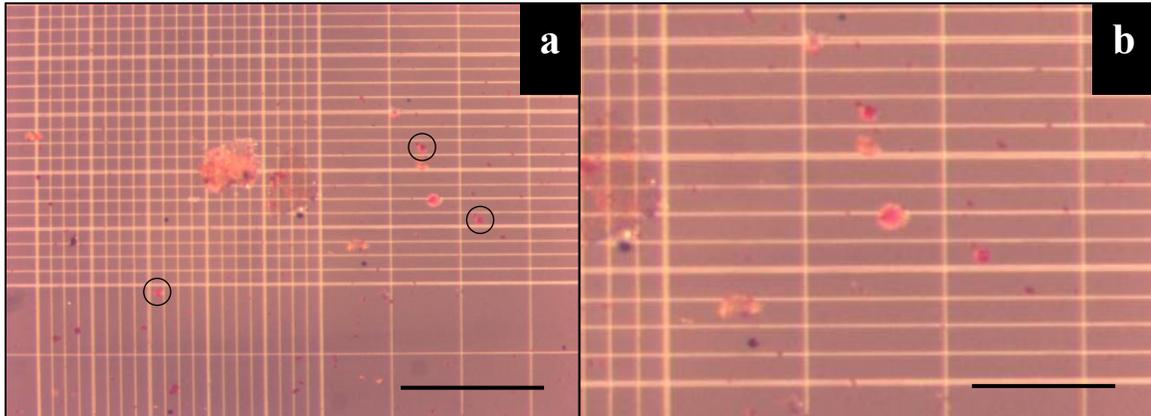
Appendix A Supplementary material for Chapter 2

Fig. S2.1 Putative pollen observed in 2007. **a** Although pollen production was extremely low, some of the debris associated with transgenic pollen from transgenic event 12 appeared to have normal shape and size. Two examples of normal size pollen are circled. The bar **a** corresponds to 1 mm and that in **b** (an enlargement of A) corresponds to 0.5 mm.

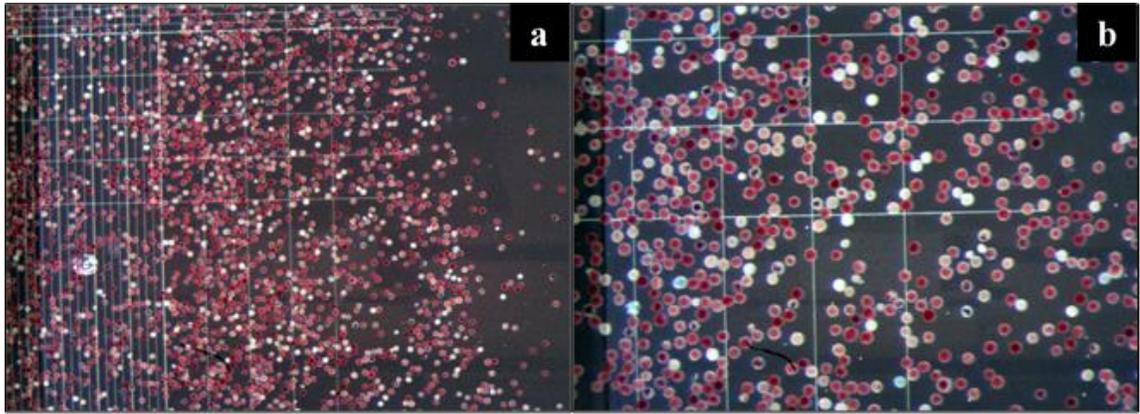


Fig. S2.2 Abundant viable pollen from the non-transgenic control in 2009. TTC-stained pollen from the control is shown (see methods). The bar in **a** corresponds to 1 mm and the bar in **b**, an enlargement of **a**, corresponds to 0.5 mm.

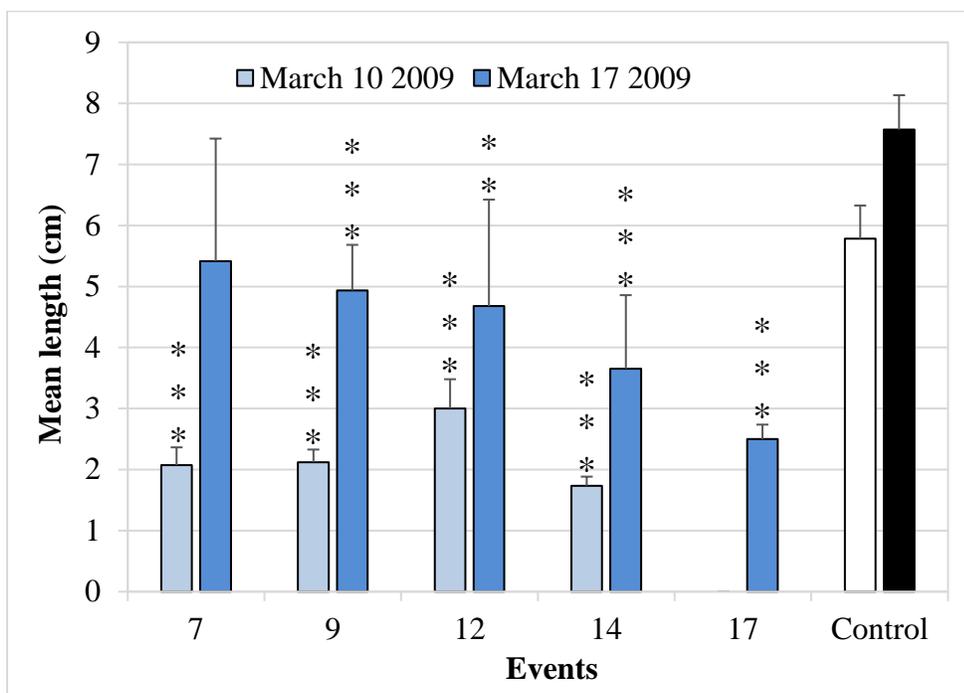


Fig. S2.3 Non-transgenic control trees had longer catkins than transgenic trees in 2009. The brackets represent 95 % confidence intervals. The asterisks indicate whether the event was significantly different than the control for the specific collection date based on a Dunnett's test (three asterisks: $P < 0.001$, two asterisks: $P < 0.01$, and one asterisk: $P < 0.05$; all rounded up).

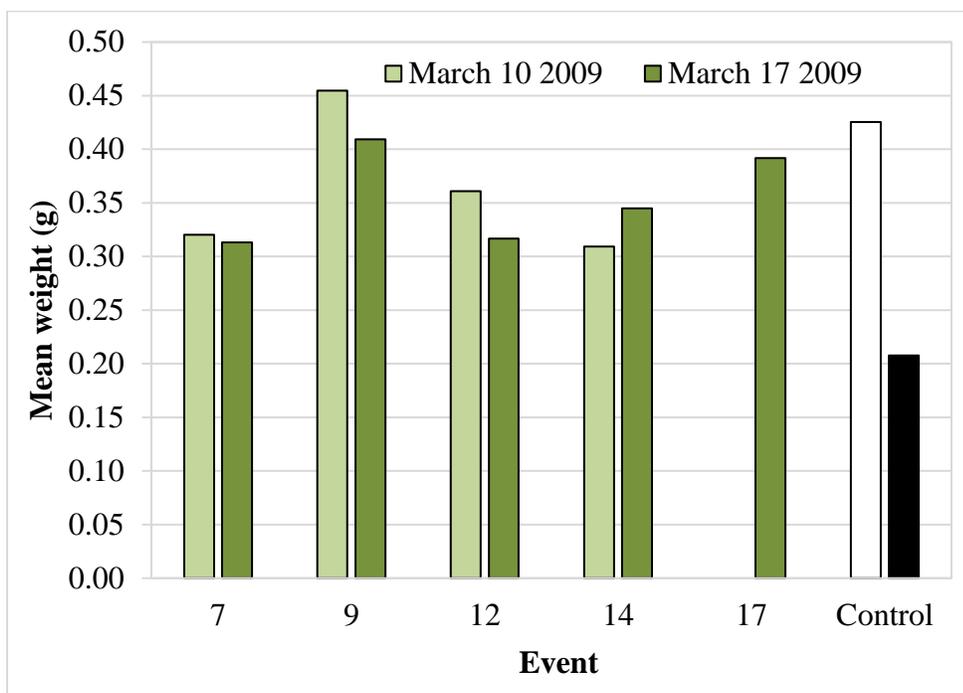


Fig. S2.4 Transgenic catkins were heavier than control catkins. A single sample t-test over all transgenic lines compared to the control value showed that catkin weight of the transgenics was significantly higher than that of the control ($P = 0.009$) for the May 17 collection date. The difference was not statistically significant for the March 10 collection date.



Fig. S2.5 Transgenic catkins were curved and dark in tone (less red). Shown are single representative catkins collected March 17th 2009. Event number is shown below each transgenic catkin.

Table S2.1 Genetic constructs used in this study.

Name	Other name	Binary vector backbone	Reference
TA29::BARNASE::NOS	-	pTTM8	Li et al. (2007)
PTD::GUS::NOS	3PG	modified pBI101	Sheppard et al. (2000)
En35S::GUS::E9	3SG	pMON10547	Perlak et al. (1993)
EnACT11::GUS::E9	3A11G	pMON10547	Perlak et al. (1993)
EnACT2::GUS::E9	3A2G	pMON10547	Perlak et al. (1993)

Table S2.2 Primers used in Polymerase Chain Reaction (PCR).

Name	Sequence (5' → 3')	Size (bp)	Comment	Reference
ACT2.001	atggtacctaggcaactatTTTTatgtatg	1228	Construction of	This study
ACT2.002	aggatccagctgcaaacacacaaaaagag		ACT2 promoter	
ACT11.001	taggtaccgctagcaaatgtcaaatggaatgcatc	1405	Construction of	This study
ACT11.002	gagaattctgtacatcctgtcaaaattgatataaa		ACT11 promoter	
NPTII-F	cttcttgacgagttcttc	340	Detection of	This study
NPTII-R	cgctgcctcgtcctg		Kan ^R gene	
TA29-pro-F02	tctcacactaagtccatgtttgc	656	Detection of	This study
TA29-term-R03	ggaaagtgaattgaccgatcagag		BARNASE gene	
V35F5	aggactattctggcttctcttac	880	Detection of GUS	Skinner et al. (2003)
GUS003	ccagactgaatgccacaggcc		gene	

Table S2.3 ANOVA table (one-way ANOVA) for the non-transgenic control trees that were distributed between the four transgenic constructs in the reporter trial for 2001.

Name	Degrees of freedom	F-value	p-value
Construct	3	2.700	0.139
Error	6		
Total	9		

Table S2.4 ANOVA tables (one-way ANOVA) for the non-transgenic control trees that were distributed between the four transgenic constructs in the reporter trial for 2003.

Name	Degrees of freedom	F-value	p-value
Construct	3	0.526	0.683
Error	5		
Total	8		

Table S2.5 ANOVA table (2-way ANOVA) with “Event” and “Block” as main effects for the sterility trial model.

Name	Degrees of freedom	F-value	Pr(>F)
Event	18	52	<0.0001
Block	2	8822522	<0.0001
Error	46		
Total	66		

Table S2.6 Dunnett's test for the sterility trial.

Linear Hypotheses	Estimate	Standard error	z value	Pr(> z)	Significance
1 - control == 0	-38.214	2.226	-17.167	< 0.001	***
2 - control == 0	-36.771	2.310	-15.917	< 0.001	***
3 - control == 0	-35.636	2.347	-15.182	< 0.001	***
4 - control == 0	-29.152	6.340	-4.598	< 0.001	***
5 - control == 0	-25.580	8.004	-3.196	0.020	*
6 - control == 0	-24.984	6.734	-3.710	0.003	**
7 - control == 0	-24.857	5.633	-4.412	< 0.001	***
8 - control == 0	-22.705	3.337	-6.804	< 0.001	***
9 - control == 0	-21.313	2.949	-7.227	< 0.001	***
10 - control == 0	-20.932	6.532	-3.204	0.020	*
11 - control == 0	-19.960	4.761	-4.192	< 0.001	***
12 - control == 0	-19.490	4.933	-3.951	0.001	**
13 - control == 0	-18.859	2.722	-6.929	< 0.001	***
14 - control == 0	-17.734	4.780	-3.710	0.003	**
15 - control == 0	-17.513	12.739	-1.375	0.900	
16 - control == 0	-17.216	4.230	-4.070	< 0.001	***
17 - control == 0	-16.863	2.767	-6.094	< 0.001	***
18 - control == 0	-12.241	3.797	-3.224	0.018	*

***: $P < 0.001$, **: $P < 0.01$, and *: $P < 0.05$

Table S2.7 ANOVA table (one-way ANOVA) with “Construct” as main event for the reporter trial model for 2001.

Name	Degrees of freedom	F-value	p-value
Construct	4	0.212	0.930
Error	36		
Total	40		

Table S2.8 ANOVA table (one-way ANOVA) with “Construct” as main event for the reporter trial model for 2003.

Name	Degrees of freedom	F-value	p-value
Construct	4	0.263	0.900
Error	36		
Total	40		

Table S2.9 Dunnett's test table for the reporter trial data from 2001.

Linear Hypotheses	Estimate	Standard error	z value	Pr(> z)
3A11G - control == 0	-3.583	127.260	-0.028	1
3A2G - control == 0	-12.603	124.383	-0.101	0.999
3PG - control == 0	-41.193	127.035	-0.324	0.935
3SG - control == 0	0.556	124.768	0.004	1

Table S2.10 Dunnett's test table for the reporter trial data from 2003.

Linear Hypotheses	Estimate	Standard error	z value	Pr(> z)
3A11G - control == 0	-1586.1	8020.9	-0.198	0.993
3A2G - control == 0	-3041.2	8075.5	-0.377	0.937
3PG - control == 0	-4206.2	8054.7	-0.522	0.848
3SG - control == 0	-593.9	8391.2	-0.071	1

Table S2.11 ANOVA tables (one-way ANOVA) with “Event” as main effect for each construct in the reporter trial for 2001.

3SG construct

Name	Degrees of freedom	F-value	p-value
Event	9	4.146	0.023
Error	9		
Total	18		

3PG construct

Name	Degrees of freedom	F-value	p-value
Event	9	7.495	0.001
Error	11		
Total	20		

3A2G construct

Name	Degrees of freedom	F-value	p-value
Event	9	17.356	1e-04
Error	10		
Total	19		

3A11G construct

Name	Degrees of freedom	F-value	p-value
Event	9	4.026	0.020
Error	10		
Total	19		

Table S2.12 ANOVA tables (one-way ANOVA) with “Event” as main effect for each construct in the reporter trial for 2003.

3SG construct

Name	Degrees of freedom	F-value	p-value
Event	9	15.816	3e-04
Error	8		
Total	17		

3PG construct

Name	Degrees of freedom	F-value	p-value
Event	9	3.954	0.022
Error	10		
Total	19		

3A2G construct

Name	Degrees of freedom	F-value	p-value
Event	9	77.816	<0.0001
Error	9		
Total	18		

3A11G construct

Name	Degrees of freedom	F-value	p-value
Event	9	3.464	0.033
Error	10		
Total	19		

Table S2.13 Dunnett's test for the sterility trial data comparing catkin mean length of control to that of transgenic events for catkin collection from March 10, 2009.

Linear Hypotheses	Estimate	Standard error	z value	Pr(> z)
Event 7 - control == 0	-3.709	0.308	-12.059	< 2e-16
Event 9 - control == 0	-3.662	0.291	-12.573	< 2e-16
Event 12 - control == 0	-2.783	0.362	-7.682	1.58e-14
Event 14 - control == 0	-4.047	0.281	-14.387	< 2e-16

Table S2.14 Dunnett's test for the sterility trial data comparing catkin mean length of control to that of transgenic events for catkin collection from March 17, 2009.

Linear Hypotheses	Estimate	Standard error	z value	Pr(> z)
Event 7 - control == 0	-2.341	1.049	-2.231	0.114
Event 9 - control == 0	-2.822	0.459	-6.152	< 0.001
Event 12 - control == 0	-3.079	0.854	-3.605	0.002
Event 14 - control == 0	-4.103	0.618	-6.636	< 0.001
Event 17 - control == 0	-5.257	0.295	-17.805	< 0.001

Table S2.15 Catkin angle per event.

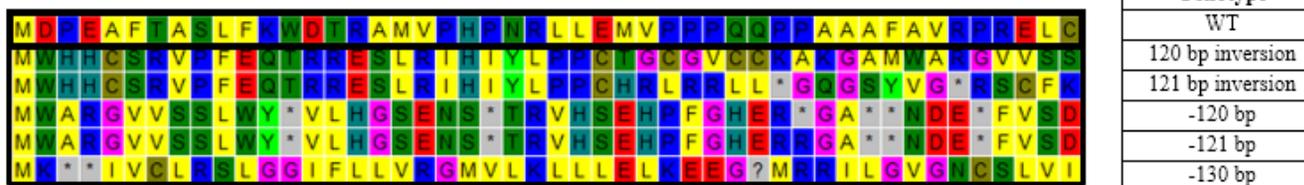
Event	Angle
7	72.83
9	90.89
12	77.83
14	130.26
17	83.28
Average	91.02
St. error	10.26

Appendix B Supplementary material for Chapter 3

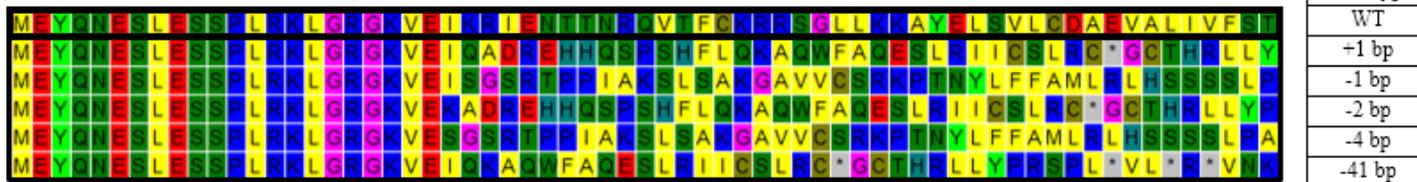
Amino acid modifications in *PLFY* for *LFY*-sg1



Amino acid modifications in *PLFY* for *LFY*-sg1sg2



Amino acid modifications in *PAG1* for *AG*-sg1sg2



Amino acid modifications in *PAG2* for *AG*-sg1sg2

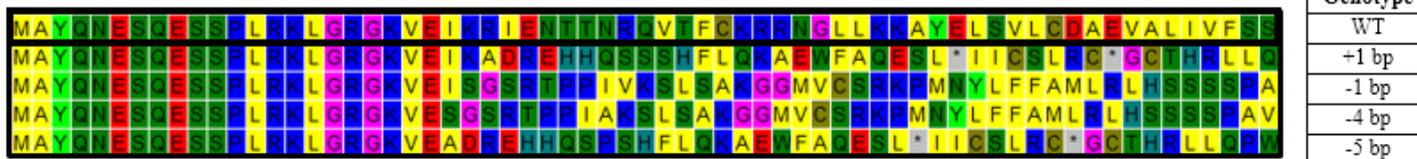


Fig. S3.1 Diversity in putative amino acid modifications to the WT peptide sequence in 717. These peptide alignments are the partial translation of the sequence alignment from the most common mutations seen in events with *LFY*-sg1, *LFY*-sg1sg2, and *AG*-sg1sg2. The first line in each alignment shows the WT sequence. Stop codons are shown with a *. The tables to the right identify the specific mutation that led to the peptide modification in each row.

Table S3.1 Partial genetic sequence of the target genes and the off-target sites. The 20bp protospacer sequences (i.e. the target sites), the off-target sites, the transcription-starting triplet (ATG), and the SNPs are underlined and in bold. SNPs are defined as: K, G or T; M, A or C; R, A or G; Y, C or T; W, A or T.

Gene-clone	Gene sequence
<i>Target Loci</i>	
<i>PLFY</i> in 717	CTGTCCAGTTC CGAAGAAACATCAAAAACCTTTAATTCTGTTAGCT TCC <u>Y</u> AATACATACAAAAAAGAAAAAAGACA <u>R</u> AAACTTGTCTCTG TTAAGGGCAGTTTTGGTATA <u>Y</u> AAATAAAACAAGAAGCTCACTTGT CTTTATATATCTACCAAATCCAAGACATGCACCAGTGAAAGATCA <u>CAGAGAGAGAGACA</u> AGGGGGCAGATAGATATGATCCGGAGGC TTTCACGGCGAGTTTGTCAAATGGGACACGAGAGCAATGGTGCC ACATCCTAACCGTCTGCTTGAAATGGT <u>GCCCCGCTCAGCAGC</u> <u>C</u> ACCGGCTGCGGCGTTTGCTGTAAGGCCAAGGGAGCTATGTGGC TAGAGGAGTTGTTTCAAGCTTATGGTATTAGGTACTACACGGCAGC GAAAATAGCTGAACTCGGGTTCACAGTGAACACCCTTTTGGACAT GAAAG <u>Y</u> GAGGAGCTTGATGAAATGATGAATAGTTTGTCTCAGAT CTTTAGGTGGGATCTTCTTGTTGGTGAGAGGTATGGTATTAAGCT GCTGTTAGAGCTGAAAGAAGAAGGCTTGATGAGGAGGATCCTAGG CGTAGGCAATTGCTCTCTGGTGATAATAATACAAATACTCTTGATG CTCTCTCCAAGAAGGTTTTGGTTAGCATTGATTCTACCTTTTAGTGT AATTAAGCTAAGCTCATACTATTACTAGCTATAGGAG <u>K</u> CCATGGC CGTTTT <u>R</u>
<i>PLFY</i> in 353	CTGTCCAGTTC CGAAGAAACATCAAAAACCTTTAATTCTGTTAGCT TCCAATACATACAAAAAAGAAAAAAGACA <u>R</u> AAACTTGTCTCTG TTAAGGGCAGTTTTGGTATA <u>Y</u> AAATAAAACAAGAAGCTCACTTGT CTTTATATATCTACCAAATCCAAGACATGCACCAGTGAAAGATCA <u>CAGAGAGAGAGACA</u> AGGGGGCAGATAGATATGATCCGGAGGC TTTCACGGCGAGTTTGTCAAATGGG <u>Y</u> ACGAGAGCAATGGTGCC ACATCCTAACCG <u>Y</u> CTGCTTGAAATGGT <u>GCCCCGCTCAGCAGC</u> <u>C</u> ACCGGCTGCGGCGTTTGCTGTAAGGCCAAGGGAGCTATGTGGC TAGAGGAGTTGTTTCAAGCTTATGGTATTAGGTACTACACGGCAGC GAAAATAGCTGAACTCGGGTTCACAGTGAACACCCTTTTGGACAT GAAAGATGAGGAGCTTGATGAAATGATGAATAGTTTGTCTCAGAT CTT <u>Y</u> AGGTGGGATCTTCTTGTTGGTGAGAGGTATGG
<i>PAG1</i> in 717	GCTAGACTGCAGCTATGGAATATCAAATGAATCCCTTGAGAGCT CCCCCTGAGGAAGCT <u>R</u> GGAA <u>GGGGAAAGGTGGAGATCAAG</u> CG GATCGAGAACACCACCAATC <u>GCCAAGTCACTTTCTGCAAA</u> AGGC GCAGTGGTTTTGCTCAAGAAAGCCTACG <u>R</u> TTATCTGTTCTTTGCGA TGCTGAGGTTGCACTCATCGTCTTCTCTACCCGCGGTCGCCTTTAT GAGTACTTAACGATAGGTAATAAATCTAATTTTAGATATATGCT TCTCTGGATCTTAAATTCTCCATGTTACAAGCCCTT

Table S3.1 Partial genetic sequence of the target genes and the off-target sites (continued).

<i>PAG1</i> in 353	GCAGCT ATG GAATATCAAATGAATCCCTTGAGAGCTCCCC M CTG AGGAAGCTGGGAAG GGGGAAAGGTGGAGATCAAG CGGATCGAGA ACACCACCAATC GCCAAGTCACTTTCTGCAAA AGGCGCAGTGGT TTGCTCAAGAAAGCCTACGAATTATCTGTTCTTTGCGATGCTGAGG TTGCACTCATCGTCTTCTACCCGCGGTGCGCTTATGAGTACTCT AACGATAGGTAATAAATCTAATTTTAGATATATGCTTCTCTGGAT CTTAAATTCTCC
<i>PAG2</i> in 717	GCTAGCAGCAGCT ATG GCATACCAAATGAATCCCAAGAGAGCTC TCCCCTGAGGAAGCTGGG RAGGGGAAAGGTGGAGATCAAG CGG ATCGAGAACACCAC MAATCGYCAAGTCACTTTCTGCAAA AGGCG GAATGGTTTGGCTCAAGAAAGCCTATGAATTATCTGTTCTTTGCGAT GCTGAGGTTGCACTCATCGTCTTCTCCAGCCGTGGACGCCTTTATG AGTACTTAACAATAGGTATATACTTAGTTCCTC W GCTCATGAATT CTCCATGTTGCAAR CCCTCTTCAAGTGCTCACAGTTGGTTTTTCTTG CTT Y CTCAT Y CAAAGGGATTGTTTTTT Y TTTT
<i>PAG2</i> in 353	ATG GCATACCAAATGAATCCCAAGAGAGCTCTCCCCTGAGGAAG CTGGGAG GGGGAAAGGTGGAGATCAAG CGGATCGAGAACACCA CMAATCGCCAAGTCACTTTCTGCAAA AGGCGGAATGGTTTGGTC AAGAAAGCCTATGAATTATCTGTTCTTTGCGATGCTGAGGTTGCAC TCATCGTCTTCTCCAGCCGTGGACGCCTTTATGAGTACTCTAACAA TAGGTATAT R CTTAGTTCCT Y G K CTCATGAATTCTCCATGTTGCAA GCCCTCTCAAGTGCTCACAGTTGGTTTTTCTTGCTTTCTCATCAA AGGGATTTG W TTTTTTCTTTTTGTTTATGCCAGGGTTAATTTTTATGG TTTTT
<i>Off-target loci</i>	
<i>UBC19</i> in 717 (Potri.001G254500)	AGAGACAATGGCAACTGTTAATGGGTATCAAGGGAATACTCCG G T GGCTGCTCCGGCGGGGACT TACCCCATCAAACAGACTGTCAC W GCGGCAAAGATTGTGACGATACGCAATCCGTGCTTAAACGGTAATTTT CTTTCTTTTTGCAT Y TGATCTGTTCTTTTCATTTGTCAAACCATGT AATATAT MWCRMGWKWTAWMWSKRKWTWTWTTWTRKGKKT TYTYTTTKYKYRWAWRTGWKWTTKTGWSWYTTTTTYTYKTK WTWWWAWM
<i>UBC20</i> in 717 (Potri.009G049600)	TATTAATGGGGTATACTCCG G T G GCTGCTCCGGCAGGGACTACC CCATCAAACAGACTGTCCCATCGGCAAAGACTGTTGATACACAA TCCGTGCTTAAACGGTATTWTTTTTTTT Y T
<i>USP36</i> in 717 (Potri.005G156900)	TCCGCTGGGCTGTTTTATAATTTAGGTGTAAATGGCGAGGTGCGGT T GCGAGAAAGGAGGAGATCAAG AGGCTGTTGGTTTTGGCAGCGG AGGAAGCCGCTAGGGCTGAGTTTGAGGCCGCGGCTTCATACGGCA CCGTTCCGGTGGTGACAAATAACTATCAATGTGCTGTTTGTTTTTG C YCGACAACGACACGGTGTGCCCGCTGTAAAGCTGTTAGATATTG GTATGT K AATTTCTGTGTTAATGCTTGATTTTTATCGGGTTTTAGTCC TTAATT G MGTTGAATTCAGGTTT G

Table S3.1 Partial genetic sequence of the target genes and the off-target sites (continued).

<p><i>STK.1</i> in 717 (Potri.013G104900)</p>	<p>ACTGTTGCATTGCCTAGCTATTCCATGCTTTTTTTGGTAGACATGA AATGTAAAAGTCAGATAAGCTAGCTATTAGGTCAAGAAAATTGCT TGATAAGAGCATATATAATATAGAAGCTTCTTTGGGTTGTGAAAG AATTGATCTTTTGTGTAGACATGGGAAGAGGAAAGATTGAGATC AAGAGGATCGAGAACACTACGAATCGTCAGGTTACTTTCTGCAAG AGAAGAAATGGGCTGTTGAAGAAAGCCTATGAATTATCTGTCCTT TGTGATGCTGAAGTCTCCCTCATCGTCTTCTCCAGCCGTGGCCGTC TCTATGAGTACGCCAACAACAAGTAACTTTACCTTCCCTAAAT</p>
<p><i>STK.2</i> in 717 (Potri.019G077200)</p>	<p>CTTTTTTTGCTCTTAATTTTGTCCATTTTCTA{T,TTAA}TTTAC TCTTTATAAAAAKATTTTTTTACCATTACTTCTCTACAGTCTTTCT CAAACTGTTGCRATTACCTTGCTATTCCATGCTTTTTTTGGTAGGCAT GACATGTAAAGGTCAGATCAGCTAGCTATTAGATTAAGAAAACTG GCGCATATATAAYACATAAGCTTGTATGGGTTATGAAAGAAACRA TCTTTTGTGTAGACATGGGAAGAGGAAAGATTGAGATCAAGAGG ATCGAGAACACCACCAATCGTCAGGTTACTTTCTGCAAGAGGAGA AATGGGCTCTTGAAGAAAGCTTATGAATTATCAGTTCTKTGTGATG CTGAAGTTGCTCTCATCGTCTTCTCTAGCCGTGGCCGTCTATGA GTACGCCAACAACAAGTAAATTTACCTTCTCCTTRTTGTCTTTTCTT TTGGATCTTGAWGGRAACCTCC{T,-}TTTCTTT</p>

Table S3.2 Table of primers, their sequence, and their specific use.

Primer name	Primer sequence (5' to 3')	Use(s)
AtU626_F1	CTTCAAAAGTCCCACATCGC	Verifying transgene sequence; event genotyping
sgRNA_R1	GCCGCCAGTGTGATGGATA	Verifying transgene sequence; event genotyping
Cas9_F1	CACGACGGAGACTACAAGGA	Verifying transgene sequence; event genotyping
Cas9_R1	TCCTTGTAGTCTCCGTCGTG	Verifying transgene sequence; event genotyping
Cas9_mid_F1	GTGGCCTATTCTGTGCTGGT	Verifying transgene sequence; event genotyping
Cas9_end_F2	CCTACAACAAGCACCGGGAT	Verifying transgene sequence; event genotyping
tnos_R2	AACGATCGGGGAAATTCGAG	Verifying transgene sequence; event genotyping
RB_F1	GAAGGCGGGAAACGACAATC	Verifying transgene sequence; event genotyping
RB_R1	CGGATAAACCTTTTCACGCC	Verifying transgene sequence; event genotyping
LFY_seq_F1	CCTGTTAAGGGCAGTTTTGG	Sequencing <i>PLFY</i> in both clones
LFY_seq_F7	TGCAGGGAACCAAATGTGTG	Sequencing <i>PLFY</i> in both clones
LFY_R2	AACCTTCTTGGGAGAGAGCA	Sequencing <i>PLFY</i> in both clones
AG_seq_F1	AGTTTGTGTTTTGGATCAGC	Sequencing <i>PAG1</i> in both clones
AG1_seq_F1	GTTGTCACTCAGTTTGTGTTTTGGA	Sequencing <i>PAG1</i> in both clones
AG1_seq_R4	GACAGCGACCACATGC	Sequencing <i>PAG1</i> in both clones
AG2_seq_F1	TGCTGTCTTCACCCAGTTTGT	Sequencing <i>PAG2</i> in both clones
AG2_seq_R5	AAAACCTTGACACCAGGCTCC	Sequencing <i>PAG2</i> in both clones
AG11_F2	TCACTCAGTTTGTGTTTTGGATCAG	Sequencing allele one of <i>PAG1</i> in 717

Table S3.2 Table of primers, their sequence, and their specific use (continued).

AG1II_F1	CACTCAGTTTGTGTTTTGGATCATC	Sequencing allele two of <i>PAG1</i> in 717
AG2I_R4	TTTGCAACATGGAGAATTCATGAGCT	Sequencing allele one of <i>PAG2</i> in 717
AG2II_R4	CTTGCAACATGGAGAATTCATGAGCA	Sequencing allele two of <i>PAG2</i> in 717
AG1I_353_F1	CCCTTGAGAGCTCCCCAC	Sequencing allele one of <i>PAG1</i> in 353
AG1II_353_F1	CCTTGAGAGCTCCCCCC	Sequencing allele two of <i>PAG1</i> in 353
AG2I_353_R2	TGCAACATGGAGAATTCATGAGC	Sequencing allele one of <i>PAG2</i> in 353
AG2II_353_R2	TGCAACATGGAGAATTCATGAGA	Sequencing allele two of <i>PAG2</i> in 353
P1G254500_F	TGTTGGTGCTTTCGATACCCT	Sequencing off-target Potri.001G254500
P1G254500_R	ACGGTTAGATAAAGAATCAGTCACA	Sequencing off-target Potri.001G254500
P9G049600_F	TGGGTTTTCTTTCTTTTGGATTCT	Sequencing off-target Potri.009G049600
P9G049600_R	AGATCACAAACCACATTCATAAACA	Sequencing off-target Potri.009G049600
OffAG_5F1	TAGGGTTTTCGAGCCTGGTG	Sequencing off-target Potri.005G156900
OffAG_5R1	TCTCCCCAGAACCAAACCTGA	Sequencing off-target Potri.005G156900
OffAG_13F1	TGGAAACAGCTTTGCACTTCC	Sequencing off-target Potri.013G104900
OffAG_13R1	ATGGTATGAAGATTTAGGGAAGGT	Sequencing off-target Potri.013G104900
OffAG_19F1	AGAAACAGATTTGCACACCCT	Sequencing off-target Potri.019G077200
OffAG_19R1	AGACCTAGTGATCTGTGAGAAAGA	Sequencing off-target Potri.019G077200

Table S3.3 Lack of mutations on target sites in empty vector controls. We sequenced the target sites corresponding to the four guide RNAs of events transformed with only the Cas9 sequence (i.e. no guide RNA) and found no mutations in both alleles of all target genes. N; number.

Transgene	Clone	Target gene	Events (N)	Gene amplicons (N)	Mutations (rate)
<i>Cas9</i> only	717	<i>PLFY</i>	32	64	0 (0%)
		<i>PAG1</i>		64	0 (0%)
		<i>PAG2</i>		64	0 (0%)
	353	<i>PLFY</i>	17	34	0 (0%)
		<i>PAG1</i>		34	0 (0%)
		<i>PAG2</i>		34	0 (0%)
Total			49	294	0 (0%)

Table S3.4 Mutation spectra of the different gene-sgRNA combinations with only one sgRNA. The most prevalent mutation type for each specific group is in bold. The “other” mutation type refers to nine, four, ten, and eight types corresponding to *LFY*-sg1, *LFY*-sg2, *AGI*-sg2, *AG2*-sg2 respectively with lower than 4.5% prevalence. “Other” is not bolded for *AGI*-sg2 because it is made up of more than one type of mutation. bp; base pairs.

	1 bp insertion	1 bp deletion	2 bp deletion	3 bp deletion	4 bp deletion	other	Total
<i>LFY</i> -sg1	58 (33.9%)	54 (31.6%)	23 (13.5%)	17 (9.9%)	4 (2.3%)	15 (8.8%)	171
<i>LFY</i> -sg2	18 (24.0%)	16 (21.3%)	33 (44.0%)	1 (1.3%)	3 (4.0%)	4 (5.3%)	75
<i>AGI</i> -sg2	18 (16.1%)	23 (20.5%)	14 (12.5%)	11 (9.8%)	21 (18.8%)	25 (22.3%)	112
<i>AG2</i> -sg2	21 (18.1%)	41 (35.3%)	8 (6.9%)	10 (8.6%)	17 (14.7%)	19 (16.4%)	116
Total	115 (24.3%)	134 (28.3%)	78 (16.5%)	39 (8.2%)	45 (9.5%)	63 (13.3%)	474

Table S3.5 Results table for the proportion comparison of all mutation spectra. Pearson's chi-squared test of independence was used to test if the mutation signature were different between gene-sgRNA combinations.

Mutation spectra comparison tested	X-squared	Degrees of freedom	P-value
All spectra	105.1	15	5.0e-04
<i>LFY</i>-sg1 vs. <i>LFY</i>-sg2	31.5	5	5.0e-04
<i>LFY</i>-sg1 vs. <i>AG1</i>-sg2	40.5	5	1.2e-07
<i>LFY</i>-sg1 vs. <i>AG2</i>-sg2	27.2	5	5.3e-05
<i>LFY</i>-sg2 vs. <i>AG1</i>-sg2	40.2	5	5.0e-04
<i>LFY</i>-sg2 vs. <i>AG2</i>-sg2	46.7	5	5.0e-04
<i>AG1</i>-sg2 vs. <i>AG2</i>-sg2	8.2	5	1.5e-01

Table S3.6 Mutation spectra generated by the same CRISPR Cas9 nuclease in the *PLFY* gene in two different hybrid poplar clones. The most prevalent mutation type for each specific group is in bold. The “other-large” and the “other-small” mutation types refer to 12 and 22 different mutations types respectively with lower than 4.5% prevalence. Chi-squared test of independence was used to determine if the mutation spectra are different.

	120 bp deletion	121 bp deletion	120 bp inversion	122 bp deletion	other-large	other-small	Total
<i>LFY</i>-sg1sg2 in 717	61 (40.1%)	8 (5.3%)	7 (4.6%)	7 (4.6%)	23 (15.1%)	46 (30.3%)	152
<i>LFY</i>-sg1sg2 in 353	32 (59.3%)	2 (3.7%)	1 (1.9%)	10 (18.5%)	6 (11.1%)	3 (5.6%)	54
Total	93 (45.1%)	10 (4.9%)	8 (3.9%)	17 (8.3%)	29 (14.1%)	49 (23.8%)	206

Table S3.7 Results table for the proportion comparison of the mutation spectra of *LFY*-sg1sg2 in two different poplar clones. Pearson's chi-squared test of independence was used to test if the mutation spectra were different between the different clones.

Mutation spectra comparison tested	X-squared	Degrees of freedom	P-value
<i>LFY</i> -sg1sg2 in 717 vs. <i>LFY</i> -sg1sg2 in 353	24.2	5	5.0e-04

Table S3.8 Mutation spectra generated by the same CRISPR Cas9 nuclease in the *PAG1* gene in two different hybrid poplar clones. The most prevalent mutation type for each specific group is in bold. The “other” mutation type refers to 20 different mutation types with lower than 4.5% prevalence. “Other” is not bolded for *AG1*-sg1sg2 in 717 because it is made up of more than one type of mutation.

	1 bp deletion	2 bp deletion	3 bp deletion	4 bp deletion	5 bp deletion	41 bp deletion	44 bp deletion	1 bp insertion	other	Total
<i>AG1</i>-sg1sg2 in 717	33 (19.3%)	31 (18.1%)	10 (5.8%)	18 (10.5%)	13 (7.6%)	15 (8.8%)	4 (2.3%)	13 (7.6%)	34 (19.9%)	171
<i>AG1</i>-sg1sg2 in 353	14 (26.4%)	7 (13.2%)	4 (7.5%)	5 (9.4%)	3 (5.7%)	8 (15.1%)	3 (5.7%)	3 (5.7%)	6 (11.3%)	53
Total	47	38	14	23	16	23	7	16	40	224

Table S3.9 Results table for the proportion comparison of the mutation spectra of *AGI-sg1sg2* in two different poplar clones. Pearson's chi-squared test of independence was used to test if the mutation spectra were different between the different clones.

Mutation spectra comparison tested	X-squared	Degrees of freedom	P-value
<i>AGI-sg1sg2</i> in 717 vs. <i>AGI-sg1sg2</i> in 353	6.9	8	0.6

Table S3.10 Mutation spectra generated by the same CRISPR Cas9 nuclease in the *PAG2* gene in two different hybrid poplar clones. The most prevalent mutation type for each specific group is in bold. The “other” mutation type refers to 11 different mutation types with lower than 4.5% prevalence.

	1 bp deletion	2 bp deletion	3 bp deletion	4 bp deletion	5 bp deletion	41 bp deletion	1 bp insertion	other	Total
AG2-sg1sg2 in 717	12 (34.3%)	2 (5.7%)	2 (5.7%)	5 (14.3%)	3 (8.6%)	2 (5.7%)	4 (11.4%)	5 (14.3%)	35
AG2-sg1sg2 in 353	23 (37.7%)	8 (13.1%)	3 (4.9%)	2 (3.3%)	3 (4.9%)	2 (3.3%)	8 (13.1%)	12 (19.7%)	61
Total	35	10	5	7	6	4	12	17	96

Table S3.11 Results table for the proportion comparison of the mutation spectra of *AG2-sg1sg2* in two different poplar clones. Pearson's chi-squared test of independence was used to test if the mutation signatures were different between the different clones.

Mutation spectra comparison tested	X-squared	Degrees of freedom	P-value
<i>AG2-sg1sg2</i> in 717 vs. <i>AG2-sg1sg2</i> in 353	6.2	7	0.5

Table S3.12 Off-target sites studied for rate of mutagenesis. The selected targets differed by three or four bases (the bases that did not match the target are shown in lowercase). PAM sites shown in lowercase and in bold. N; number.

Target sgRNA	Target gene	Off-target site sequence	Gene name in <i>Arabidopsis</i>	Poplar gene ID	Mismatches (N)	Events (N)	Mutations (N)
<i>LFY</i> -sg1	<i>PLFY</i>	GtCCCCGCCggAGCAGCCAC cgg	Ubiquitin-conjugating enzyme 19 (UBC19)	Potri.001G254500	3	19	0
		GtCCCtGCCggAGCAGCCAC cgg	Ubiquitin-conjugating enzyme 20 (UBC20)	Potri.009G049600	4	19	0
<i>AG</i> -sg2	<i>PAG1</i> , <i>PAG2</i>	cGaGAAAGGaGGAGATCAAG gagg	Ubiquitin-specific protease 16 (USP36)	Potri.005G156900	3	39	0
		GaGGAAAGaTtGAGATCAAG gagg	<i>SEEDSTICK</i> (<i>STK</i>)	Potri.013G104900	3	39	0
		GaGGAAAGaTtGAGATCAAG gagg	<i>SEEDSTICK</i> (<i>STK</i>)	Potri.019G077200	3	39	0

Appendix C Supplementary material for Chapter 4

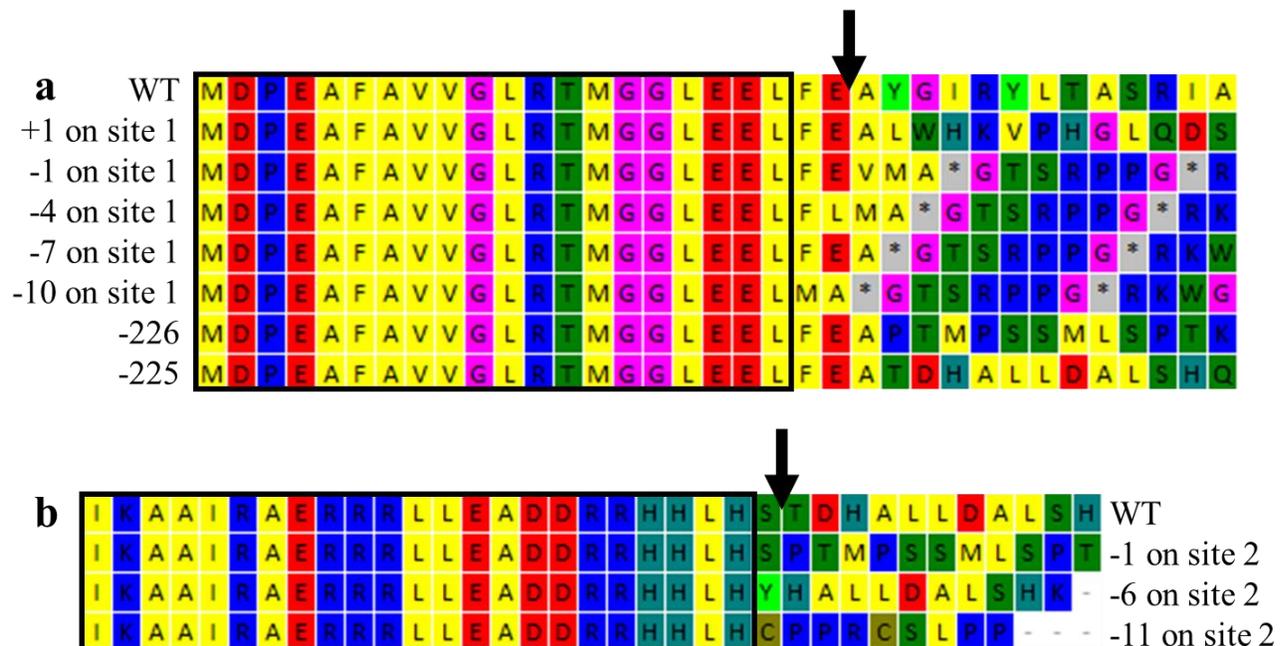
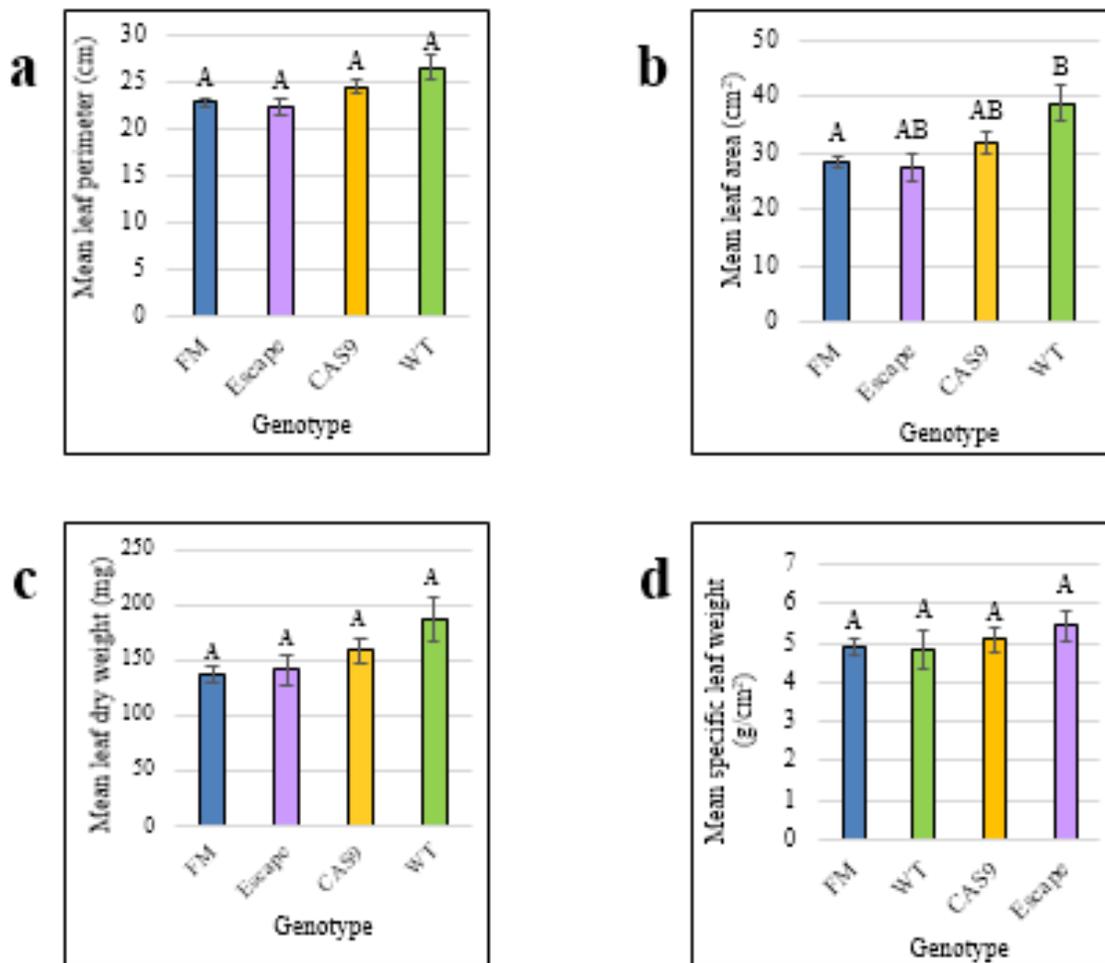


Fig. S4.1 Examples of partial peptide alignment of the N-terminal motif in mutants observed. **(a)** Peptide sequence of the first 35 amino acids of the N-terminal motif in WT and FM events transformed with *ELFY*-sg1 (i.e., events with mutations on site one) and *ELFY*-sg1sg2 (i.e., events with a deletion from site one through site two) are shown. The last two peptide sequences shown have large deletions that reduced their exon size from 109 to 35. Mutations on site one are expected to have amino acids modifications after the arrow. **(b)** Peptide sequence of the last 35 amino acids of the N-terminal motif in WT and FM events transformed with *ELFY*-sg2 (i.e., events with mutations on site 2). The black rectangles show the amino acids that remained unmodified. The black arrows indicate where the modifications to the peptide sequence are expected to happen for each target. All of these peptide sequences belonged to confirmed FM events.



e

Contrast	Perimeter (cm)	P-values	Area (cm ²)	P-values	Dry weight (mg)	P-values	Specific weight (g/cm ²)	P-values
Cas9 - escape	2.120	0.352	4.198	0.522	16.700	0.848	-0.348	0.883
Cas9 - FM	1.687	0.292	3.372	0.452	21.277	0.528	0.186	0.954
Cas9 - WT	-2.006	0.575	-7.161	0.283	-28.374	0.695	0.259	0.967
Escape - FM	-0.433	0.975	-0.826	0.988	4.587	0.994	0.534	0.602
Escape - WT	-4.126	0.113	-11.359	0.070	-45.064	0.398	0.607	0.761
FM - WT	-3.693	0.097	-10.533	0.049	-49.651	0.218	0.073	0.999

Fig. S4.2 Leaf phenotypes of potted plants in WT trial. (a) Mean leaf perimeter of predicted FM plants and the three controls (escapes, WT, and Cas9-only). (b) Mean leaf area. (c) Mean leaf dry weight. (d) Mean specific leaf weight. (e) Table of estimated mean differences and p-values corresponding to the t-test on the means of each contrast for leaf perimeter, leaf area, leaf dry weight, and specific leaf weight. Error bars represent \pm SE of means. Different letters above bars indicated statistical significance below the 5% level based on Student's t-test. Cas9, transgenic but no sgRNAs. Escape, non-transgenic but *Agrobacterium* cocultivated and regenerated. FM, flowering mutant. WT, wild type, not cocultivated but micropropagated.

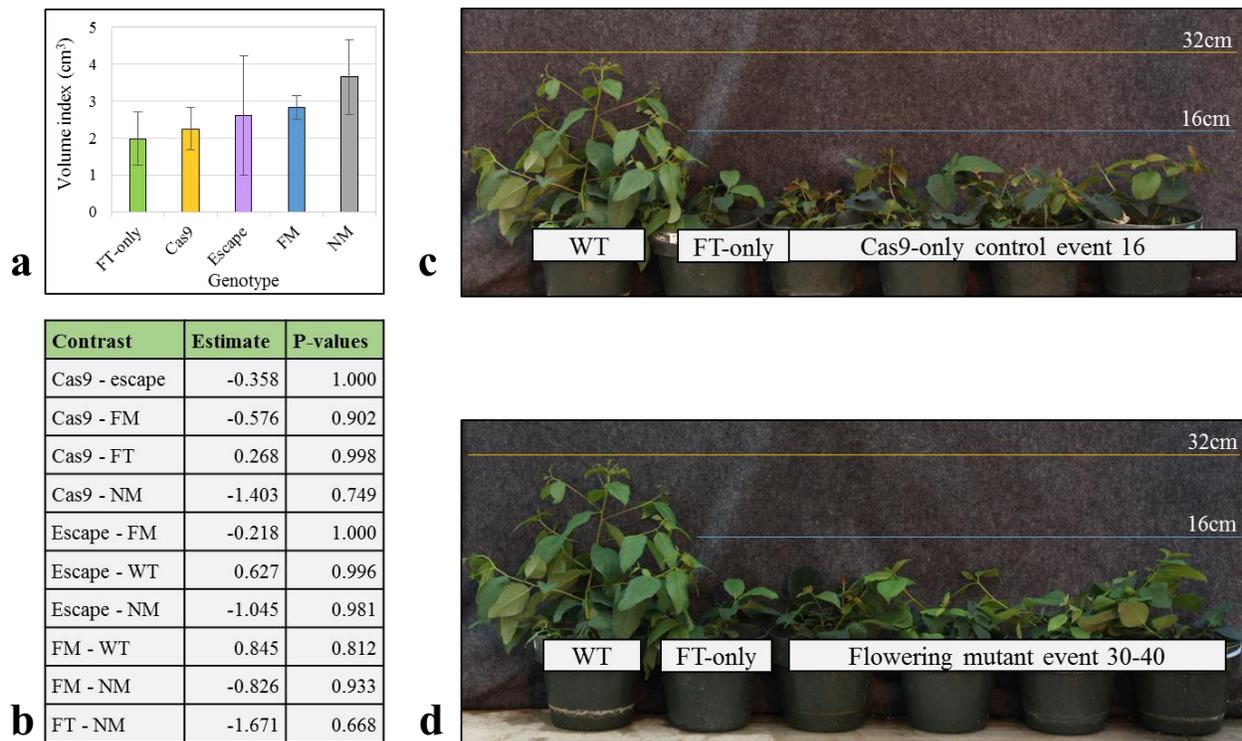
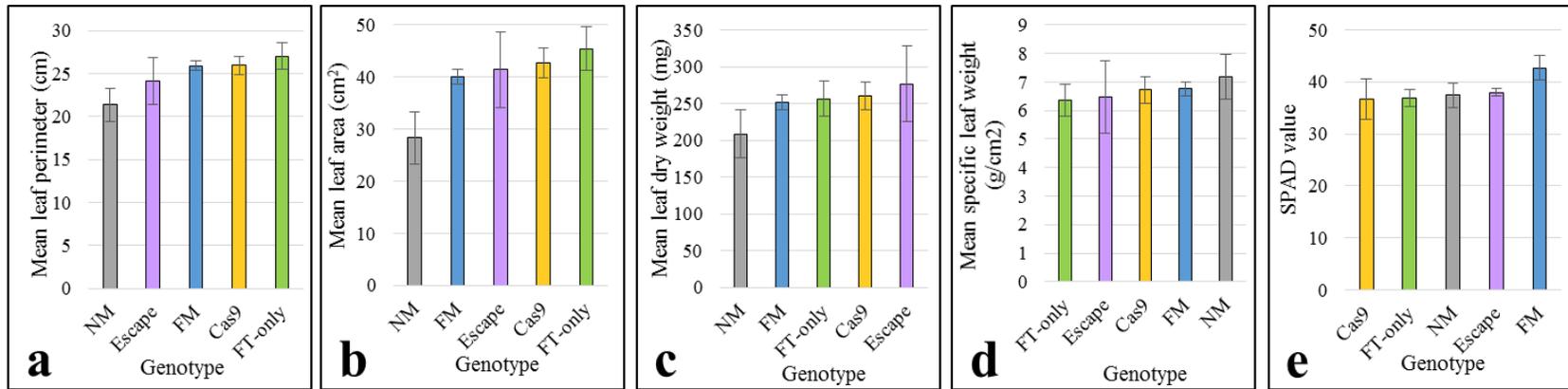


Fig. S4.3 Stem growth was reduced in plants that flowered precociously due to *AtFT* overexpression, but did not differ due to *ELFY* mutagenesis. **(a)** Mean stem volume index (height x diameter²) and standard error for the flowering mutants, the non-mutated events, and the three control groups. Error bars represent \pm SE of means. Cas9, transgenic lines that do not contain sgRNAs. Escape, non-transgenic but *Agrobacterium* cocultivated and regenerated lines. FM, flowering mutant. FT-only, original flowering background (i.e., *AtFT*-only). NM, transformed with Cas9 and sgRNA(s) but peptide sequence is similar to WT. **(b)** Table of estimated mean differences and p-values corresponding to a Student's t-test on the means for each contrast. **(c)** Image of potted reference WT ramet, *AtFT*-only flowering reference, and the four ramets corresponding to Cas9-only event 16. **(d)** Image of potted reference WT ramet, *AtFT*-only flowering reference, and the four ramets of FM event 167. The yellow and blue lines in both photographs are at 32 and 16 cm height respectively.



Fig. S4.4 Developmental sequence of flower formation in the greenhouse. **(a)** Flower buds and flowers of *AtFT* Cas9 events. The entire sequence spans approximately four months. **(b)** Flower buds of FM events. The entire sequence spans approximately seven months.



Contrast	Perimeter (cm)	P-values	Area (cm ²)	P-values	Dry weight (mg)	P-values	Specific weight (g/cm ²)	P-values	SPAD	P-values
Cas9 - escape	1.860	0.969	1.347	1.000	-16.738	0.998	0.244	1.000	0.171	1.000
Cas9 - FM	0.065	1.000	2.644	0.917	8.710	0.993	-0.038	1.000	-1.109	0.968
Cas9 - FT	-1.072	0.980	-2.734	0.981	3.667	1.000	0.355	0.987	-5.827	0.281
Cas9 - NM	4.631	0.233	14.435	0.117	51.280	0.654	-0.462	0.986	-0.515	1.000
Escape - FM	-1.795	0.967	1.297	1.000	25.448	0.980	-0.283	1.000	-1.280	0.998
Escape - FT	-2.932	0.886	-4.081	0.988	20.405	0.996	0.110	1.000	-5.998	0.684
Escape - NM	2.771	0.919	13.088	0.585	68.018	0.793	-0.706	0.999	-0.686	1.000
FM - FT	-1.137	0.960	-5.378	0.737	-5.043	1.000	0.393	0.964	-4.718	0.359
FM - NM	4.565	0.167	11.791	0.187	42.570	0.723	-0.424	0.986	0.594	0.999
FT - NM	5.703	0.182	17.169	0.099	47.613	0.763	-0.817	0.915	5.312	0.531

f

Fig. S4.5 Leaf phenotypes of potted plants in *FT* trial.

Fig. S4.5 Leaf phenotypes of potted plants in *FT* trial. **(a)** Mean leaf perimeter of predicted flowering mutants, the non-mutated events, and the three control groups (escapes, WT, and Cas9-only). **(b)** Mean leaf area. **(c)** Mean leaf dry weight. **(d)** Mean specific leaf weight. **(e)** SPAD, a proxy for chlorophyll content. **(f)** Table of estimated mean differences and p-values corresponding to the Student's t-test on the means of each contrast for leaf perimeter, leaf area, leaf dry weight, specific leaf weight, and SPAD. Error bars represent \pm SE of means. Cas9, transgenic lines that do not contain sgRNAs. Escape, non-transgenic but *Agrobacterium* cocultivated and regenerated lines. FM, flowering mutant. FT-only, original flowering background (i.e., *AtFT*-only). NM, transformed with Cas9 and sgRNA(s) but peptide sequence is similar to WT.

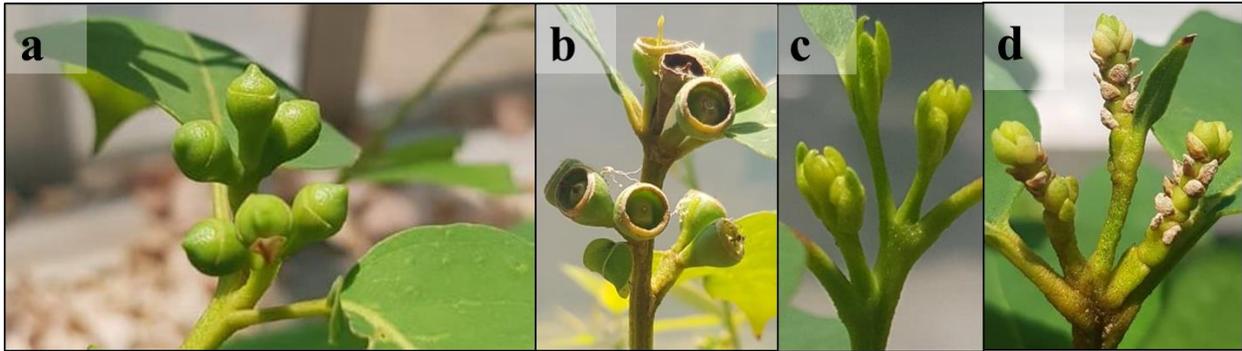


Fig. S4.6 Flower buds and flowers of *AtFT*-only and FM events in a greenhouse trial at the University of Pretoria in South Africa. The flowering controls (i.e., *AtFT*-only and *AtFT* Cas9) and FM events had essentially the same flowering phenotypes seen in Oregon. (a) Flowers buds from *AtFT*-only event with bracts and calicine opercula shed. (b) Developing seed capsules from *AtFT*-only event with stamens shed and stigmas dried out. (c) Three umbels early in development with early buds from FM event. (d) Umbel with four mutant flowering buds from FM event with layers of pedicel-like and bract-like organs. The bract-like organs dry out and eventually fall off.

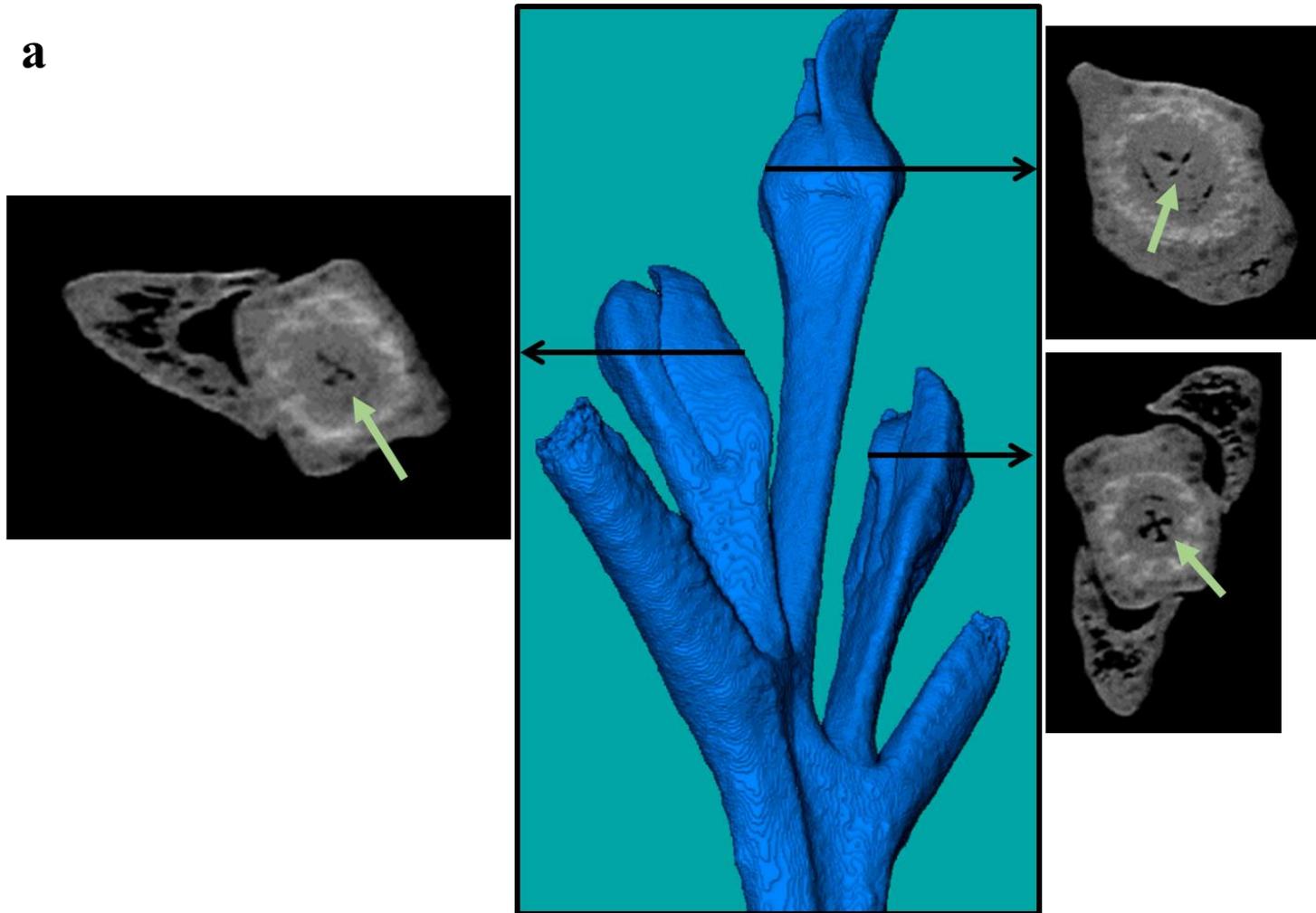


Fig. S4.7 3D representation of X-ray projections of inflorescences.

b

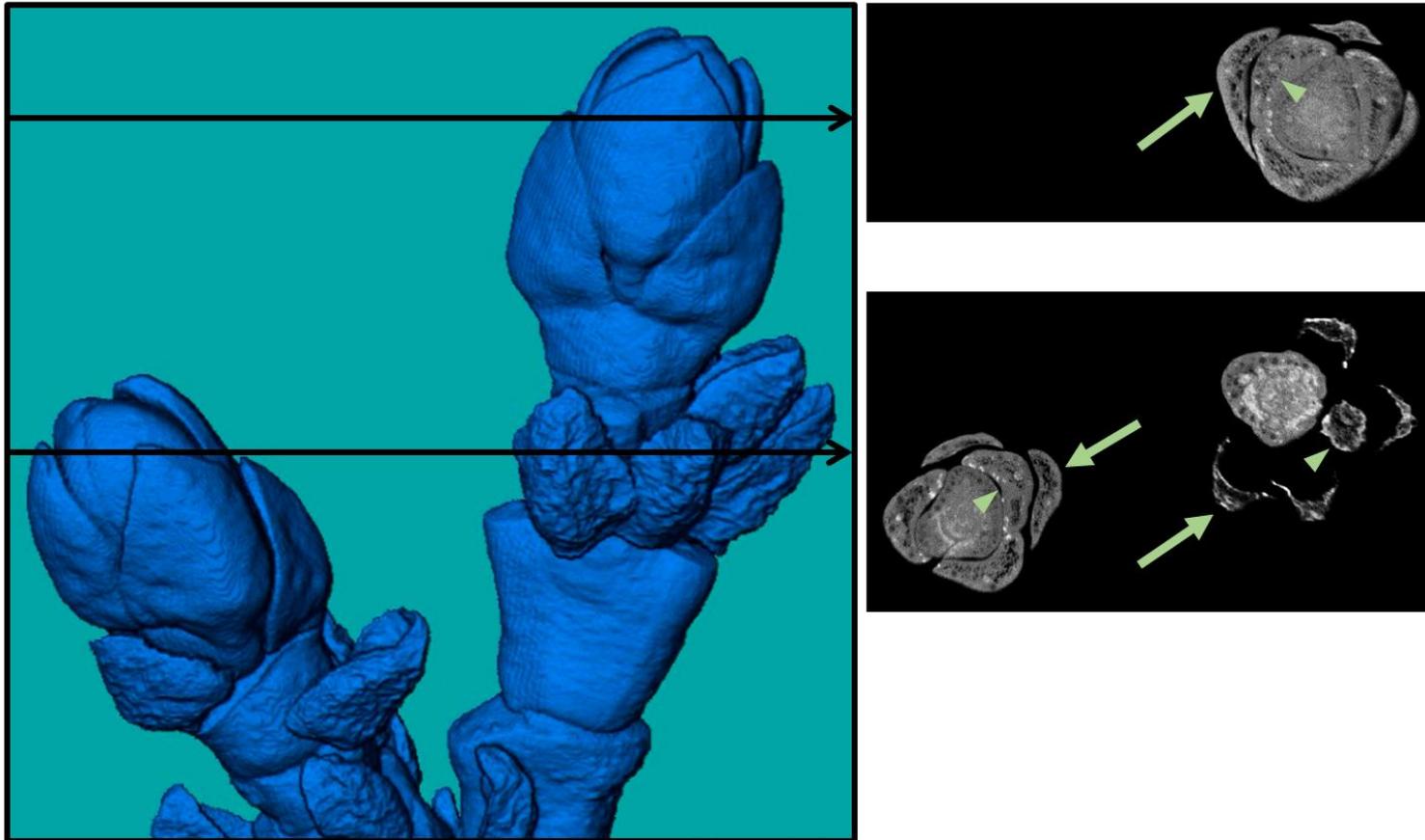


Fig. S4.7 3D representation of X-ray projections of inflorescences. **(a)** 3D representation of a young WT inflorescence after X-ray scanning. Arrows point to four locules where carpels were in development. The inflorescence was harvested from *AtFT*-only event 30-62. **(b)** 3D representation of a ten-month-old inflorescence from FM event 30-16 after X-ray scanning. Slices show repeated bract-like organs. Arrow points the outer most bract and arrowheads point to the next bract in the repeated succession. No reproductive organ development was seen in these or other cross-sections from the image.

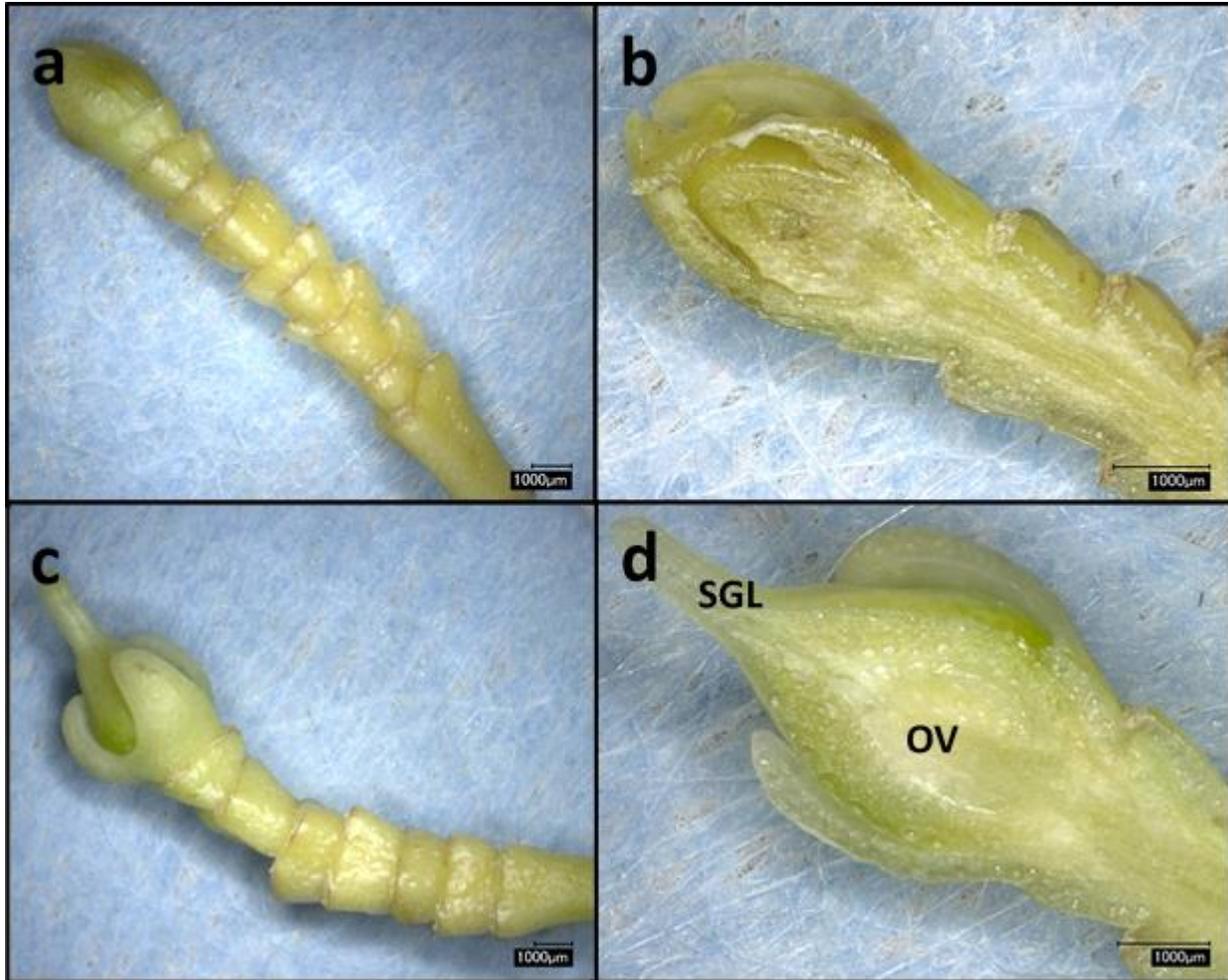


Fig. S4.8 Underdeveloped organs appeared occasionally in organless mutant plants. Six-month-old buds in event 4-66. **(a)** Long-lived bud with nine visible layered pedicel-like organs. **(b)** No reproductive organs seen in cross-section of bud in (a). **(c)** Long-lived bud with seven visible layered pedicel-like organs, bract-like organs covering a hypanthium-like structure. **(d)** Undeveloped ovary with ovules and a stigma-like organ seen inside dissected bud in (c). No male reproductive organs were visible. OV, ovary. SGL, stigma-like.

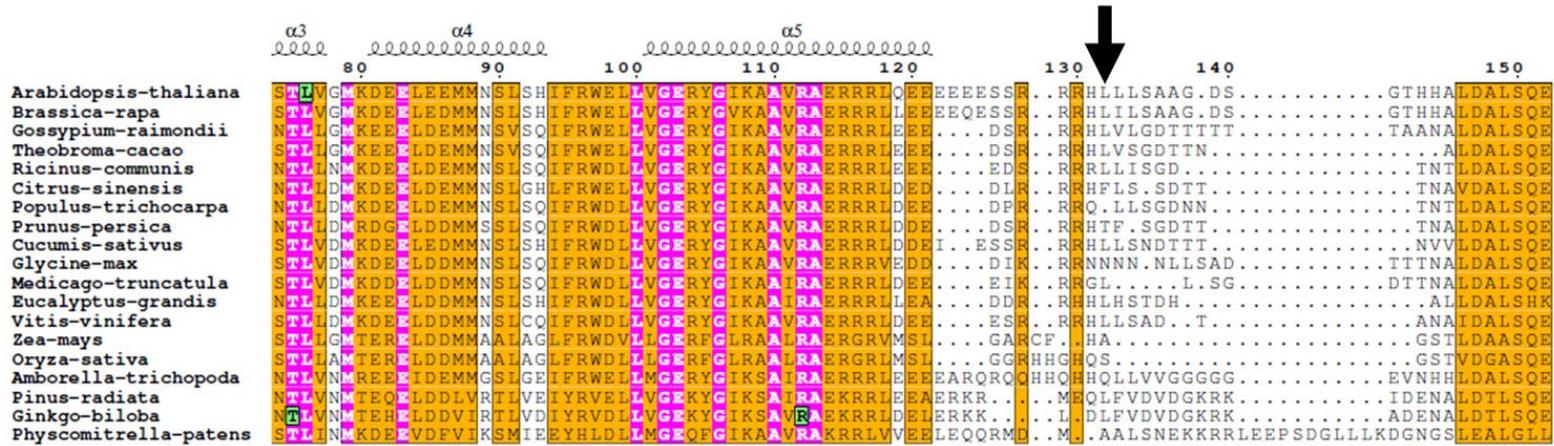
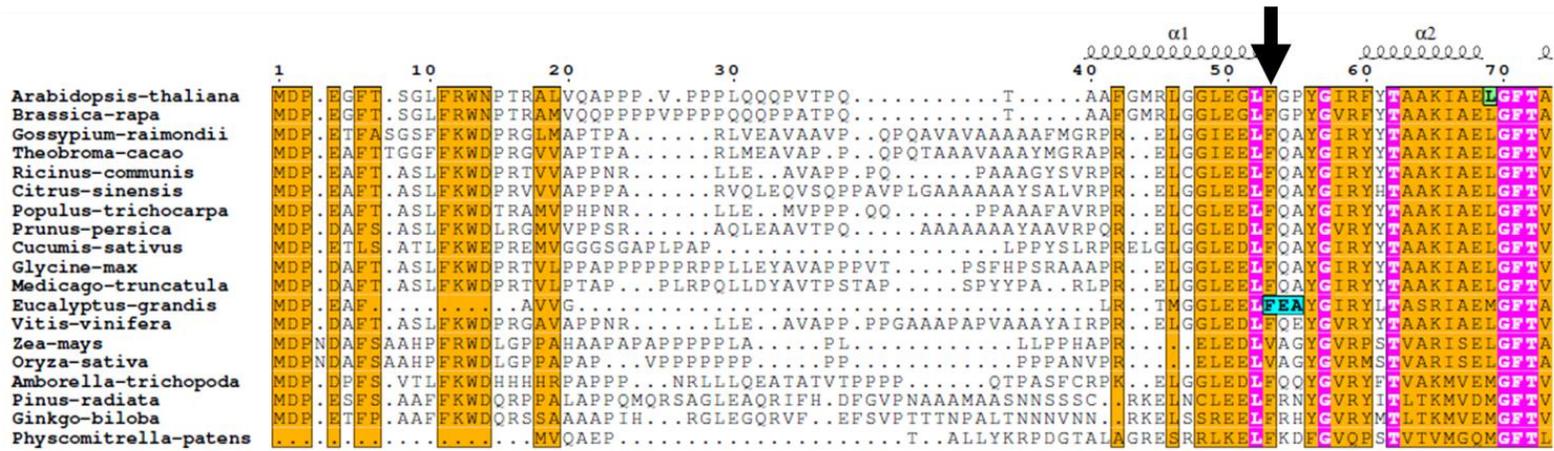


Fig. S4.9 Peptide alignment of the N-terminal domain in *LFY* and orthologous transcription factors. The three amino acids removed in event 4-8 are in bold and inside cyan boxes. The two amino acids removed in event 30-6 are the glutamic acid and alanine missing from event 4-8. The amino acids modified by Siriwardana and Lamb (2012) and Sayou *et al.* (2016) are in bold and inside green boxes. Orange and pink indicate 70 and 100% similarity across all the sequences respectively. Periods indicate no amino acid. The black arrows point to the location of the target sites where modifications to the peptide sequence would be expected to occur.

Table S4.1 Primers used for genotyping and sequencing.

Primer name	Primer sequence (5' to 3')	Use(s)	Reference
AtU626_F1	CTTCAAAAGTCCCACATCGC	Verifying transgene sequence; event genotyping	Elorriaga et al. 2018
sgRNA_R1	GCCGCCAGTGTGATGGATA	Verifying transgene sequence; event genotyping	Elorriaga et al. 2018
Cas9_F1	CACGACGGAGACTACAAGGA	Verifying transgene sequence; event genotyping	Elorriaga et al. 2018
RB_R1	CGGATAAACCTTTTCACGCC	Verifying transgene sequence; event genotyping	Elorriaga et al. 2018
Egrandis_F3	GCGCGAGAATGGATCCAGAA	Amplifying the <i>LFY</i> allele from <i>E. grandis</i>	This study
Egrandis_R1	GAGGCCGAGTTAAGTTACCTTT	Amplifying the <i>LFY</i> allele from <i>E. grandis</i>	This study
Euro_F3	GCGCGAGAATGGATCCAGAG	Amplifying the <i>LFY</i> allele from <i>E. urophylla</i>	This study
Euro_R1	GAGGCCGAGTTAAGGTACCTTG	Amplifying the <i>LFY</i> allele from <i>E. urophylla</i>	This study

Table S4.2 Gene names and IDs for qPCR experiments. The orthologs in *Eucalyptus* were determined from a previously published floral transcriptome Vining et al. (2015) and an unpublished RNA-seq database.

Gene name in <i>Arabidopsis</i>	Gene ID in <i>Arabidopsis</i>	Gene ID in <i>Eucalyptus</i>
<i>AGAMOUS (AG)</i>	At4G18960	Eucgr.E02863
<i>APETALA 3 (AP3)</i>	At3G54340	Eucgr.F01615
<i>CAULIFLOWER (CAL)</i>	At1G26310	Eucgr.I02059
<i>FLOWERING LOCUS T (FT)</i>	At1G65480	Eucgr.B01458
<i>FRUITFULL (FUL)</i>	At5G60910	Eucgr.B00634
<i>FRUITFULL (FUL)</i>	At5G60910	Eucgr.K02547
<i>glyceraldehyde-3-phosphate dehydrogenase (GAPDH)</i>	At1G13440	Eucgr.H04673
<i>LEAFY (LFY)</i>	At5G61850	Eucgr.K02192
<i>PISTILLATA (PI)</i>	At5G20240	Eucgr.E01007
<i>SEEDSTICK (STK)</i>	At4G09960	Eucgr.F02981
<i>SHATTERPROOF 2 (SHP2)</i>	At2G42830	Eucgr.K01195
<i>SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 3 (SPL3)</i>	At2G33810	Eucgr.D02505
<i>SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 (SPL9)</i>	At2G42200	Eucgr.K01828

Table S4.3 Predicted loss-of-function (LOF) rates based on the number of frame-shifts, large deletions (i.e. ≥ 222 bp), and deletions of essential amino acids. LOF, loss-of-function (i.e., FM or flowering mutant). WT, predicted wild-type peptide.

Population	Total events (alleles)	Predicted phenotype	N° events
FT LFY- CRISPR	59 (118)	LOF	53 (90%)
		WT	6 (10%)
WT LFY- CRISPR	9 (18)	LOF	9 (100 %)
		WT	0 (0%)
All eucalypt	68 (136)	LOF	62 (91%)
		WT	6 (9%)

Table S4.4 Phenotypes seen in FM (flowering mutant) events kept in the GH. Each row corresponds to one FM event. CRISPR corresponds to the specific CRISPR Cas9 nuclease in the FM event. The mutations seen in each allele and the flower phenotype of the FM event are also specified. The total number of ramets and the number of ramets that flowered is also recorded. EARLY ORGANS, sterile underdeveloped ovules and/or stamens present after only two or three layered pedicels. LATE ORGANS, sterile underdeveloped ovules present after three to five layered pedicels. ORGANLESS, no underdeveloped reproductive organs in five layered pedicels. NONE, plants did not flower.

Event	CRISPR	Egrandis site 1	Egrandis site 2	Euro site 1	Euro site 2	N° of ramets	N° of ramets that flowered	Floral phenotype
4-46	EgLFY-sg1sg2	5bp deletion, 1bp insertion + chimera	1bp deletion	228bp deletion		5	5 (100.0 %)	EARLY ORGANS
4-55	EgLFY-sg1sg2	225bp deletion		222bp insertion		6	6 (100.0 %)	EARLY ORGANS
30-30	EgLFY-sg1sg2	7bp deletion	2bp deletion	228bp deletion		3	3 (100.0 %)	EARLY ORGANS
30-33	EgLFY-sg1sg2	261bp deletion		264bp deletion		5	5 (100.0%)	EARLY ORGANS
30-42	EgLFY-sg1sg2	Inversion (225bp)		228bp deletion		2	2 (100.0 %)	EARLY ORGANS
4-8	EgLFY-sg1	9bp deletion	NA	1bp insertion	NA	3	3 (100.0 %)	LATE ORGANS
4-33	EgLFY-sg1	1bp insertion	NA	7bp deletion	NA	4	4 (100.0 %)	LATE ORGANS
4-54	EgLFY-sg1	4bp deletion	NA	1bp deletion	NA	6	6 (100.0 %)	LATE ORGANS
4-59	EgLFY-sg1	32bp deletion	NA	1bp insertion	NA	6	6 (100.0 %)	LATE ORGANS
4-60	EgLFY-sg1	1bp insertion	NA	10bp deletion	NA	4	4 (100.0 %)	LATE ORGANS
30-6	EgLFY-sg1	6bp deletion	NA	3bp deletion	NA	1	1 (100.0 %)	LATE ORGANS
30-11	EgLFY-sg1	4bp deletion	NA	4bp deletion	NA	6	6 (100.0 %)	LATE ORGANS
4-9	EgLFY-sg1sg2	31bp deletion	3bp deletion	229bp deletion		6	6 (100.0 %)	LATE ORGANS
4-66	EgLFY-sg1sg2	7bp deletion	1bp deletion	8bp deletion	1bp insertion	6	6 (100.0 %)	LATE ORGANS
4-10	EgLFY-sg1	1bp insertion	NA	10bp deletion	NA	4	4 (100.0 %)	ORGANLESS
4-12	EgLFY-sg1	4bp deletion	NA	2bp deletion	NA	3	3 (100.0 %)	ORGANLESS

Table S4.4 Phenotypes seen in FM (flowering mutant) events kept in the GH (continued).

4-34	EgLFY-sg1	1bp insertion	NA	7bp deletion	NA	1	1 (100.0 %)	ORGANLESS
30-2	EgLFY-sg1	1bp insertion	NA	10bp deletion	NA	3	3 (100.0 %)	ORGANLESS
30-10	EgLFY-sg1	8bp deletion	NA	5bp deletion	NA	5	5 (100.0 %)	ORGANLESS
4-4	EgLFY-sg2	NA	1bp deletion	NA	7bp deletion	2	2 (100.0 %)	ORGANLESS
4-74	EgLFY-sg2	NA	1bp deletion	NA	1bp deletion	3	3 (100.0 %)	ORGANLESS
30-4	EgLFY-sg2	NA	2bp deletion	NA	5bp deletion	2	2 (100.0 %)	ORGANLESS
30-5	EgLFY-sg2	NA	1bp deletion	NA	11bp deletion	3	3 (100.0 %)	ORGANLESS
30-31	EgLFY-sg2	NA	1bp insertion	NA	23bp deletion	2	2 (100.0 %)	ORGANLESS
30-45	EgLFY-sg2	NA	1bp deletion	NA	1bp deletion	3	3 (100.0 %)	ORGANLESS
4-37	EgLFY-sg1sg2	(261 bp deletion) and (2bp insertion)		36bp insertion	1bp deletion	3	3 (100.0 %)	ORGANLESS
4-65	EgLFY-sg1sg2	22bp deletion	1bp deletion	Inversion (228 bp)		4	2 (50.0 %)	ORGANLESS
30-1	EgLFY-sg1sg2	Inversion (225bp)		8bp deletion	3bp deletion	5	5 (100.0 %)	ORGANLESS
30-2	EgLFY-sg1sg2	2bp deletion	1bp deletion	3bp deletion	1bp deletion	5	5 (100.0 %)	ORGANLESS
30-3	EgLFY-sg1sg2	2bp insertion	1bp insertion	7bp deletion	1bp deletion	1	1 (100.0 %)	ORGANLESS
30-16	EgLFY-sg1sg2	226bp deletion		1bp insertion	1bp deletion	1	1 (100.0 %)	ORGANLESS
30-40	EgLFY-sg1sg2	4bp deletion	1bp deletion	228bp deletion		2	2 (100.0 %)	ORGANLESS
4-1	EgLFY-sg2	NA	3bp deletion	NA	3bp deletion	7	7 (100.0 %)	WT – fertile
4-7	EgLFY-sg2	NA	15bp insertion	NA	16bp deletion	3	2 (66.7 %)	WT – fertile
4-72	EgLFY-sg2	NA	3bp deletion	NA	5bp deletion	6	6 (100.0 %)	WT – fertile
30-41	EgLFY-sg2	NA	1bp deletion	NA	6bp deletion	6	6 (100.0 %)	WT – fertile

Table S4.4 Phenotypes seen in FM (flowering mutant) events kept in the GH (continued).

4-17	EgLFY-sg1	7bp deletion	NA	7bp deletion	NA	6	6 (100.0 %)	NONE
4-18	EgLFY-sg1	7bp deletion	NA	7bp deletion	NA	6	6 (100.0 %)	NONE
4-88	EgLFY-sg1	13bp deletion	NA	7bp deletion	NA	6	6 (100.0 %)	NONE
4-24	EgLFY-sg1sg2	Inversion (225bp)		12bp deletion	61bp deletion	4	4 (100.0 %)	NONE
4-41	EgLFY-sg1sg2	225bp deletion		1bp insertion	NA	4	4 (100.0 %)	NONE
30-19	EgLFY-sg1sg2	226bp deletion		264bp deletion		3	3 (100.0 %)	NONE

Methods S4.1 Allele-specific PCR recipe and thermocycler program.

Each 20 μl reaction contained 0.15 μl of Econotaq (www.lucigen.com), 0.15 μl of bovine serum albumin (1% BSA), 1 μl of forward and reverse primers (10 μM each), 1 μl of deoxynucleotide triphosphate (dNTP) mix (2.5 μM each), 2 μl of 10X Econotaq Reaction Buffer, 13.7 μl of water, and 2 μl of DNA template (total between 100 and 200 ng). The thermocycler program conditions were: 5 min at 95°C; 35 cycles of 30s at 95°C, 30s at 57°C and 30s at 72°C; and 10 min at 72°C.

Appendix D Supplementary material for Chapter 5

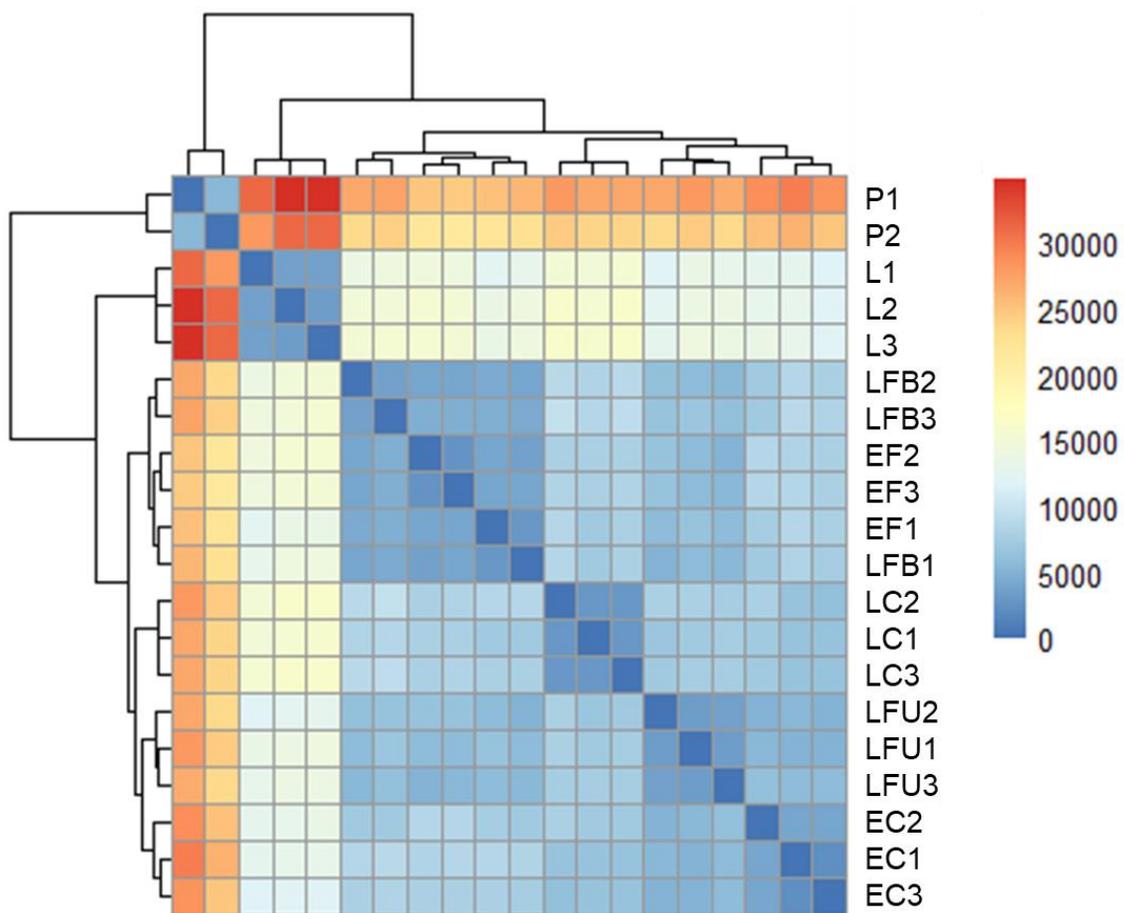


Fig. S5.1 Poisson distance clustering of tissues.

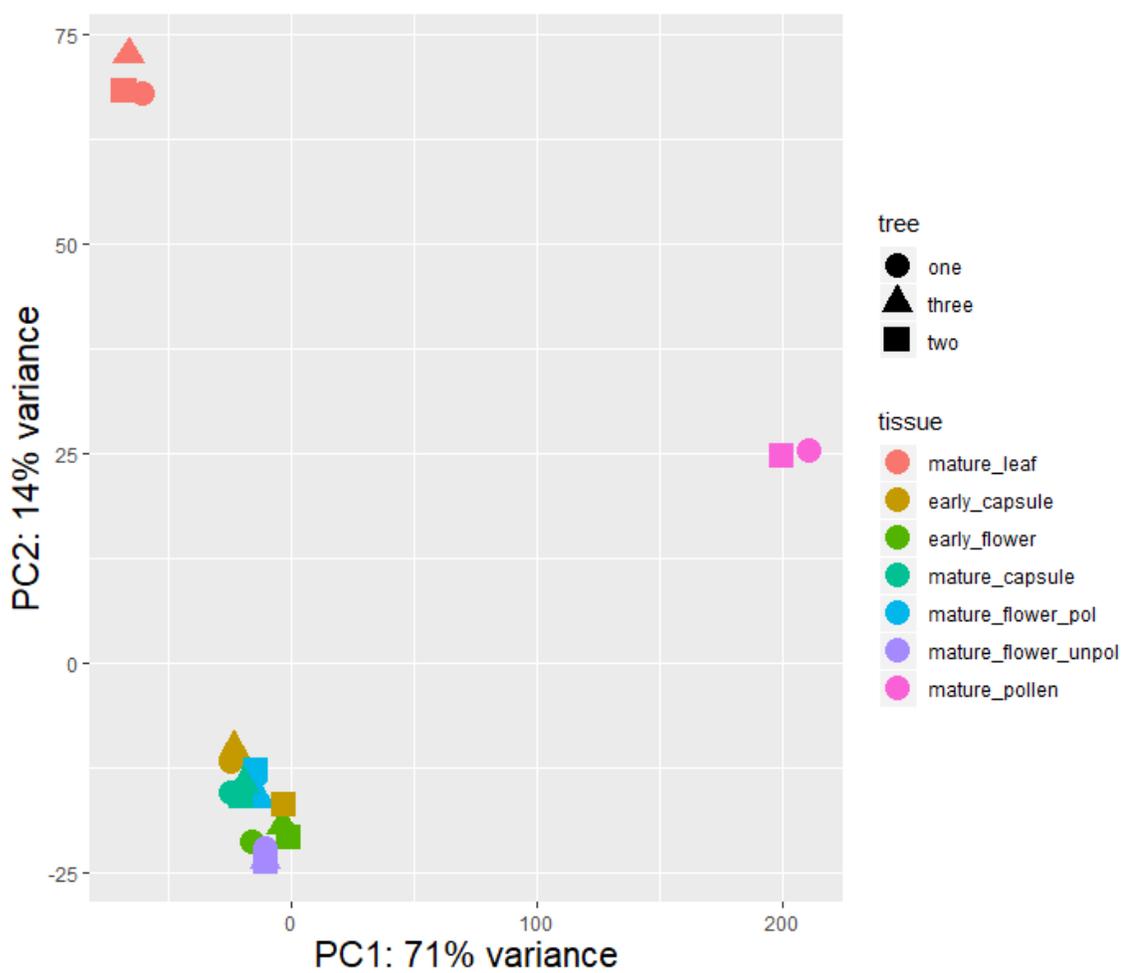


Fig. S5.2 Examination of variation among tissues including pollen and leaf.

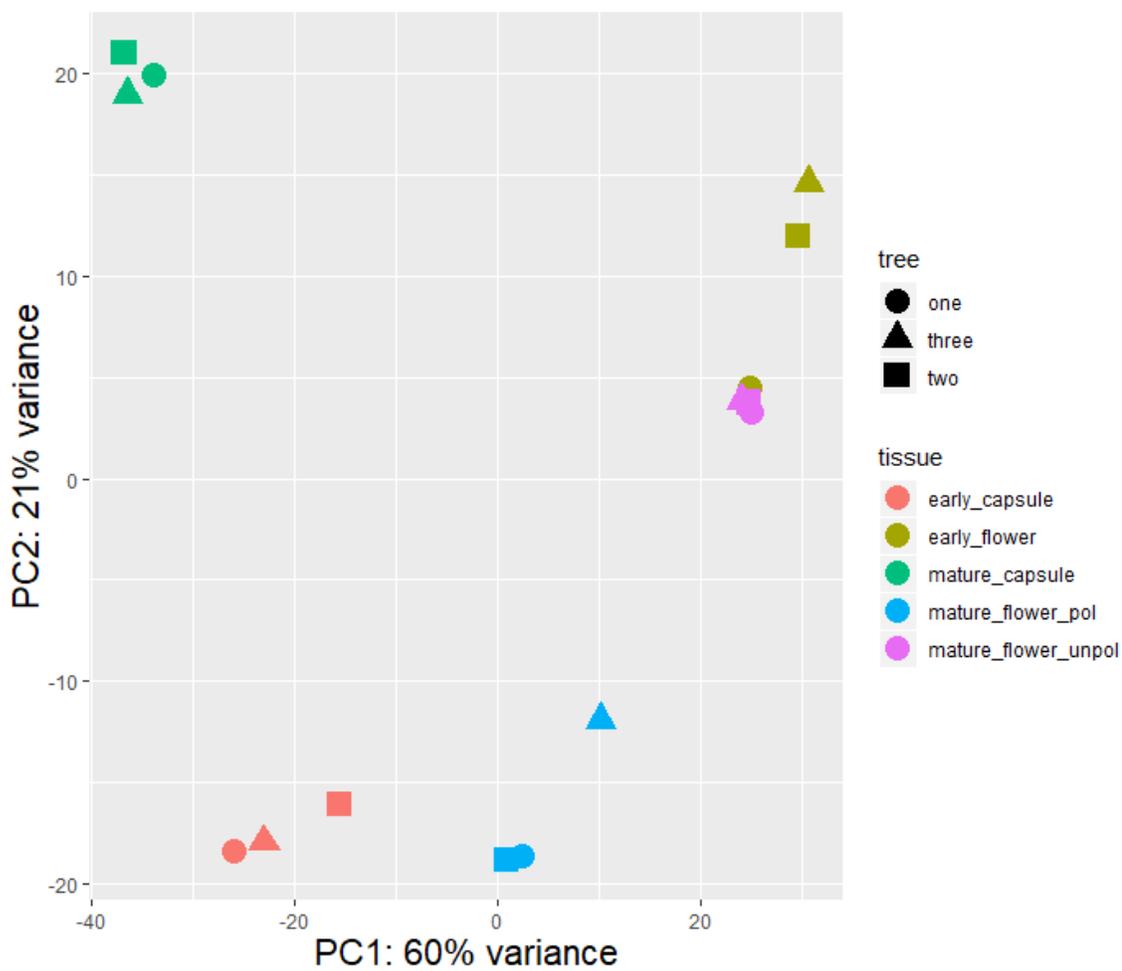


Fig. S5.3 Examination of variation among tissues excluding pollen and leaf.

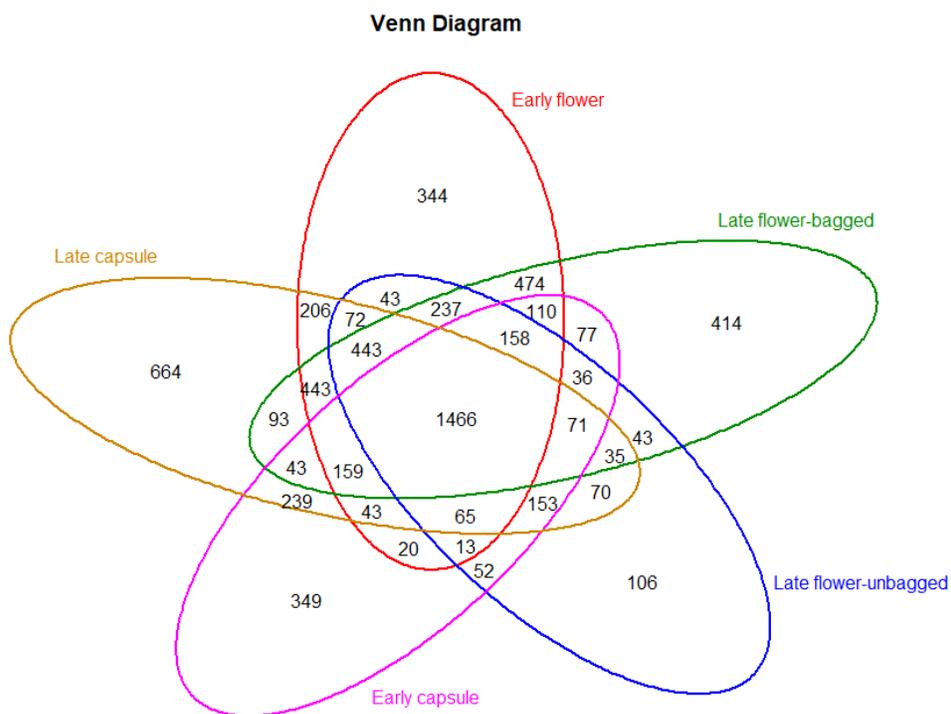


Fig. S5.5 Venn diagram of downregulated genes in flowers and capsules (LFC cutoff < -1, FDR < 0.05).

Table S5.1 The 20 most upregulated genes in early flower vs leaf.

EF vs L	brief explanation
Eucgr.A00869	(1 of 4) 5.4.2.12 - Phosphoglycerate mutase (2,3-diphosphoglycerate-independent) / Phosphoglyceromutase
Eucgr.A01881	(1 of 29) PTHR11746//PTHR11746:SF99 - O-METHYLTRANSFERASE // SUBFAMILY NOT NAMED
Eucgr.A02619	(1 of 6) 1.1.1.331 - Secoisolariciresinol dehydrogenase
Eucgr.A02978	(1 of 12) 4.2.1.78 - (S)-norcoclaurine synthase / (S)-norlaudanoline synthase
Eucgr.B01406	(1 of 7) PTHR31752:SF2 - AUXIN EFFLUX CARRIER COMPONENT 8-RELATED
Eucgr.B03515	(1 of 1) PTHR11945//PTHR11945:SF216 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.C03863	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.E01594	(1 of 1) 1.13.11.68 - 9-cis-beta-carotene 9',10'-cleaving dioxygenase
Eucgr.F01487	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.F03472	(1 of 14) PF04520 - Senescence regulator (Senescence_reg)
Eucgr.H01188	(1 of 9) PTHR30540//PTHR30540:SF25 - OSMOTIC STRESS POTASSIUM TRANSPORTER // SUBFAMILY NOT NAMED
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.H04154	(1 of 1) PTHR10791:SF52 - BIDIRECTIONAL SUGAR TRANSPORTER SWEET9
Eucgr.I02058	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.I02144	(1 of 15) PTHR22835:SF262 - GDSL ESTERASE/LIPASE 5-RELATED
Eucgr.J01518	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.J02217	(1 of 1) PTHR10641:SF469 - TRANSCRIPTION FACTOR MYB21-RELATED
Eucgr.K02547	(1 of 2) PTHR11945//PTHR11945:SF166 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.K03319	(1 of 1) K17911 - beta-carotene isomerase (DWARF27)
Eucgr.L03070	(1 of 2) 3.2.1.154 - Fructan beta-(2,6)-fructosidase / Levanase

Table S5.2 The 20 most upregulated genes in late flower bagged vs leaf.

LFB vs L	brief explanation
Eucgr.A00548	(1 of 2) PF00205//PF02776 - Thiamine pyrophosphate enzyme, central domain (TPP_enzyme_M) // Thiamine pyrophosphate enzyme, N-terminal TPP binding domain (TPP_enzyme_N)
Eucgr.A01881	(1 of 29) PTHR11746//PTHR11746:SF99 - O-METHYLTRANSFERASE // SUBFAMILY NOT NAMED
Eucgr.A02978	(1 of 12) 4.2.1.78 - (S)-norcoclaurine synthase / (S)-norlaudanoline synthase
Eucgr.B01406	(1 of 7) PTHR31752:SF2 - AUXIN EFFLUX CARRIER COMPONENT 8-RELATED
Eucgr.B03515	(1 of 1) PTHR11945//PTHR11945:SF216 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.C03863	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.E01594	(1 of 1) 1.13.11.68 - 9-cis-beta-carotene 9',10'-cleaving dioxygenase
Eucgr.F01487	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.F03472	(1 of 14) PF04520 - Senescence regulator (Senescence_reg)
Eucgr.G01437	(1 of 9) 1.14.13.93 - (+)-abscisic acid 8'-hydroxylase / ABA 8'-hydroxylase
Eucgr.H01188	(1 of 9) PTHR30540//PTHR30540:SF25 - OSMOTIC STRESS POTASSIUM TRANSPORTER // SUBFAMILY NOT NAMED
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.H04154	(1 of 1) PTHR10791:SF52 - BIDIRECTIONAL SUGAR TRANSPORTER SWEET9
Eucgr.I02058	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.I02144	(1 of 15) PTHR22835:SF262 - GDSL ESTERASE/LIPASE 5-RELATED
Eucgr.J01518	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.J02217	(1 of 1) PTHR10641:SF469 - TRANSCRIPTION FACTOR MYB21-RELATED
Eucgr.K02547	(1 of 2) PTHR11945//PTHR11945:SF166 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.K03319	(1 of 1) K17911 - beta-carotene isomerase (DWARF27)
Eucgr.L03070	(1 of 2) 3.2.1.154 - Fructan beta-(2,6)-fructosidase / Levanase

Table S5.3 The 20 most upregulated genes in late flower unconfined vs leaf.

LFU vs L	brief explanation
Eucgr.B03515	(1 of 1) PTHR11945//PTHR11945:SF216 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.C03863	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.C04253	(1 of 64) PF02298 - Plastocyanin-like domain (Cu_bind_like)
Eucgr.D01152	(1 of 7) K15404 - aldehyde decarbonylase (K15404)
Eucgr.F02981	(1 of 2) PTHR11945:SF170 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL11
Eucgr.F03472	(1 of 14) PF04520 - Senescence regulator (Senescence_reg)
Eucgr.G01685	(1 of 4) PTHR31321:SF10 - PECTINESTERASE 11-RELATED
Eucgr.H00268	(1 of 10) PTHR11926:SF271 - UDP-GLYCOSYLTRANSFERASE 74D1
Eucgr.H01188	(1 of 9) PTHR30540//PTHR30540:SF25 - OSMOTIC STRESS POTASSIUM TRANSPORTER // SUBFAMILY NOT NAMED
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.H04154	(1 of 1) PTHR10791:SF52 - BIDIRECTIONAL SUGAR TRANSPORTER SWEET9
Eucgr.H04370	(1 of 18) PTHR11732//PTHR11732:SF253 - ALDO/KETO REDUCTASE // SUBFAMILY NOT NAMED
Eucgr.I01800	(1 of 10) K18980 - 2-methylene-furan-3-one reductase (EO)
Eucgr.I02058	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.J01446	(1 of 3) 1.1.1.39 - Malate dehydrogenase (decarboxylating) / Pyruvic-malic carboxylase
Eucgr.J01518	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.J02217	(1 of 1) PTHR10641:SF469 - TRANSCRIPTION FACTOR MYB21-RELATED
Eucgr.J02997	(1 of 5) K17302 - coatomer
Eucgr.K02547	(1 of 2) PTHR11945//PTHR11945:SF166 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.K02656	(1 of 3) 1.17.1.3 - Leucoanthocyanidin reductase / Leucocyanidin reductase

Table S5.4 The 20 most upregulated genes in early capsule vs leaf.

EC vs L	brief explanation
Eucgr.A01798	(1 of 1) PTHR12771//PTHR12771:SF17 - ENGULFMENT AND CELL MOTILITY // SUBFAMILY NOT NAMED
Eucgr.B00628	(1 of 14) PF04833 - COBRA-like protein (COBRA)
Eucgr.B03515	(1 of 1) PTHR11945//PTHR11945:SF216 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.C03863	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.C03865	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.D01152	(1 of 7) K15404 - aldehyde decarboxylase (K15404)
Eucgr.E01271	(1 of 24) K13457 - disease resistance protein RPM1 (RPM1)
Eucgr.E01616	(1 of 8) PTHR31692:SF3 - BETA EXPANSIN 6-RELATED
Eucgr.F00501	(1 of 10) PTHR11206:SF102 - MATE EFFLUX FAMILY PROTEIN
Eucgr.F02981	(1 of 2) PTHR11945:SF170 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL11
Eucgr.F03472	(1 of 14) PF04520 - Senescence regulator (Senescence_reg)
Eucgr.G02639	(1 of 7) PTHR11941//PTHR11941:SF70 - ENOYL-COA HYDRATASE-RELATED // SUBFAMILY NOT NAMED
Eucgr.G03101	(1 of 2) PF01344//PF10539 - Kelch motif (Kelch_1) // Development and cell death domain (Dev_Cell_Death)
Eucgr.H00717	(1 of 8) PTHR11802:SF96 - SERINE CARBOXYPEPTIDASE-LIKE 20-RELATED
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.H04197	(1 of 5) PTHR22835:SF160 - GDSL ESTERASE/LIPASE LTL1
Eucgr.I02058	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.J01446	(1 of 3) 1.1.1.39 - Malate dehydrogenase (decarboxylating) / Pyruvic-malic carboxylase
Eucgr.K02547	(1 of 2) PTHR11945//PTHR11945:SF166 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.L02598	(1 of 16) PF05498 - Rapid ALkalinization Factor (RALF) (RALF)

Table S5.5 The 20 most upregulated genes in late capsule vs leaf.

LC vs L	brief explanation
Eucgr.B03515	(1 of 1) PTHR11945//PTHR11945:SF216 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.C00175	(1 of 1) PTHR31062:SF47 - XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE PROTEIN 21-RELATED
Eucgr.C03863	(1 of 3) PTHR11814:SF54 - SULFATE TRANSPORTER 3.5-RELATED
Eucgr.D01152	(1 of 7) K15404 - aldehyde decarbonylase (K15404
Eucgr.D01671	(1 of 2) PTHR31989:SF39 - NAC DOMAIN-CONTAINING PROTEIN 12
Eucgr.D01819	(1 of 1) PTHR10641//PTHR10641:SF667 - MYB-LIKE DNA-BINDING PROTEIN MYB // SUBFAMILY NOT NAMED
Eucgr.D01872	(1 of 18) PF03763 - Remorin
Eucgr.E00023	(1 of 4) PTHR24298:SF46 - CYTOCHROME P450 78A6-RELATED
Eucgr.E01250	(1 of 1) 1.23.1.2 - (+)-lariciresinol reductase / Pinoresinol/lariciresinol reductase
Eucgr.F01583	
Eucgr.F02733	(M=2) PF07366 - SnoaL-like polyketide cyclase
Eucgr.F02981	(1 of 2) PTHR11945:SF170 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL11
Eucgr.G01437	(1 of 9) 1.14.13.93 - (+)-abscisic acid 8'-hydroxylase / ABA 8'-hydroxylase
Eucgr.G03101	(1 of 2) PF01344//PF10539 - Kelch motif (Kelch_1) // Development and cell death domain (Dev_Cell_Death)
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.H03155	(1 of 12) PF06886 - Targeting protein for Xklp2 (TPX2) (TPX2)
Eucgr.I00923	(1 of 2) K09838 - zeaxanthin epoxidase (ZEP)
Eucgr.I02058	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.K02547	(1 of 2) PTHR11945//PTHR11945:SF166 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.L01734	(1 of 8) PTHR11709:SF93 - LACCASE-10-RELATED

Table S5.6 The 20 most upregulated genes in mature pollen vs leaf.

P vs L	brief explanation
Eucgr.A01065	(1 of 17) 3.2.1.15//3.2.1.67 - Polygalacturonase / Pectinase // Galacturan 1
Eucgr.A01967	(1 of 7) 3.2.1.151 - Xyloglucan-specific endo-beta-1
Eucgr.A02403	(1 of 2) PTHR34672:SF2 - ARABINOGALACTAN PEPTIDE 23-RELATED
Eucgr.C03978	(1 of 17) PF06749 - Protein of unknown function (DUF1218) (DUF1218)
Eucgr.D02268	(1 of 2) PTHR24206:SF22 - PROTEIN PLIM2A
Eucgr.D02451	(1 of 3) PTHR12290:SF18 - PROTEIN CORNICHON HOMOLOG 3-RELATED
Eucgr.E00014	(1 of 2) PTHR10286:SF10 - F11O4.12
Eucgr.E00162	(1 of 4) PTHR11913:SF30 - ACTIN-DEPOLYMERIZING FACTOR 10-RELATED
Eucgr.E01763	(1 of 16) PF05498 - Rapid ALkalinization Factor (RALF) (RALF)
Eucgr.E02463	(1 of 1) PTHR38378:SF1 - MYOSIN HEAVY CHAIN-LIKE PROTEIN
Eucgr.E02748	(1 of 2) PTHR31614:SF2 - F28N24.16 PROTEIN
Eucgr.F00977	(1 of 17) PF00234 - Protease inhibitor/seed storage/LTP family (Tryp_alpha_amyl)
Eucgr.F01074	(1 of 2) PF06764 - Protein of unknown function (DUF1223) (DUF1223)
Eucgr.F03264	(1 of 2) PTHR31614:SF2 - F28N24.16 PROTEIN
Eucgr.H01456	(1 of 2) PF09253 - Pollen allergen ole e 6 (Ole-e-6)
Eucgr.I01300	(1 of 5) 3.1.3.2//3.1.3.41//3.1.3.60 - Acid phosphatase / Phosphomonoesterase // 4-nitrophenylphosphatase // Phosphoenolpyruvate phosphatase
Eucgr.I02095	(1 of 2) PTHR34191:SF2 - STRESS-INDUCED PROTEIN KIN1-RELATED
Eucgr.J01446	(1 of 3) 1.1.1.39 - Malate dehydrogenase (decarboxylating) / Pyruvic-malic carboxylase
Eucgr.K00661	(1 of 1) PTHR31375:SF23 - EXOPOLYGALACTURONASE / GALACTURAN 1
Eucgr.L02598	(1 of 16) PF05498 - Rapid ALkalinization Factor (RALF) (RALF)

Table S5.7 The 20 most upregulated genes in cluster 1 when comparing amongst the flower and fruit tissues.

Genes in cluster 1	brief explanation
Eucgr.A02142	(1 of 1) PTHR12266:SF8 - CATION/CALCIUM EXCHANGER 2
Eucgr.B02783	(1 of 1) PTHR11662:SF235 - ANION TRANSPORTER 3, CHLOROPLASTIC-RELATED
Eucgr.C00665	(1 of 2) 5.3.99.5 - Thromboxane-A synthase / Thromboxane synthetase
Eucgr.C00786	(1 of 45) PF12708 - Pectate lyase superfamily protein (Pectate_lyase_3)
Eucgr.C02284	(1 of 1) 6.2.1.34 - Trans-feruloyl-CoA synthase / Trans-feruloyl-CoA synthetase
Eucgr.F00374	(1 of 41) PF10250 - GDP-fucose protein O-fucosyltransferase (O-FucT)
Eucgr.F01595	(1 of 4) PTHR23324:SF43 - PHOSPHATIDYLINOSITOL/PHOSPHATIDYLCHOLINE TRANSFER PROTEIN SFH6-RELATED
Eucgr.F02557	(1 of 1) K18368 - caffeoylshikimate esterase (CSE)
Eucgr.G01800	(1 of 5) PTHR31618:SF1 - MECHANOSENSITIVE ION CHANNEL PROTEIN 6-RELATED
Eucgr.G02924	(1 of 1) PTHR11615//PTHR11615:SF148 - NITRATE, FROMATE, IRON DEHYDROGENASE // SUBFAMILY NOT NAMED
Eucgr.G03056	(1 of 7) K05857 - phosphatidylinositol phospholipase C, delta (PLCD)
Eucgr.H00442	
Eucgr.H04617	(1 of 2) PTHR11945//PTHR11945:SF196 - MADS BOX PROTEIN // SUBFAMILY NOT NAMED
Eucgr.H04739	(1 of 3) PTHR14155:SF94 - RING-H2 FINGER PROTEIN ATL33-RELATED
Eucgr.I01397	(1 of 3) PTHR21576:SF17 - MAJOR FACILITATOR FAMILY PROTEIN
Eucgr.I01902	(1 of 21) PTHR23070:SF5 - AAA-TYPE ATPASE FAMILY PROTEIN-RELATED
Eucgr.I01905	(1 of 21) PTHR23070:SF5 - AAA-TYPE ATPASE FAMILY PROTEIN-RELATED
Eucgr.J01079	(1 of 9) 4.3.1.24 - Phenylalanine ammonia-lyase
Eucgr.K02331	(1 of 2) PTHR11695:SF449 - ALCOHOL DEHYDROGENASE-LIKE 6
Eucgr.L01848	(1 of 3) K17725 - sulfur dioxygenase (ETHE1)

Table S5.8 The 20 most upregulated genes in cluster 2 when comparing amongst the flower and fruit tissues.

Genes in cluster 2	brief explanation
Eucgr.B02126	(1 of 8) PF04576 - Zein-binding (Zein-binding)
Eucgr.C00150	(1 of 3) K01602 - ribulose-bisphosphate carboxylase small chain (rbcS)
Eucgr.C00899	(1 of 3) PTHR10071//PTHR10071:SF191 - TRANSCRIPTION FACTOR GATA GATA BINDING FACTOR // SUBFAMILY NOT NAMED
Eucgr.C03853	(1 of 6) K12355 - coniferyl-aldehyde dehydrogenase (REF1)
Eucgr.D00291	(1 of 2) PTHR10543:SF42 - 9-CIS-EPOXYCAROTENOID DIOXYGENASE NCED2, CHLOROPLASTIC-RELATED
Eucgr.D01384	(1 of 2) PTHR23257//PTHR23257:SF475 - SERINE-THREONINE PROTEIN KINASE // SUBFAMILY NOT NAMED
Eucgr.E03663	(1 of 17) PF00954//PF07714//PF08276 - S-locus glycoprotein domain (S_locus_glycop) // Protein tyrosine kinase (Pkinase_Tyr) // PAN-like domain (PAN_2)
Eucgr.F00458	(1 of 4) PTHR23029:SF40 - PHOSPHOGLYCERATE MUTASE-LIKE PROTEIN
Eucgr.F02398	(1 of 2) 1.14.13.78 - Ent-kaurene oxidase
Eucgr.F02959	(1 of 26) PF01476 - LysM domain (LysM)
Eucgr.F04466	(1 of 2) PTHR10836//PTHR10836:SF46 - GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE // SUBFAMILY NOT NAMED
Eucgr.G00854	(1 of 2) PTHR21337//PTHR21337:SF7 - PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE 1, 2 // SUBFAMILY NOT NAMED
Eucgr.G02223	(1 of 34) K00083 - cinnamyl-alcohol dehydrogenase (E1.1.1.195)
Eucgr.H00152	(1 of 1) PTHR22870:SF87 - REGULATOR OF CHROMOSOME CONDENSATION REPEAT-CONTAINING PROTEIN-RELATED
Eucgr.H04728	(1 of 5) PTHR22778:SF25 - ALPHA/BETA-HYDROLASES SUPERFAMILY PROTEIN
Eucgr.I00449	(1 of 37) 2.4.1.128 - Scopoletin glucosyltransferase
Eucgr.I01002	(1 of 10) PTHR10992:SF820 - METHYLESTERASE 1-RELATED
Eucgr.I01803	(1 of 10) K18980 - 2-methylene-furan-3-one reductase (EO)
Eucgr.J00678	(1 of 193) PF00646 - F-box domain (F-box)
Eucgr.K00311	(1 of 6) 3.2.1.17 - Lysozyme / Muramidase

Table S5.9 The 20 most upregulated genes in cluster 3 when comparing amongst the flower and fruit tissues.

Genes in cluster 3	brief explanation
Eucgr.A01633	(1 of 12) K01674 - carbonic anhydrase (cah)
Eucgr.A01877	(1 of 29) PTHR11746//PTHR11746:SF99 - O-METHYLTRANSFERASE // SUBFAMILY NOT NAMED
Eucgr.A02311	(1 of 4) 1.1.1.8 - Glycerol-3-phosphate dehydrogenase (NAD(+)) / NADH-dihydroxyacetone phosphate reductase
Eucgr.B00086	(1 of 11) PTHR31642:SF43 - OMEGA-HYDROXYPALMITATE O-FERULOYL TRANSFERASE
Eucgr.B00724	(1 of 2) PTHR31744:SF3 - GENOMIC DNA, CHROMOSOME 3, P1 CLONE: MYF24
Eucgr.B03663	(1 of 2) PTHR10024:SF243 - ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE-LIKE PROTEIN
Eucgr.C02185	(1 of 37) K16296 - serine carboxypeptidase-like clade I [EC:3.4.16.-] (SCPL-I)
Eucgr.C02352	(1 of 27) PTHR11802:SF29 - SERINE CARBOXYPEPTIDASE-LIKE 1-RELATED
Eucgr.C04222	(1 of 2) PTHR18896:SF11 - PHOSPHOLIPASE D ALPHA 1-RELATED
Eucgr.D00325	(1 of 8) PTHR31376:SF17 - PURINE PERMEASE 10-RELATED
Eucgr.F02644	(1 of 1) PF00394//PF04782//PF07731//PF07732 - Multicopper oxidase (Cu-oxidase) // Protein of unknown function (DUF632) (DUF632) // Multicopper oxidase (Cu-oxidase_2) // Multicopper oxidase (Cu-oxida...
Eucgr.G02700	(1 of 1) PTHR22835//PTHR22835:SF224 - ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Eucgr.H01970	(1 of 2) PTHR22835//PTHR22835:SF147 - ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Eucgr.H04937	(1 of 27) PTHR11709//PTHR11709:SF118 - MULTI-COPPER OXIDASE // SUBFAMILY NOT NAMED
Eucgr.I01041	(1 of 37) K16296 - serine carboxypeptidase-like clade I [EC:3.4.16.-] (SCPL-I)
Eucgr.I02438	(1 of 7) PTHR23500:SF43 - SUGAR TRANSPORTER ERD6-LIKE 4-RELATED
Eucgr.I02625	
Eucgr.J01662	(1 of 1) PTHR11972:SF73 - RESPIRATORY BURST OXIDASE HOMOLOG PROTEIN D
Eucgr.K01218	(1 of 2) PTHR11003//PTHR11003:SF140 - POTASSIUM CHANNEL, SUBFAMILY K // SUBFAMILY NOT NAMED
Eucgr.K02656	(1 of 3) 1.17.1.3 - Leucoanthocyanidin reductase / Leucocyanidin reductase

Table S5.10 The 20 most upregulated genes in cluster 4 when comparing amongst the flower and fruit tissues.

Genes in cluster 4	brief explanation
Eucgr.B03975	(1 of 4) PTHR23316:SF22 - ARMADILLO/BETA-CATENIN-LIKE REPEATS-CONTAINING PROTEIN-RELATED
Eucgr.C00757	(1 of 10) PF04784/PF14389 - Protein of unknown function, DUF547 (DUF547) // Leucine-zipper of ternary complex factor MIP1 (Lzipper-MIP1)
Eucgr.C01681	(1 of 2) PTHR27001:SF62 - PROTEIN KINASE FAMILY PROTEIN-RELATED
Eucgr.D00400	(1 of 109) PF04578/PF13968 - Protein of unknown function, DUF594 (DUF594) // Domain of unknown function (DUF4220) (DUF4220)
Eucgr.D00969	(1 of 1) PTHR10579/PTHR10579:SF47 - CALCIUM-ACTIVATED CHLORIDE CHANNEL REGULATOR // SUBFAMILY NOT NAMED
Eucgr.D01696	(1 of 1) PTHR13902/PTHR13902:SF63 - SERINE/THREONINE-PROTEIN KINASE WNK WITH NO LYSINE -RELATED // SUBFAMILY NOT NAMED
Eucgr.E00461	(1 of 1) PTHR11709:SF82 - MONOCOPPER OXIDASE-LIKE PROTEIN SKU5
Eucgr.F01987	(1 of 1) PTHR32472:SF12 - P-LOOP CONTAINING NUCLEOSIDE TRIPHOSPHATE HYDROLASES SUPERFAMILY PROTEIN
Eucgr.F03356	(1 of 2) PTHR13140:SF384 - MYOSIN-2
Eucgr.G00064	(1 of 66) PTHR23155/PTHR23155:SF554 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Eucgr.G03140	(1 of 1) PTHR27000:SF197 - INFLORESCENCE AND ROOT APICES RECEPTOR-LIKE KINASE
Eucgr.H03531	(1 of 1) 2.5.1.79 - Thermospermine synthase
Eucgr.H04336	(1 of 4) PTHR31734:SF11 - AUXIN-RESPONSIVE PROTEIN IAA1-RELATED
Eucgr.I00229	
Eucgr.I00565	(1 of 2) PTHR27001:SF105 - PROTEIN KINASE FAMILY PROTEIN-RELATED
Eucgr.I01402	(1 of 1) PTHR10795/PTHR10795:SF442 - PROPROTEIN CONVERTASE SUBTILISIN/KEXIN // SUBFAMILY NOT NAMED
Eucgr.J00581	(1 of 2) PTHR32295:SF12 - PROTEIN IQ-DOMAIN 15-RELATED
Eucgr.J03181	(1 of 1) PTHR10759/PTHR10759:SF4 - 60S RIBOSOMAL PROTEIN L34 // SUBFAMILY NOT NAMED
Eucgr.K01221	(1 of 2) PTHR27001:SF221 - PROTEIN KINASE FAMILY PROTEIN
Eucgr.K02134	(1 of 1) PTHR36760:SF1 - GENOMIC DNA, CHROMOSOME 3, P1 CLONE: MJL12

Table S5.11 The 20 most upregulated genes in cluster 1 when comparing EF, LFB, and LFU.

Genes in cluster 1	brief explanation
Eucgr.B00366	(1 of 5) PTHR13935//PTHR13935:SF59 - ACHAETE-SCUTE TRANSCRIPTION FACTOR-RELATED // SUBFAMILY NOT NAMED
Eucgr.C00426	
Eucgr.E00451	(1 of 5) 2.4.2.7 - Adenine phosphoribosyltransferase / Transphosphoribosidase
Eucgr.F00131	(1 of 3) PTHR23201:SF10 - GIBBERELLIN-REGULATED GASA/GAST/SNAKIN FAMILY PROTEIN-RELATED
Eucgr.F00219	(1 of 2) PTHR21493//PTHR21493:SF104 - CGI-141-RELATED/LIPASE CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Eucgr.F01487	(1 of 230) KOG0156 - Cytochrome P450 CYP2 subfamily
Eucgr.G00854	(1 of 2) PTHR21337//PTHR21337:SF7 - PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE 1, 2 // SUBFAMILY NOT NAMED
Eucgr.G01774	(1 of 2) PTHR10641:SF588 - TRANSCRIPTION REPRESSOR MYB4
Eucgr.H01202	(1 of 2) PTHR10168:SF57 - GLUTAREDOXIN-C13-RELATED
Eucgr.H01259	(1 of 2) PTHR12320:SF16 - PROTEIN PHOSPHATASE 2C 80-RELATED
Eucgr.H03170	(1 of 1) PTHR31734:SF30 - AUXIN-RESPONSIVE PROTEIN IAA14-RELATED
Eucgr.H04728	(1 of 5) PTHR22778:SF25 - ALPHA/BETA-HYDROLASES SUPERFAMILY PROTEIN
Eucgr.I00582	(1 of 6) PTHR10551//PTHR10551:SF13 - FASCIN // SUBFAMILY NOT NAMED
Eucgr.I00659	(1 of 1) PTHR31062:SF48 - XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE PROTEIN 6-RELATED
Eucgr.I01241	(1 of 2) 1.17.1.2 - 4-hydroxy-3-methylbut-2-enyl diphosphate reductase / HMBPP reductase
Eucgr.J00662	(1 of 1) 2.7.1.148 - 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase / CMK
Eucgr.J02319	
Eucgr.K02397	
Eucgr.K02977	(1 of 2) PTHR22952//PTHR22952:SF145 - CAMP-RESPONSE ELEMENT BINDING PROTEIN-RELATED // SUBFAMILY NOT NAMED
Eucgr.K03319	(1 of 1) K17911 - beta-carotene isomerase (DWARF27)

Table S5.12 The 20 most upregulated genes in cluster 2 when comparing EF, LFB, and LFU

Genes in cluster 2	brief explanation
Eucgr.A01002	(1 of 2) PTHR22883:SF53 - PROTEIN S-ACYLTRANSFERASE 18
Eucgr.A01877	(1 of 29) PTHR11746//PTHR11746:SF99 - O-METHYLTRANSFERASE // SUBFAMILY NOT NAMED
Eucgr.B02627	(1 of 260) PF00249 - Myb-like DNA-binding domain (Myb_DNA-binding)
Eucgr.B03663	(1 of 2) PTHR10024:SF243 - ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE-LIKE PROTEIN
Eucgr.B03746	(1 of 3) K13083 - flavonoid 3',5'-hydroxylase (CYP75A)
Eucgr.C03822	(1 of 25) PTHR27001:SF89 - PROLINE-RICH RECEPTOR-LIKE PROTEIN KINASE PERK1-RELATED
Eucgr.E00104	(1 of 2) PTHR31339:SF12 - POLYGALACTURONASE-LIKE PROTEIN
Eucgr.F00421	(1 of 2) K10395 - kinesin family member 4/21/27 (KIF4_21_27)
Eucgr.F02227	(1 of 19) PF14309 - Domain of unknown function (DUF4378) (DUF4378)
Eucgr.F04000	(1 of 2) PF00389//PF01842//PF07991 - D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain (2-Hacid_dh) // ACT domain (ACT) // Acetohydroxy acid isomeroeductase, NADPH-binding domain (IlvN...
Eucgr.G03199	(1 of 3) K09754 - coumaroylquininate(coumaroylshikimate) 3'-monooxygenase (CYP98A3, C3'H)
Eucgr.H02748	(1 of 2) K03164 - DNA topoisomerase II [EC:5.99.1.3] (TOP2)
Eucgr.H03531	(1 of 1) 2.5.1.79 - Thermospermine synthase
Eucgr.H04498	(1 of 7) K15718 - linoleate 9S-lipoxygenase (LOX1_5)
Eucgr.I02677	(1 of 2) K08917 - light-harvesting complex II chlorophyll a/b binding protein 6 (LHCB6)
Eucgr.J02238	(1 of 1) K11790 - denticleless (DTL, CDT2, DCAF2)
Eucgr.K03381	(1 of 4) PTHR13683:SF379 - ASPARTYL PROTEASE-LIKE PROTEIN
Eucgr.K03427	(1 of 1) KOG4270 - GTPase-activator protein
Eucgr.L00816	(1 of 4) PTHR22835//PTHR22835:SF189 - ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED
Eucgr.L01962	(1 of 57) PF00314 - Thaumatin family (Thaumatin)

Table S5.13 The 20 most upregulated genes in cluster 3 when comparing EF, LFB, and LFU

Genes in cluster 3	brief explanation
Eucgr.A00237	(1 of 1) PTHR14233//PTHR14233:SF16 - DUF914-RELATED // SUBFAMILY NOT NAMED
Eucgr.B00504	(1 of 1) PTHR11062//PTHR11062:SF99 - EXOSTOSIN HEPARAN SULFATE GLYCOSYLTRANSFERASE -RELATED // SUBFAMILY NOT NAMED
Eucgr.B03594	(1 of 1) KOG2049 - Translational repressor MPT5/PUF4 and related RNA-binding proteins (Puf superfamily)
Eucgr.B04000	(1 of 1) PTHR11668//PTHR11668:SF267 - SERINE/THREONINE PROTEIN PHOSPHATASE // SUBFAMILY NOT NAMED
Eucgr.C01762	(1 of 2) PF12854//PF13041//PF14432 - PPR repeat (PPR_1) // PPR repeat family (PPR_2) // DYW family of nucleic acid deaminases (DYW_deaminase)
Eucgr.C02986	(1 of 7) K00850 - 6-phosphofructokinase 1 (pfkA, PFK)
Eucgr.C03851	
Eucgr.E00626	(1 of 1) 6.3.5.2 - GMP synthase (glutamine-hydrolyzing) / GMP synthetase (glutamine-hydrolyzing)
Eucgr.E03667	(1 of 73) PF00954//PF01453//PF07714//PF08276 - S-locus glycoprotein domain (S_locus_glycop) // D-mannose binding lectin (B_lectin) // Protein tyrosine kinase (Pkinase_Tyr) // PAN-like domain (PAN_2)
Eucgr.F01480	(1 of 3) 6.3.4.2 - CTP synthase (glutamine hydrolyzing) / UTP-- ammonia ligase
Eucgr.F01642	(1 of 1) K03256 - tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit (TRM6, GCD10)
Eucgr.F03654	(1 of 7) PTHR22950:SF228 - AMINO ACID PERMEASE 7-RELATED
Eucgr.G02061	(1 of 1) PTHR30540//PTHR30540:SF5 - OSMOTIC STRESS POTASSIUM TRANSPORTER // SUBFAMILY NOT NAMED
Eucgr.H00216	(1 of 4) PTHR31734:SF11 - AUXIN-RESPONSIVE PROTEIN IAA1-RELATED
Eucgr.H00650	(1 of 4) 2.2.1.2 - Transaldolase / Glycerone transferase
Eucgr.H01306	(1 of 1) PTHR23429:SF4 - GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE 4, CHLOROPLASTIC
Eucgr.H04094	(1 of 2) PTHR11453:SF44 - BORON TRANSPORTER 1-RELATED
Eucgr.H04418	(1 of 1) K14289 - exportin-5 (XPO5)
Eucgr.J02452	(1 of 27) K08472 - mlo protein (MLO)
Eucgr.L00778	

Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0055114	P	oxidation-reduction process	127	1849	8.60E-09
GO:0044710	P	single-organism metabolic process	169	2807	1.40E-07
GO:0055085	P	transmembrane transport	71	889	1.60E-07
GO:0006810	P	transport	104	1552	7.20E-07
GO:0051234	P	establishment of localization	104	1559	8.80E-07
GO:0051179	P	localization	104	1566	1.10E-06
GO:0009765	P	photosynthesis, light harvesting	9	22	1.80E-06
GO:0015979	P	photosynthesis	16	89	3.00E-06
GO:0019684	P	photosynthesis, light reaction	10	34	5.70E-06
GO:0006091	P	generation of precursor metabolites and energy	14	90	5.10E-05
GO:0042592	P	homeostatic process	16	130	0.00019
GO:0019725	P	cellular homeostasis	15	124	0.00035
GO:0006873	P	cellular ion homeostasis	5	12	0.00037
GO:0050801	P	ion homeostasis	5	13	0.0005
GO:0044699	P	single-organism process	219	4462	0.0011
GO:0065008	P	regulation of biological quality	17	176	0.0015
GO:0005976	P	polysaccharide metabolic process	12	105	0.0021
GO:0030243	P	cellulose metabolic process	7	41	0.0024
GO:0030244	P	cellulose biosynthetic process	7	41	0.0024
GO:0055082	P	cellular chemical homeostasis	5	20	0.0025
GO:0006073	P	cellular glucan metabolic process	11	95	0.0028
GO:0044042	P	glucan metabolic process	11	95	0.0028
GO:0048878	P	chemical homeostasis	5	21	0.003
GO:0044264	P	cellular polysaccharide metabolic process	11	96	0.003
GO:2001141	P	regulation of RNA biosynthetic process	54	903	0.0042
GO:0006355	P	regulation of transcription, DNA-templated	54	903	0.0042
GO:1903506	P	regulation of nucleic acid-templated transcription	54	903	0.0042
GO:0051252	P	regulation of RNA metabolic process	54	905	0.0043
GO:0018904	P	ether metabolic process	6	35	0.0048
GO:0006662	P	glycerol ether metabolic process	6	35	0.0048
GO:0031323	P	regulation of cellular metabolic process	55	931	0.0048

Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis (continued).

GO:0010556	P	regulation of macromolecule biosynthetic process	54	911	0.0049
GO:2000112	P	regulation of cellular macromolecule biosynthetic process	54	911	0.0049
GO:0009889	P	regulation of biosynthetic process	54	913	0.0051
GO:0031326	P	regulation of cellular biosynthetic process	54	913	0.0051
GO:0019219	P	regulation of nucleobase-containing compound metabolic process	54	913	0.0051
GO:0044262	P	cellular carbohydrate metabolic process	14	151	0.0053
GO:0005975	P	carbohydrate metabolic process	42	673	0.0056
GO:0051171	P	regulation of nitrogen compound metabolic process	54	919	0.0058
GO:0010468	P	regulation of gene expression	54	924	0.0064
GO:0019222	P	regulation of metabolic process	55	946	0.0065
GO:0080090	P	regulation of primary metabolic process	54	927	0.0068
GO:0051186	P	cofactor metabolic process	13	143	0.0082
GO:0060255	P	regulation of macromolecule metabolic process	54	939	0.0085
GO:0045454	P	cell redox homeostasis	10	103	0.013
GO:0009250	P	glucan biosynthetic process	7	59	0.014
GO:0051273	P	beta-glucan metabolic process	7	59	0.014
GO:0051274	P	beta-glucan biosynthetic process	7	59	0.014
GO:0033692	P	cellular polysaccharide biosynthetic process	7	60	0.015
GO:0000271	P	polysaccharide biosynthetic process	7	60	0.015
GO:0097659	P	nucleic acid-templated transcription	56	1016	0.016
GO:0006351	P	transcription, DNA-templated	56	1016	0.016
GO:0032774	P	RNA biosynthetic process	56	1018	0.016
GO:0034654	P	nucleobase-containing compound biosynthetic process	59	1093	0.019
GO:0019438	P	aromatic compound biosynthetic process	61	1154	0.025
GO:0070726	P	cell wall assembly	3	14	0.027
GO:0030198	P	extracellular matrix organization	3	14	0.027

Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis (continued).

GO:0071668	P	plant-type cell wall assembly	3	14	0.027
GO:0009664	P	plant-type cell wall organization	3	14	0.027
GO:0010215	P	cellulose microfibril organization	3	14	0.027
GO:0043062	P	extracellular structure organization	3	14	0.027
GO:0006732	P	coenzyme metabolic process	10	117	0.027
GO:0008152	P	metabolic process	408	9388	0.029
GO:0006812	P	cation transport	23	363	0.03
GO:0006790	P	sulfur compound metabolic process	5	40	0.03
GO:0071669	P	plant-type cell wall organization or biogenesis	3	15	0.031
GO:0009832	P	plant-type cell wall biogenesis	3	15	0.031
GO:0006811	P	ion transport	28	469	0.033
GO:0003333	P	amino acid transmembrane transport	5	42	0.035
GO:1903825	P	organic acid transmembrane transport	5	42	0.035
GO:1905039	P	carboxylic acid transmembrane transport	5	42	3.50E-02
GO:0015849	P	organic acid transport	5	42	3.50E-02
GO:0046496	P	nicotinamide nucleotide metabolic process	6	57	3.60E-02
GO:0019362	P	pyridine nucleotide metabolic process	6	57	3.60E-02
GO:0006733	P	oxidoreduction coenzyme metabolic process	6	58	3.80E-02
GO:1901362	P	organic cyclic compound biosynthetic process	62	1207	3.90E-02
GO:0042546	P	cell wall biogenesis	3	17	4.10E-02
GO:0040007	P	growth	3	17	4.10E-02
GO:0016049	P	cell growth	3	17	0.041
GO:0072524	P	pyridine-containing compound metabolic process	6	61	0.046
GO:0044272	P	sulfur compound biosynthetic process	3	18	0.047
GO:0018130	P	heterocycle biosynthetic process	59	1159	0.049
GO:0016491	F	oxidoreductase activity	142	2023	2.20E-10
GO:0005215	F	transporter activity	78	1001	9.60E-08
GO:0046914	F	transition metal ion binding	101	1550	3.40E-06

Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis (continued).

GO:0046872	F	metal ion binding	125	2137	3.10E-05
GO:0043169	F	cation binding	125	2143	3.50E-05
GO:0043167	F	ion binding	128	2230	5.60E-05
GO:0022857	F	transmembrane transporter activity	56	783	6.40E-05
GO:0005506	F	iron ion binding	46	611	9.60E-05
GO:0008194	F	UDP-glycosyltransferase activity	16	135	0.00028
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	43	628	0.001
GO:0015035	F	protein disulfide oxidoreductase activity	11	83	0.0011
GO:0015036	F	disulfide oxidoreductase activity	11	83	0.0011
GO:0020037	F	heme binding	47	709	0.0011
GO:0046906	F	tetrapyrrole binding	47	712	0.0012
GO:0016759	F	cellulose synthase activity	7	37	0.0014
GO:0016760	F	cellulose synthase (UDP-forming) activity	7	37	0.0014
GO:0035251	F	UDP-glucosyltransferase activity	10	77	0.002
GO:0046527	F	glucosyltransferase activity	10	78	0.0022
GO:0016758	F	transferase activity, transferring hexosyl groups	41	629	0.003
GO:0016757	F	transferase activity, transferring glycosyl groups	45	727	0.0048
GO:0022891	F	substrate-specific transmembrane transporter activity	30	434	0.0049
GO:0016667	F	oxidoreductase activity, acting on a sulfur group of donors	12	120	0.0055
GO:0015018	F	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase activity	3	7	0.0058
GO:0004332	F	fructose-bisphosphate aldolase activity	3	7	0.0058
GO:0022892	F	substrate-specific transporter activity	32	478	0.0058
GO:0015020	F	glucuronosyltransferase activity	3	8	0.0077
GO:0005342	F	organic acid transmembrane transporter activity	5	29	9.70E-03
GO:0015171	F	amino acid transmembrane transporter activity	5	29	9.70E-03
GO:0046943	F	carboxylic acid transmembrane transporter activity	5	29	9.70E-03

Table S5.14 Gene ontology significant terms for cluster one in the seed capsule development cluster analysis (continued).

GO:0003700	F	transcription factor activity, sequence-specific DNA binding	30	458	9.70E-03
GO:0001071	F	nucleic acid binding transcription factor activity	30	458	0.0097
GO:0016832	F	aldehyde-lyase activity	3	9	0.01
GO:0008171	F	O-methyltransferase activity	11	115	0.01
GO:0005507	F	copper ion binding	12	137	0.014
GO:0045735	F	nutrient reservoir activity	5	33	0.015
GO:0005253	F	anion channel activity	4	23	0.02
GO:0008308	F	voltage-gated anion channel activity	4	23	0.02
GO:0022832	F	voltage-gated channel activity	4	23	0.02
GO:0005244	F	voltage-gated ion channel activity	4	23	0.02
GO:0008509	F	anion transmembrane transporter activity	6	49	0.02
GO:0015075	F	ion transmembrane transporter activity	25	390	0.021
GO:0005509	F	calcium ion binding	10	119	0.03
GO:0051087	F	chaperone binding	3	15	0.031
GO:0003824	F	catalytic activity	409	9441	0.035
GO:0043565	F	sequence-specific DNA binding	16	237	0.04
GO:0005544	F	calcium-dependent phospholipid binding	3	17	0.041
GO:0008270	F	zinc ion binding	42	780	0.045
GO:0016798	F	hydrolase activity, acting on glycosyl bonds	27	463	0.046
GO:0016638	F	oxidoreductase activity, acting on the CH-NH2 group of donors	3	18	0.047
GO:0016020	C	membrane	159	2184	8.50E-13
GO:0031224	C	intrinsic component of membrane	81	1053	8.90E-08
GO:0016021	C	integral component of membrane	78	1038	3.80E-07
GO:0044425	C	membrane part	88	1243	7.00E-07
GO:0009523	C	photosystem II	5	34	0.017
GO:0009521	C	photosystem	6	52	0.025
GO:0009654	C	photosystem II oxygen evolving complex	4	26	0.028
GO:0031225	C	anchored component of membrane	3	15	0.031
GO:0034357	C	photosynthetic membrane	6	57	0.036
GO:0009579	C	thylakoid	6	58	0.038
GO:0044436	C	thylakoid part	6	58	0.038

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0044710	P	single-organism metabolic process	174	2807	2.10E-17
GO:0044699	P	single-organism process	220	4462	9.30E-12
GO:0055114	P	oxidation-reduction process	107	1849	5.90E-09
GO:0044281	P	small molecule metabolic process	50	610	8.10E-09
GO:0006952	P	defense response	20	158	8.30E-07
GO:0044711	P	single-organism biosynthetic process	37	457	9.90E-07
GO:0006082	P	organic acid metabolic process	31	366	3.20E-06
GO:0006720	P	isoprenoid metabolic process	9	32	3.90E-06
GO:0008299	P	isoprenoid biosynthetic process	9	32	3.90E-06
GO:0009607	P	response to biotic stimulus	16	124	8.30E-06
GO:0019752	P	carboxylic acid metabolic process	28	328	8.40E-06
GO:0008152	P	metabolic process	357	9388	1.20E-05
GO:0043436	P	oxoacid metabolic process	28	337	1.30E-05
GO:0006767	P	water-soluble vitamin metabolic process	7	23	3.10E-05
GO:0006766	P	vitamin metabolic process	7	23	3.10E-05
GO:0042364	P	water-soluble vitamin biosynthetic process	7	23	3.10E-05
GO:0009110	P	vitamin biosynthetic process	7	23	3.10E-05
GO:0051186	P	cofactor metabolic process	16	143	4.10E-05
GO:1901564	P	organonitrogen compound metabolic process	50	849	6.30E-05
GO:0009058	P	biosynthetic process	102	2143	6.70E-05
GO:0044712	P	single-organism catabolic process	12	91	9.10E-05
GO:0005984	P	disaccharide metabolic process	8	39	9.50E-05
GO:0016114	P	terpenoid biosynthetic process	5	11	9.70E-05
GO:0006721	P	terpenoid metabolic process	5	11	9.70E-05
GO:0032787	P	monocarboxylic acid metabolic process	15	143	0.00014
GO:0008610	P	lipid biosynthetic process	17	179	0.00015
GO:0006629	P	lipid metabolic process	30	441	0.0002
GO:0006732	P	coenzyme metabolic process	13	117	0.00023

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis (continued).

GO:0044262	P	cellular carbohydrate metabolic process	15	151	0.00024
GO:0009311	P	oligosaccharide metabolic process	8	46	0.00026
GO:0044282	P	small molecule catabolic process	6	24	0.00029
GO:0044763	P	single-organism cellular process	101	2241	0.00049
GO:0005975	P	carbohydrate metabolic process	39	673	0.00054
GO:1901576	P	organic substance biosynthetic process	92	2025	0.00071
GO:0009063	P	cellular amino acid catabolic process	4	10	0.00074
GO:0016052	P	carbohydrate catabolic process	8	56	0.00084
GO:0044249	P	cellular biosynthetic process	90	1984	0.00086
GO:0006771	P	riboflavin metabolic process	4	11	0.00099
GO:0042727	P	flavin-containing compound biosynthetic process	4	11	0.00099
GO:0042726	P	flavin-containing compound metabolic process	4	11	0.00099
GO:0009231	P	riboflavin biosynthetic process	4	11	0.00099
GO:0005985	P	sucrose metabolic process	5	21	0.0011
GO:0044723	P	single-organism carbohydrate metabolic process	18	238	0.0012
GO:0044724	P	single-organism carbohydrate catabolic process	7	47	0.0014
GO:0006090	P	pyruvate metabolic process	7	47	0.0014
GO:0046395	P	carboxylic acid catabolic process	4	13	0.0016
GO:0044255	P	cellular lipid metabolic process	16	208	0.0019
GO:0009116	P	nucleoside metabolic process	12	132	0.0019
GO:1901657	P	glycosyl compound metabolic process	12	132	0.0019
GO:0006558	P	L-phenylalanine metabolic process	3	6	0.0022
GO:0016054	P	organic acid catabolic process	4	15	0.0025
GO:0065008	P	regulation of biological quality	14	176	0.0027
GO:0009072	P	aromatic amino acid family metabolic process	5	27	0.003
GO:0055086	P	nucleobase-containing small molecule metabolic process	14	181	0.0034
GO:0019725	P	cellular homeostasis	11	124	0.0035
GO:1901362	P	organic cyclic compound biosynthetic process	57	1207	0.0036

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis.

GO:0006520	P	cellular amino acid metabolic process	12	148	0.0046
GO:0042592	P	homeostatic process	11	130	0.0049
GO:0046351	P	disaccharide biosynthetic process	4	19	0.0053
GO:0044283	P	small molecule biosynthetic process	13	172	0.0056
GO:0051188	P	cofactor biosynthetic process	8	80	0.0063
GO:1901360	P	organic cyclic compound metabolic process	80	1866	0.0069
GO:0045454	P	cell redox homeostasis	9	103	0.0086
GO:0033013	P	tetrapyrrole metabolic process	4	23	0.0094
GO:0009312	P	oligosaccharide biosynthetic process	4	23	0.0094
GO:0009064	P	glutamine family amino acid metabolic process	4	25	0.012
GO:0006413	P	translational initiation	4	25	0.012
GO:0006950	P	response to stress	30	590	0.013
GO:0018130	P	heterocycle biosynthetic process	52	1159	0.013
GO:0006855	P	drug transmembrane transport	8	93	0.014
GO:0015893	P	drug transport	8	93	0.014
GO:0042493	P	response to drug	8	93	0.014
GO:0046031	P	ADP metabolic process	5	41	0.014
GO:0009179	P	purine ribonucleoside diphosphate metabolic process	5	41	0.014
GO:0006757	P	ATP generation from ADP	5	41	0.014
GO:0009185	P	ribonucleoside diphosphate metabolic process	5	41	0.014
GO:0009135	P	purine nucleoside diphosphate metabolic process	5	41	0.014
GO:0006096	P	glycolytic process	5	41	0.014
GO:0009108	P	coenzyme biosynthetic process	6	58	0.015
GO:0006807	P	nitrogen compound metabolic process	96	2383	0.016
GO:1901575	P	organic substance catabolic process	14	224	0.018
GO:0072524	P	pyridine-containing compound metabolic process	6	61	0.018
GO:0042221	P	response to chemical	14	226	0.019
GO:0006163	P	purine nucleotide metabolic process	9	120	0.02

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis.

GO:0042440	P	pigment metabolic process	3	16	0.02
GO:0006165	P	nucleoside diphosphate phosphorylation	5	46	0.021
GO:0046939	P	nucleotide phosphorylation	5	46	0.021
GO:0072521	P	purine-containing compound metabolic process	9	122	0.022
GO:0009605	P	response to external stimulus	4	31	0.023
GO:0005992	P	trehalose biosynthetic process	3	17	0.023
GO:0009132	P	nucleoside diphosphate metabolic process	5	48	0.025
GO:1901605	P	alpha-amino acid metabolic process	6	66	0.025
GO:0046128	P	purine ribonucleoside metabolic process	8	105	0.026
GO:0042278	P	purine nucleoside metabolic process	8	105	0.026
GO:0033014	P	tetrapyrrole biosynthetic process	3	18	2.70E-02
GO:0005991	P	trehalose metabolic process	3	18	2.70E-02
GO:0046483	P	heterocycle metabolic process	74	1823	2.80E-02
GO:0009117	P	nucleotide metabolic process	10	149	0.028
GO:1901565	P	organonitrogen compound catabolic process	5	50	0.029
GO:0009119	P	ribonucleoside metabolic process	8	108	0.029
GO:0006725	P	cellular aromatic compound metabolic process	74	1829	0.03
GO:0009628	P	response to abiotic stimulus	3	19	0.03
GO:0006753	P	nucleoside phosphate metabolic process	10	151	0.03
GO:0009150	P	purine ribonucleotide metabolic process	8	109	0.031
GO:0009259	P	ribonucleotide metabolic process	8	109	0.031
GO:0044271	P	cellular nitrogen compound biosynthetic process	64	1559	0.033
GO:0009056	P	catabolic process	14	244	0.033
GO:0019438	P	aromatic compound biosynthetic process	49	1154	0.036
GO:0009308	P	amine metabolic process	4	37	0.039
GO:1901135	P	carbohydrate derivative metabolic process	15	274	0.039
GO:0019693	P	ribose phosphate metabolic process	8	116	0.041

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis (continued).

GO:0046496	P	nicotinamide nucleotide metabolic process	5	57	0.045
GO:0019362	P	pyridine nucleotide metabolic process	5	57	0.045
GO:0006733	P	oxidoreduction coenzyme metabolic process	5	58	0.047
GO:0016491	F	oxidoreductase activity	121	2023	6.30E-11
GO:0048037	F	cofactor binding	36	469	4.30E-06
GO:0019842	F	vitamin binding	7	27	7.40E-05
GO:0043167	F	ion binding	104	2230	0.00012
GO:0050662	F	coenzyme binding	26	361	0.00024
GO:0030976	F	thiamine pyrophosphate binding	6	23	0.00024
GO:0003824	F	catalytic activity	349	9441	0.00028
GO:0005506	F	iron ion binding	36	611	0.00066
GO:0046872	F	metal ion binding	96	2137	0.00077
GO:0043169	F	cation binding	96	2143	0.00084
GO:0015291	F	secondary active transmembrane transporter activity	14	154	0.00085
GO:0022804	F	active transmembrane transporter activity	21	291	0.00089
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	13	140	0.0011
GO:0016616	F	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	12	123	0.0011
GO:0016701	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	7	45	0.0011
GO:0000287	F	magnesium ion binding	14	166	0.0016
GO:0016758	F	transferase activity, transferring hexosyl groups	35	629	0.002
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	34	628	0.0034
GO:0016829	F	lyase activity	16	229	0.0047
GO:0016157	F	sucrose synthase activity	4	19	0.0053
GO:0015035	F	protein disulfide oxidoreductase activity	8	83	0.0077

Table S5.15 Gene ontology significant terms for cluster two in the seed capsule development cluster analysis (continued).

GO:0015036	F	disulfide oxidoreductase activity	8	83	0.0077
GO:0030955	F	potassium ion binding	3	11	0.0086
GO:0031420	F	alkali metal ion binding	3	11	0.0086
GO:0004743	F	pyruvate kinase activity	3	11	0.0086
GO:0071949	F	FAD binding	6	51	0.0086
GO:0016702	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	5	36	0.0088
GO:0019205	F	nucleobase-containing compound kinase activity	4	23	0.0094
GO:0016757	F	transferase activity, transferring glycosyl groups	36	727	0.0098
GO:0043168	F	anion binding	8	89	0.011
GO:0020037	F	heme binding	35	709	0.011
GO:0015297	F	antiporter activity	10	128	0.012
GO:0051213	F	dioxygenase activity	5	39	0.012
GO:0046906	F	tetrapyrrole binding	35	712	0.012
GO:0090484	F	drug transporter activity	8	93	0.014
GO:0015238	F	drug transmembrane transporter activity	8	93	0.014
GO:0030170	F	pyridoxal phosphate binding	6	65	0.024
GO:0003743	F	translation initiation factor activity	4	36	0.036
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	4	36	0.036
GO:0046983	F	protein dimerization activity	17	320	0.037
GO:0008171	F	O-methyltransferase activity	8	115	0.04
GO:0050661	F	NADP binding	4	38	0.042
GO:0016830	F	carbon-carbon lyase activity	5	57	0.045
GO:0015103	F	inorganic anion transmembrane transporter activity	4	40	0.048
GO:0016667	F	oxidoreductase activity, acting on a sulfur group of donors	8	120	0.048
GO:0048046	C	apoplast	4	36	0.036
GO:0016021	C	integral component of membrane	44	1038	0.046

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0007059	P	chromosome segregation	17	31	1.30E-11
GO:0007049	P	cell cycle	26	99	5.30E-11
GO:0000278	P	mitotic cell cycle	18	41	5.90E-11
GO:0022402	P	cell cycle process	22	71	1.20E-10
GO:0055114	P	oxidation-reduction process	149	1849	4.00E-10
GO:1903047	P	mitotic cell cycle process	16	39	1.50E-09
GO:0000280	P	nuclear division	15	33	1.60E-09
GO:1901136	P	carbohydrate derivative catabolic process	15	36	4.20E-09
GO:0007067	P	mitotic nuclear division	14	31	6.10E-09
GO:0048285	P	organelle fission	15	38	7.50E-09
GO:0098813	P	nuclear chromosome segregation	13	27	1.20E-08
GO:0044699	P	single-organism process	284	4462	4.10E-08
GO:0046348	P	amino sugar catabolic process	13	31	4.30E-08
GO:1901072	P	glucosamine-containing compound catabolic process	13	31	4.30E-08
GO:1901071	P	glucosamine-containing compound metabolic process	13	31	4.30E-08
GO:0006026	P	aminoglycan catabolic process	13	31	4.30E-08
GO:0006030	P	chitin metabolic process	13	31	4.30E-08
GO:0006032	P	chitin catabolic process	13	31	4.30E-08
GO:0006040	P	amino sugar metabolic process	13	31	4.30E-08
GO:0044710	P	single-organism metabolic process	195	2807	4.50E-08
GO:0000819	P	sister chromatid segregation	12	26	6.20E-08
GO:0000070	P	mitotic sister chromatid segregation	12	26	6.20E-08
GO:0006022	P	aminoglycan metabolic process	13	35	1.30E-07
GO:0006260	P	DNA replication	14	50	7.80E-07
GO:0016998	P	cell wall macromolecule catabolic process	13	43	9.40E-07
GO:0044036	P	cell wall macromolecule metabolic process	13	44	1.20E-06
GO:0005975	P	carbohydrate metabolic process	62	673	1.80E-06
GO:1901565	P	organonitrogen compound catabolic process	13	50	3.90E-06
GO:0008152	P	metabolic process	513	9388	4.40E-06

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0007076	P	mitotic chromosome condensation	7	11	7.80E-06
GO:0030261	P	chromosome condensation	7	11	7.80E-06
GO:0071554	P	cell wall organization or biogenesis	17	94	9.90E-06
GO:0007017	P	microtubule-based process	18	107	1.30E-05
GO:0051726	P	regulation of cell cycle	11	41	1.70E-05
GO:1902589	P	single-organism organelle organization	17	102	2.50E-05
GO:0010564	P	regulation of cell cycle process	8	23	5.30E-05
GO:0051276	P	chromosome organization	15	96	0.00014
GO:0006261	P	DNA-dependent DNA replication	5	8	0.00018
GO:0045930	P	negative regulation of mitotic cell cycle	5	8	0.00018
GO:0007093	P	mitotic cell cycle checkpoint	5	8	0.00018
GO:0007346	P	regulation of mitotic cell cycle	7	22	0.00025
GO:0051304	P	chromosome separation	6	15	0.00025
GO:0071103	P	DNA conformation change	9	39	0.00026
GO:0006928	P	movement of cell or subcellular component	11	59	0.00029
GO:0007018	P	microtubule-based movement	11	59	0.00029
GO:0006508	P	proteolysis	51	634	0.00034
GO:0000075	P	cell cycle checkpoint	6	18	0.00056
GO:0006323	P	DNA packaging	7	27	0.0007
GO:0007094	P	mitotic spindle assembly checkpoint	4	6	0.0007
GO:0033046	P	negative regulation of sister chromatid segregation	4	6	0.0007
GO:0033048	P	negative regulation of mitotic sister chromatid segregation	4	6	0.0007
GO:0051784	P	negative regulation of nuclear division	4	6	0.0007
GO:1902100	P	negative regulation of metaphase/anaphase transition of cell cycle	4	6	0.0007
GO:0045841	P	negative regulation of mitotic metaphase/anaphase transition	4	6	0.0007
GO:2000816	P	negative regulation of mitotic sister chromatid separation	4	6	0.0007
GO:0071173	P	spindle assembly checkpoint	4	6	0.0007

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0071174	P	mitotic spindle checkpoint	4	6	0.0007
GO:0051985	P	negative regulation of chromosome segregation	4	6	0.0007
GO:0045839	P	negative regulation of mitotic nuclear division	4	6	0.0007
GO:0031577	P	spindle checkpoint	4	6	0.0007
GO:2001251	P	negative regulation of chromosome organization	4	6	0.0007
GO:0051301	P	cell division	5	12	0.00073
GO:1901988	P	negative regulation of cell cycle phase transition	4	7	0.0011
GO:1901991	P	negative regulation of mitotic cell cycle phase transition	4	7	0.0011
GO:0010948	P	negative regulation of cell cycle process	4	7	0.0011
GO:0010965	P	regulation of mitotic sister chromatid separation	5	14	0.0013
GO:0045786	P	negative regulation of cell cycle	5	14	0.0013
GO:1902099	P	regulation of metaphase/anaphase transition of cell cycle	5	14	0.0013
GO:0051783	P	regulation of nuclear division	5	14	0.0013
GO:0044784	P	metaphase/anaphase transition of cell cycle	5	14	0.0013
GO:0033044	P	regulation of chromosome organization	5	14	0.0013
GO:0033045	P	regulation of sister chromatid segregation	5	14	0.0013
GO:0033047	P	regulation of mitotic sister chromatid segregation	5	14	0.0013
GO:0051983	P	regulation of chromosome segregation	5	14	0.0013
GO:0007088	P	regulation of mitotic nuclear division	5	14	0.0013
GO:0030071	P	regulation of mitotic metaphase/anaphase transition	5	14	0.0013
GO:0051306	P	mitotic sister chromatid separation	5	14	0.0013
GO:0007091	P	metaphase/anaphase transition of mitotic cell cycle	5	14	0.0013
GO:0048523	P	negative regulation of cellular process	6	22	0.0013

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0010639	P	negative regulation of organelle organization	4	8	0.0015
GO:0051129	P	negative regulation of cellular component organization	4	8	0.0015
GO:0044262	P	cellular carbohydrate metabolic process	17	151	0.0016
GO:1901990	P	regulation of mitotic cell cycle phase transition	5	15	0.0016
GO:1901987	P	regulation of cell cycle phase transition	5	15	0.0016
GO:0044770	P	cell cycle phase transition	5	15	0.0016
GO:0044772	P	mitotic cell cycle phase transition	5	15	0.0016
GO:0006996	P	organelle organization	23	240	0.002
GO:0015979	P	photosynthesis	12	89	0.002
GO:0006270	P	DNA replication initiation	3	4	0.0028
GO:0000910	P	cytokinesis	4	10	0.0029
GO:0009765	P	photosynthesis, light harvesting	5	22	0.0066
GO:0009057	P	macromolecule catabolic process	14	134	0.0073
GO:1901575	P	organic substance catabolic process	20	224	0.0077
GO:0006073	P	cellular glucan metabolic process	11	95	0.0085
GO:0044042	P	glucan metabolic process	11	95	0.0085
GO:0048519	P	negative regulation of biological process	6	34	0.0087
GO:0016043	P	cellular component organization	29	372	0.009
GO:0000226	P	microtubule cytoskeleton organization	5	24	0.009
GO:0044264	P	cellular polysaccharide metabolic process	11	96	0.0091
GO:0044248	P	cellular catabolic process	16	168	0.0095
GO:0033043	P	regulation of organelle organization	6	38	1.40E-02
GO:0006631	P	fatty acid metabolic process	10	89	1.40E-02
GO:0006633	P	fatty acid biosynthetic process	8	63	1.50E-02
GO:0050832	P	defense response to fungus	3	9	1.50E-02
GO:0009620	P	response to fungus	3	9	1.50E-02
GO:0005976	P	polysaccharide metabolic process	11	105	1.60E-02
GO:0009056	P	catabolic process	20	244	1.70E-02

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0034637	P	cellular carbohydrate biosynthetic process	9	80	1.90E-02
GO:0006979	P	response to oxidative stress	13	138	2.00E-02
GO:0072330	P	monocarboxylic acid biosynthetic process	8	67	2.00E-02
GO:0006259	P	DNA metabolic process	18	216	2.00E-02
GO:0051128	P	regulation of cellular component organization	6	43	2.20E-02
GO:0043207	P	response to external biotic stimulus	3	11	2.30E-02
GO:0051707	P	response to other organism	3	11	2.30E-02
GO:0009617	P	response to bacterium	3	11	2.30E-02
GO:0042742	P	defense response to bacterium	3	11	2.30E-02
GO:0098542	P	defense response to other organism	3	11	0.023
GO:0044723	P	single-organism carbohydrate metabolic process	19	238	0.025
GO:0071840	P	cellular component organization or biogenesis	29	412	0.029
GO:0019684	P	photosynthesis, light reaction	5	34	0.03
GO:0008610	P	lipid biosynthetic process	15	179	0.031
GO:0006013	P	mannose metabolic process	3	13	0.033
GO:0016051	P	carbohydrate biosynthetic process	9	90	0.035
GO:0006855	P	drug transmembrane transport	9	93	0.041
GO:0015893	P	drug transport	9	93	0.041
GO:0042493	P	response to drug	9	93	0.041
GO:1901135	P	carbohydrate derivative metabolic process	20	274	0.046
GO:0032787	P	monocarboxylic acid metabolic process	12	143	0.049
GO:0003824	F	catalytic activity	572	9441	1.30E-16
GO:0016787	F	hydrolase activity	191	2363	4.70E-13
GO:0004185	F	serine-type carboxypeptidase activity	29	105	1.40E-12
GO:0004180	F	carboxypeptidase activity	29	107	2.10E-12
GO:0070008	F	serine-type exopeptidase activity	29	110	3.80E-12
GO:0005507	F	copper ion binding	31	137	2.00E-11
GO:0008238	F	exopeptidase activity	29	128	8.80E-11

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0004553	F	hydrolase activity, hydrolyzing O-glycosyl compounds	57	443	1.30E-10
GO:0016798	F	hydrolase activity, acting on glycosyl bonds	58	463	2.30E-10
GO:0016491	F	oxidoreductase activity	158	2023	8.90E-10
GO:0004568	F	chitinase activity	13	31	4.30E-08
GO:0008236	F	serine-type peptidase activity	34	244	1.40E-07
GO:0017171	F	serine hydrolase activity	34	244	1.40E-07
GO:0008061	F	chitin binding	8	12	1.30E-06
GO:0008017	F	microtubule binding	15	69	4.80E-06
GO:0015631	F	tubulin binding	15	70	5.60E-06
GO:0070011	F	peptidase activity, acting on L-amino acid peptides	49	574	0.00012
GO:0008233	F	peptidase activity	50	596	0.00016
GO:0008092	F	cytoskeletal protein binding	15	97	0.00016
GO:0003777	F	microtubule motor activity	11	59	0.00029
GO:0004601	F	peroxidase activity	17	131	0.00039
GO:0016684	F	oxidoreductase activity, acting on peroxide as acceptor	17	133	0.00045
GO:0050664	F	oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor	4	7	0.0011
GO:0003774	F	motor activity	11	71	0.0011
GO:0020037	F	heme binding	53	709	1.30E-03
GO:0016209	F	antioxidant activity	17	148	0.0013
GO:0046906	F	tetrapyrrole binding	53	712	0.0014
GO:0048037	F	cofactor binding	37	469	0.0029
GO:0003993	F	acid phosphatase activity	5	18	0.0032
GO:0016616	F	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	13	123	0.0087
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	44	628	0.0091
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	14	140	0.01
GO:0046914	F	transition metal ion binding	94	1550	0.01
GO:0005506	F	iron ion binding	42	611	0.014

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0016788	F	hydrolase activity, acting on ester bonds	35	498	0.018
GO:0050662	F	coenzyme binding	27	361	0.018
GO:0043565	F	sequence-specific DNA binding	19	237	0.024
GO:0016757	F	transferase activity, transferring glycosyl groups	47	727	0.025
GO:0015297	F	antiporter activity	12	128	0.025
GO:0016747	F	transferase activity, transferring acyl groups other than amino-acyl groups	23	306	0.026
GO:0017111	F	nucleoside-triphosphatase activity	33	481	0.028
GO:0016844	F	strictosidine synthase activity	3	12	0.028
GO:0016843	F	amine-lyase activity	3	12	0.028
GO:0016462	F	pyrophosphatase activity	34	505	0.032
GO:0016746	F	transferase activity, transferring acyl groups	26	365	0.033
GO:0004190	F	aspartic-type endopeptidase activity	11	119	0.034
GO:0070001	F	aspartic-type peptidase activity	11	119	0.034
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	5	36	0.036
GO:0016818	F	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	34	512	0.038
GO:0016840	F	carbon-nitrogen lyase activity	3	14	0.039
GO:0016759	F	cellulose synthase activity	5	37	0.04
GO:0016760	F	cellulose synthase (UDP-forming) activity	5	37	0.04
GO:0015291	F	secondary active transmembrane transporter activity	13	154	0.04
GO:0090484	F	drug transporter activity	9	93	0.041
GO:0015238	F	drug transmembrane transporter activity	9	93	0.041
GO:0016651	F	oxidoreductase activity, acting on NAD(P)H	5	38	0.044
GO:0005694	C	chromosome	14	80	8.10E-05
GO:0044427	C	chromosomal part	12	66	0.00019
GO:0015630	C	microtubule cytoskeleton	15	100	0.00021
GO:0005871	C	kinesin complex	11	58	0.00025
GO:0005875	C	microtubule associated complex	11	73	0.0014

Table S5.16 Gene ontology significant terms for cluster three in the seed capsule development cluster analysis (continued).

GO:0000775	C	chromosome, centromeric region	4	8	0.0015
GO:0000796	C	condensin complex	3	3	0.0017
GO:0098687	C	chromosomal region	4	9	0.0021
GO:0000793	C	condensed chromosome	5	18	0.0032
GO:0044430	C	cytoskeletal part	15	135	0.0033
GO:0044815	C	DNA packaging complex	6	29	0.0044
GO:0005856	C	cytoskeleton	15	146	0.0064
GO:1990204	C	oxidoreductase complex	6	34	0.0087
GO:0030312	C	external encapsulating structure	9	71	0.01
GO:0005618	C	cell wall	9	71	0.01
GO:0005874	C	microtubule	4	19	0.019
GO:0043234	C	protein complex	41	616	0.024
GO:0099512	C	supramolecular fiber	4	22	0.028
GO:0099513	C	polymeric cytoskeletal fiber	4	22	0.028
GO:0048046	C	apoplast	5	36	0.036
GO:0071944	C	cell periphery	12	138	0.04
GO:0009654	C	photosystem II oxygen evolving complex	4	26	0.045
GO:0009521	C	photosystem	6	52	0.046
GO:0043232	C	intracellular non-membrane-bounded organelle	33	506	0.049
GO:0043228	C	non-membrane-bounded organelle	33	506	0.049

Table S5.17 Gene ontology significant terms for cluster four in the seed capsule development cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0006928	P	movement of cell or subcellular component	24	59	1.10E-10
GO:0007018	P	microtubule-based movement	24	59	1.10E-10
GO:0007017	P	microtubule-based process	29	107	3.60E-09
GO:0006073	P	cellular glucan metabolic process	18	95	0.0002
GO:0044042	P	glucan metabolic process	18	95	0.0002
GO:0044264	P	cellular polysaccharide metabolic process	18	96	0.00022
GO:0005976	P	polysaccharide metabolic process	18	105	0.00057
GO:0006260	P	DNA replication	11	50	0.0012
GO:0023052	P	signaling	64	660	0.0026
GO:0044700	P	single organism signaling	64	660	0.0026
GO:0007165	P	signal transduction	64	660	0.0026
GO:0030243	P	cellulose metabolic process	9	41	0.0033
GO:0030244	P	cellulose biosynthetic process	9	41	0.0033
GO:0009250	P	glucan biosynthetic process	11	59	0.0036
GO:0051273	P	beta-glucan metabolic process	11	59	0.0036
GO:0051274	P	beta-glucan biosynthetic process	11	59	0.0036
GO:0007154	P	cell communication	83	916	0.0037
GO:0033692	P	cellular polysaccharide biosynthetic process	11	60	0.0041
GO:0000271	P	polysaccharide biosynthetic process	11	60	0.0041
GO:0044262	P	cellular carbohydrate metabolic process	20	151	0.0046
GO:0034637	P	cellular carbohydrate biosynthetic process	13	80	0.0047
GO:0005975	P	carbohydrate metabolic process	63	673	0.0057
GO:0032502	P	developmental process	10	56	0.0071
GO:0016051	P	carbohydrate biosynthetic process	13	90	0.011
GO:0006468	P	protein phosphorylation	181	2322	0.013
GO:0051716	P	cellular response to stimulus	71	819	0.017
GO:0044707	P	single-multicellular organism process	8	49	0.023
GO:0007275	P	multicellular organism development	8	49	0.023

Table S5.17 Gene ontology significant terms for cluster four in the seed capsule development cluster analysis (continued).

GO:0044767	P	single-organism developmental process	8	49	0.023
GO:0048856	P	anatomical structure development	8	49	0.023
GO:0006261	P	DNA-dependent DNA replication	3	8	0.027
GO:0006813	P	potassium ion transport	6	33	0.031
GO:0016310	P	phosphorylation	182	2402	0.032
GO:0050794	P	regulation of cellular process	135	1741	0.034
GO:0050789	P	regulation of biological process	136	1766	4.00E-02
GO:0000910	P	cytokinesis	3	10	4.20E-02
GO:0032559	F	adenyl ribonucleotide binding	286	2874	2.90E-12
GO:0030554	F	adenyl nucleotide binding	286	2878	3.40E-12
GO:0003774	F	motor activity	28	71	5.60E-12
GO:0043531	F	ADP binding	96	683	6.50E-11
GO:0003777	F	microtubule motor activity	24	59	1.10E-10
GO:0001882	F	nucleoside binding	297	3120	1.10E-10
GO:0032555	F	purine ribonucleotide binding	296	3110	1.20E-10
GO:0032550	F	purine ribonucleoside binding	296	3110	1.20E-10
GO:0001883	F	purine nucleoside binding	296	3110	1.20E-10
GO:0032549	F	ribonucleoside binding	296	3119	1.60E-10
GO:0017076	F	purine nucleotide binding	296	3124	1.90E-10
GO:0032553	F	ribonucleotide binding	296	3136	2.80E-10
GO:0008017	F	microtubule binding	25	69	3.20E-10
GO:0015631	F	tubulin binding	25	70	4.10E-10
GO:0097367	F	carbohydrate derivative binding	297	3163	4.40E-10
GO:0000166	F	nucleotide binding	319	3460	5.20E-10
GO:1901265	F	nucleoside phosphate binding	319	3460	5.20E-10
GO:0036094	F	small molecule binding	320	3495	9.90E-10
GO:0008092	F	cytoskeletal protein binding	27	97	7.70E-09
GO:0005515	F	protein binding	318	3740	1.30E-06
GO:0016787	F	hydrolase activity	210	2363	1.10E-05
GO:0016462	F	pyrophosphatase activity	61	505	1.70E-05
GO:0016818	F	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	61	512	2.50E-05
GO:0017111	F	nucleoside-triphosphatase activity	58	481	2.90E-05

Table S5.17 Gene ontology significant terms for cluster four in the seed capsule development cluster analysis (continued).

GO:0016817	F	hydrolase activity, acting on acid anhydrides	62	528	3.20E-05
GO:0004252	F	serine-type endopeptidase activity	20	103	6.60E-05
GO:0005524	F	ATP binding	190	2191	0.00011
GO:0008236	F	serine-type peptidase activity	34	244	0.00012
GO:0017171	F	serine hydrolase activity	34	244	0.00012
GO:0003887	F	DNA-directed DNA polymerase activity	6	10	0.00026
GO:0005092	F	GDP-dissociation inhibitor activity	5	6	0.0003
GO:0005094	F	Rho GDP-dissociation inhibitor activity	5	6	0.0003
GO:0016798	F	hydrolase activity, acting on glycosyl bonds	52	463	0.00036
GO:0005488	F	binding	718	9943	0.00037
GO:0034061	F	DNA polymerase activity	6	11	0.00038
GO:0004553	F	hydrolase activity, hydrolyzing O-glycosyl compounds	49	443	0.00071
GO:0035639	F	purine ribonucleoside triphosphate binding	200	2427	0.0009
GO:1901363	F	heterocyclic compound binding	428	5745	0.0022
GO:0097159	F	organic cyclic compound binding	428	5745	0.0022
GO:0016759	F	cellulose synthase activity	8	37	0.0059
GO:0016760	F	cellulose synthase (UDP-forming) activity	8	37	0.0059
GO:0004175	F	endopeptidase activity	31	290	0.0095
GO:0005544	F	calcium-dependent phospholipid binding	5	17	0.0098
GO:0004672	F	protein kinase activity	183	2344	0.012
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	7	36	0.016
GO:0030695	F	GTPase regulator activity	5	20	0.017
GO:0008289	F	lipid binding	11	77	0.02
GO:0005507	F	copper ion binding	16	137	0.028
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	16	140	0.032
GO:0005543	F	phospholipid binding	7	43	0.033
GO:0016773	F	phosphotransferase activity, alcohol group as acceptor	185	2459	0.037

Table S5.17 Gene ontology significant terms for cluster four in the seed capsule development cluster analysis (continued).

GO:0016772	F	transferase activity, transferring phosphorus-containing groups	198	2651	0.04
GO:0060589	F	nucleoside-triphosphatase regulator activity	5	26	0.04
GO:0042623	F	ATPase activity, coupled	17	157	0.042
GO:0035251	F	UDP-glucosyltransferase activity	10	77	4.20E-02
GO:0046527	F	glucosyltransferase activity	10	78	4.50E-02
GO:0005871	C	kinesin complex	23	58	4.00E-10
GO:0005875	C	microtubule associated complex	23	73	1.50E-08
GO:0044430	C	cytoskeletal part	31	135	2.90E-08
GO:0015630	C	microtubule cytoskeleton	26	100	4.70E-08
GO:0005856	C	cytoskeleton	31	146	1.40E-07
GO:0016459	C	myosin complex	4	12	0.015
GO:0048046	C	apoplast	7	36	0.016
GO:0005576	C	extracellular region	9	55	0.017

Table S5.18 Gene ontology significant terms for cluster one in the flowers-only cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0044710	P	single-organism metabolic process	83	2807	1.20E-09
GO:0006720	P	isoprenoid metabolic process	9	32	7.60E-09
GO:0008299	P	isoprenoid biosynthetic process	9	32	7.60E-09
GO:0006629	P	lipid metabolic process	22	441	2.00E-06
GO:0016114	P	terpenoid biosynthetic process	5	11	2.80E-06
GO:0006721	P	terpenoid metabolic process	5	11	2.80E-06
GO:0008610	P	lipid biosynthetic process	13	179	6.70E-06
GO:0044255	P	cellular lipid metabolic process	13	208	3.00E-05
GO:0044699	P	single-organism process	98	4462	4.00E-05
GO:0055114	P	oxidation-reduction process	50	1849	5.40E-05
GO:0044711	P	single-organism biosynthetic process	19	457	0.00011
GO:0044281	P	small molecule metabolic process	22	610	0.00023
GO:0006082	P	organic acid metabolic process	15	366	0.00066
GO:0009072	P	aromatic amino acid family metabolic process	4	27	0.0011
GO:0019752	P	carboxylic acid metabolic process	13	328	0.002
GO:0043436	P	oxoacid metabolic process	13	337	0.0025
GO:0009058	P	biosynthetic process	49	2143	0.0026
GO:0042440	P	pigment metabolic process	3	16	0.0027
GO:1901362	P	organic cyclic compound biosynthetic process	31	1207	0.0034
GO:1901576	P	organic substance biosynthetic process	46	2025	0.004
GO:1901360	P	organic cyclic compound metabolic process	43	1866	0.0043
GO:0044249	P	cellular biosynthetic process	45	1984	0.0046
GO:0008152	P	metabolic process	165	9388	0.005
GO:0019438	P	aromatic compound biosynthetic process	29	1154	0.0062
GO:2001141	P	regulation of RNA biosynthetic process	24	903	0.0065
GO:0006355	P	regulation of transcription, DNA-templated	24	903	0.0065
GO:1903506	P	regulation of nucleic acid-templated transcription	24	903	0.0065

Table S5.18 Gene ontology significant terms for cluster one in the flowers-only cluster analysis (continued).

GO:0051252	P	regulation of RNA metabolic process	24	905	0.0067
GO:0010556	P	regulation of macromolecule biosynthetic process	24	911	0.0072
GO:2000112	P	regulation of cellular macromolecule biosynthetic process	24	911	0.0072
GO:0009889	P	regulation of biosynthetic process	24	913	0.0074
GO:0031326	P	regulation of cellular biosynthetic process	24	913	0.0074
GO:0019219	P	regulation of nucleobase-containing compound metabolic process	24	913	0.0074
GO:0044282	P	small molecule catabolic process	3	24	0.0074
GO:0032787	P	monocarboxylic acid metabolic process	7	143	0.0075
GO:0051171	P	regulation of nitrogen compound metabolic process	24	919	0.008
GO:0006725	P	cellular aromatic compound metabolic process	41	1829	0.0083
GO:0010468	P	regulation of gene expression	24	924	0.0085
GO:0080090	P	regulation of primary metabolic process	24	927	0.0088
GO:0031323	P	regulation of cellular metabolic process	24	931	0.0093
GO:0060255	P	regulation of macromolecule metabolic process	24	939	0.01
GO:0019222	P	regulation of metabolic process	24	946	0.011
GO:0018130	P	heterocycle biosynthetic process	28	1159	0.012
GO:0006952	P	defense response	7	158	0.012
GO:0009607	P	response to biotic stimulus	6	124	0.014
GO:0097659	P	nucleic acid-templated transcription	25	1016	0.014
GO:0006351	P	transcription, DNA-templated	25	1016	0.014
GO:0032774	P	RNA biosynthetic process	25	1018	0.014
GO:0044712	P	single-organism catabolic process	5	91	0.015
GO:0034654	P	nucleobase-containing compound biosynthetic process	26	1093	0.017
GO:0045454	P	cell redox homeostasis	5	103	0.023

Table S5.18 Gene ontology significant terms for cluster one in the flowers-only cluster analysis (continued).

GO:0046483	P	heterocycle metabolic process	38	1823	3.00E-02
GO:0005975	P	carbohydrate metabolic process	17	673	3.10E-02
GO:0006090	P	pyruvate metabolic process	3	47	3.90E-02
GO:0019725	P	cellular homeostasis	5	124	0.045
GO:1901565	P	organonitrogen compound catabolic process	3	50	0.045
GO:0016491	F	oxidoreductase activity	58	2023	2.00E-06
GO:0030976	F	thiamine pyrophosphate binding	5	23	5.40E-05
GO:0000287	F	magnesium ion binding	11	166	7.50E-05
GO:0019842	F	vitamin binding	5	27	0.0001
GO:0043167	F	ion binding	56	2230	0.00013
GO:0046872	F	metal ion binding	52	2137	0.00049
GO:0043169	F	cation binding	52	2143	0.00052
GO:0019205	F	nucleobase-containing compound kinase activity	4	23	0.00066
GO:0016702	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	4	36	0.0029
GO:0051213	F	dioxygenase activity	4	39	0.0038
GO:0046983	F	protein dimerization activity	12	320	0.0044
GO:0016701	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	4	45	0.0061
GO:0001071	F	nucleic acid binding transcription factor activity	14	458	0.012
GO:0003700	F	transcription factor activity, sequence-specific DNA binding	14	458	0.012
GO:0016835	F	carbon-oxygen lyase activity	6	132	0.018
GO:0010333	F	terpene synthase activity	5	99	0.02
GO:0016838	F	carbon-oxygen lyase activity, acting on phosphates	5	102	0.022
GO:0016829	F	lyase activity	8	229	0.027
GO:0015035	F	protein disulfide oxidoreductase activity	4	83	0.042
GO:0015036	F	disulfide oxidoreductase activity	4	83	0.042
GO:0020037	F	heme binding	17	709	0.046
GO:0046906	F	tetrapyrrole binding	17	712	0.048

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0007017	P	microtubule-based process	40	107	3.10E-13
GO:0015979	P	photosynthesis	36	89	7.50E-13
GO:0006928	P	movement of cell or subcellular component	29	59	3.10E-12
GO:0007018	P	microtubule-based movement	29	59	3.10E-12
GO:0009765	P	photosynthesis, light harvesting	18	22	1.10E-10
GO:0006260	P	DNA replication	24	50	3.30E-10
GO:0019684	P	photosynthesis, light reaction	20	34	6.70E-10
GO:0005975	P	carbohydrate metabolic process	104	673	3.10E-09
GO:0007049	P	cell cycle	30	99	1.60E-08
GO:0022402	P	cell cycle process	25	71	2.20E-08
GO:0000278	P	mitotic cell cycle	19	41	3.50E-08
GO:0007059	P	chromosome segregation	16	31	1.40E-07
GO:0055114	P	oxidation-reduction process	213	1849	3.10E-07
GO:1903047	P	mitotic cell cycle process	17	39	3.70E-07
GO:0006091	P	generation of precursor metabolites and energy	25	90	9.50E-07
GO:0048285	P	organelle fission	15	38	4.70E-06
GO:0000280	P	nuclear division	14	33	5.00E-06
GO:0044710	P	single-organism metabolic process	290	2807	1.10E-05
GO:0007067	P	mitotic nuclear division	13	31	1.20E-05
GO:0044699	P	single-organism process	434	4462	1.20E-05
GO:0098813	P	nuclear chromosome segregation	12	27	1.60E-05
GO:1901136	P	carbohydrate derivative catabolic process	13	36	4.30E-05
GO:0006261	P	DNA-dependent DNA replication	7	8	4.70E-05
GO:0046348	P	amino sugar catabolic process	12	31	4.90E-05
GO:0006026	P	aminoglycan catabolic process	12	31	4.90E-05
GO:1901072	P	glucosamine-containing compound catabolic process	12	31	4.90E-05
GO:1901071	P	glucosamine-containing compound metabolic process	12	31	4.90E-05
GO:0006030	P	chitin metabolic process	12	31	4.90E-05
GO:0006032	P	chitin catabolic process	12	31	4.90E-05
GO:0006040	P	amino sugar metabolic process	12	31	4.90E-05

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:0008152	P	metabolic process	835	9388	4.90E-05
GO:0000819	P	sister chromatid segregation	11	26	5.30E-05
GO:0000070	P	mitotic sister chromatid segregation	11	26	5.30E-05
GO:0006508	P	proteolysis	82	634	6.30E-05
GO:0044042	P	glucan metabolic process	21	95	0.00012
GO:0006073	P	cellular glucan metabolic process	21	95	0.00012
GO:0006022	P	aminoglycan metabolic process	12	35	0.00013
GO:0044264	P	cellular polysaccharide metabolic process	21	96	0.00014
GO:0030261	P	chromosome condensation	7	11	0.00019
GO:0007076	P	mitotic chromosome condensation	7	11	0.00019
GO:0071554	P	cell wall organization or biogenesis	20	94	0.00027
GO:0051301	P	cell division	7	12	0.00028
GO:0071103	P	DNA conformation change	12	39	0.00029
GO:0045930	P	negative regulation of mitotic cell cycle	6	8	0.0003
GO:0007093	P	mitotic cell cycle checkpoint	6	8	0.0003
GO:0005976	P	polysaccharide metabolic process	21	105	0.0004
GO:0010564	P	regulation of cell cycle process	9	23	0.0004
GO:0051726	P	regulation of cell cycle	12	41	0.00042
GO:0016998	P	cell wall macromolecule catabolic process	12	43	0.00061
GO:1901565	P	organonitrogen compound catabolic process	13	50	0.00064
GO:0044262	P	cellular carbohydrate metabolic process	26	151	0.00067
GO:0000910	P	cytokinesis	6	10	0.00071
GO:0044036	P	cell wall macromolecule metabolic process	12	44	0.00072
GO:0006259	P	DNA metabolic process	33	216	0.0009
GO:1901991	P	negative regulation of mitotic cell cycle phase transition	5	7	0.0012
GO:0010948	P	negative regulation of cell cycle process	5	7	0.0012
GO:1901988	P	negative regulation of cell cycle phase transition	5	7	0.0012

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:0007346	P	regulation of mitotic cell cycle	8	22	0.0012
GO:1902589	P	single-organism organelle organization	19	102	0.0015
GO:0006270	P	DNA replication initiation	4	4	0.0017
GO:0000075	P	cell cycle checkpoint	7	18	0.0018
GO:0051276	P	chromosome organization	18	96	0.0019
GO:0045786	P	negative regulation of cell cycle	6	14	0.0026
GO:0006323	P	DNA packaging	8	27	0.0036
GO:0006996	P	organelle organization	33	240	0.0041
GO:0051784	P	negative regulation of nuclear division	4	6	0.0045
GO:0051985	P	negative regulation of chromosome segregation	4	6	0.0045
GO:1902100	P	negative regulation of metaphase/anaphase transition of cell cycle	4	6	0.0045
GO:0033046	P	negative regulation of sister chromatid segregation	4	6	0.0045
GO:0045841	P	negative regulation of mitotic metaphase/anaphase transition	4	6	0.0045
GO:2000816	P	negative regulation of mitotic sister chromatid separation	4	6	0.0045
GO:0031577	P	spindle checkpoint	4	6	0.0045
GO:0007094	P	mitotic spindle assembly checkpoint	4	6	0.0045
GO:2001251	P	negative regulation of chromosome organization	4	6	0.0045
GO:0045839	P	negative regulation of mitotic nuclear division	4	6	0.0045
GO:0071174	P	mitotic spindle checkpoint	4	6	0.0045
GO:0071173	P	spindle assembly checkpoint	4	6	0.0045
GO:0033048	P	negative regulation of mitotic sister chromatid segregation	4	6	0.0045
GO:0000226	P	microtubule cytoskeleton organization	7	24	6.80E-03
GO:0016043	P	cellular component organization	45	372	7.90E-03
GO:0010639	P	negative regulation of organelle organization	4	8	9.30E-03
GO:0051129	P	negative regulation of cellular component organization	4	8	9.30E-03

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:1901990	P	regulation of mitotic cell cycle phase transition	5	15	1.40E-02
GO:0051304	P	chromosome separation	5	15	1.40E-02
GO:1901987	P	regulation of cell cycle phase transition	5	15	1.40E-02
GO:0044770	P	cell cycle phase transition	5	15	1.40E-02
GO:0044772	P	mitotic cell cycle phase transition	5	15	1.40E-02
GO:0048523	P	negative regulation of cellular process	6	22	1.50E-02
GO:0006979	P	response to oxidative stress	19	138	2.50E-02
GO:0009415	P	response to water	3	6	2.50E-02
GO:0006814	P	sodium ion transport	3	6	2.50E-02
GO:0006265	P	DNA topological change	4	12	2.70E-02
GO:0030244	P	cellulose biosynthetic process	8	41	2.70E-02
GO:0030243	P	cellulose metabolic process	8	41	2.70E-02
GO:0009628	P	response to abiotic stimulus	5	19	2.90E-02
GO:0048519	P	negative regulation of biological process	7	34	3.00E-02
GO:0006013	P	mannose metabolic process	4	13	3.30E-02
GO:0071840	P	cellular component organization or biogenesis	45	412	3.40E-02
GO:0007088	P	regulation of mitotic nuclear division	4	14	4.00E-02
GO:0051783	P	regulation of nuclear division	4	14	4.00E-02
GO:0030071	P	regulation of mitotic metaphase/anaphase transition	4	14	4.00E-02
GO:0051983	P	regulation of chromosome segregation	4	14	4.00E-02
GO:0051306	P	mitotic sister chromatid separation	4	14	4.00E-02
GO:0033044	P	regulation of chromosome organization	4	14	0.04
GO:0044784	P	metaphase/anaphase transition of cell cycle	4	14	0.04
GO:0033045	P	regulation of sister chromatid segregation	4	14	0.04
GO:0033047	P	regulation of mitotic sister chromatid segregation	4	14	0.04
GO:0010965	P	regulation of mitotic sister chromatid separation	4	14	0.04

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:1902099	P	regulation of metaphase/anaphase transition of cell cycle	4	14	0.04
GO:0007091	P	metaphase/anaphase transition of mitotic cell cycle	4	14	0.04
GO:0001101	P	response to acid chemical	3	8	0.043
GO:0016787	F	hydrolase activity	326	2363	1.20E-20
GO:0003824	F	catalytic activity	936	9441	1.40E-18
GO:0008017	F	microtubule binding	34	69	3.90E-14
GO:0016798	F	hydrolase activity, acting on glycosyl bonds	94	463	4.40E-14
GO:0015631	F	tubulin binding	34	70	5.40E-14
GO:0004553	F	hydrolase activity, hydrolyzing O-glycosyl compounds	91	443	6.30E-14
GO:0003777	F	microtubule motor activity	29	59	3.10E-12
GO:0008092	F	cytoskeletal protein binding	35	97	2.10E-11
GO:0003774	F	motor activity	30	71	2.60E-11
GO:0005507	F	copper ion binding	40	137	1.70E-10
GO:0016491	F	oxidoreductase activity	239	2023	7.10E-09
GO:0008236	F	serine-type peptidase activity	48	244	1.70E-07
GO:0017171	F	serine hydrolase activity	48	244	1.70E-07
GO:0004185	F	serine-type carboxypeptidase activity	28	105	4.40E-07
GO:0004180	F	carboxypeptidase activity	28	107	6.00E-07
GO:0070008	F	serine-type exopeptidase activity	28	110	9.60E-07
GO:0017111	F	nucleoside-triphosphatase activity	70	481	7.80E-06
GO:0008238	F	exopeptidase activity	28	128	1.20E-05
GO:0070011	F	peptidase activity, acting on L-amino acid peptides	79	574	1.30E-05
GO:0008233	F	peptidase activity	81	596	1.50E-05
GO:0016462	F	pyrophosphatase activity	71	505	1.90E-05
GO:0016818	F	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	71	512	2.80E-05
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	13	36	4.30E-05
GO:0004568	F	chitinase activity	12	31	4.90E-05
GO:0016817	F	hydrolase activity, acting on acid anhydrides	71	528	6.70E-05

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:0016757	F	transferase activity, transferring glycosyl groups	89	727	1.80E-04
GO:0008061	F	chitin binding	7	12	2.80E-04
GO:0004601	F	peroxidase activity	24	131	4.90E-04
GO:0016684	F	oxidoreductase activity, acting on peroxide as acceptor	24	133	0.00059
GO:0004252	F	serine-type endopeptidase activity	20	103	0.00075
GO:0016788	F	hydrolase activity, acting on ester bonds	62	498	0.0011
GO:0005544	F	calcium-dependent phospholipid binding	7	17	0.0014
GO:0016758	F	transferase activity, transferring hexosyl groups	74	629	0.0016
GO:0003993	F	acid phosphatase activity	7	18	0.0018
GO:0016209	F	antioxidant activity	24	148	0.0022
GO:0004175	F	endopeptidase activity	39	290	0.0027
GO:0048037	F	cofactor binding	55	469	0.0063
GO:0050664	F	oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor	4	7	0.0066
GO:0003896	F	DNA primase activity	3	3	0.007
GO:0050662	F	coenzyme binding	44	361	0.0076
GO:0016655	F	oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor	4	9	0.013
GO:0004190	F	aspartic-type endopeptidase activity	18	119	0.013
GO:0070001	F	aspartic-type peptidase activity	18	119	0.013
GO:0003887	F	DNA-directed DNA polymerase activity	4	10	0.017
GO:0016760	F	cellulose synthase (UDP-forming) activity	8	37	0.017
GO:0016759	F	cellulose synthase activity	8	37	0.017
GO:0016651	F	oxidoreductase activity, acting on NAD(P)H	8	38	0.019
GO:0034061	F	DNA polymerase activity	4	11	0.022
GO:0016872	F	intramolecular lyase activity	3	6	0.025
GO:0003924	F	GTPase activity	11	66	0.027
GO:0003916	F	DNA topoisomerase activity	4	12	0.027

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:0004096	F	catalase activity	3	7	0.033
GO:0004332	F	fructose-bisphosphate aldolase activity	3	7	0.033
GO:0005543	F	phospholipid binding	8	43	0.034
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	18	140	0.047
GO:0005871	C	kinesin complex	29	58	2.20E-12
GO:0015630	C	microtubule cytoskeleton	36	100	1.20E-11
GO:0005875	C	microtubule associated complex	29	73	1.80E-10
GO:0044430	C	cytoskeletal part	37	135	3.50E-09
GO:0005856	C	cytoskeleton	38	146	7.20E-09
GO:0034357	C	photosynthetic membrane	19	57	2.10E-06
GO:0044436	C	thylakoid part	19	58	2.60E-06
GO:0009579	C	thylakoid	19	58	2.60E-06
GO:0009521	C	photosystem	17	52	8.80E-06
GO:0005694	C	chromosome	21	80	1.40E-05
GO:0044815	C	DNA packaging complex	12	29	2.90E-05
GO:0044427	C	chromosomal part	18	66	3.80E-05
GO:0048046	C	apoplast	13	36	4.30E-05
GO:0030312	C	external encapsulating structure	18	71	8.40E-05
GO:0005618	C	cell wall	18	71	8.40E-05
GO:0043234	C	protein complex	77	616	0.00027
GO:0042651	C	thylakoid membrane	11	33	0.00029
GO:0005576	C	extracellular region	14	55	0.00048
GO:0009522	C	photosystem I	7	15	0.00079
GO:0009654	C	photosystem II oxygen evolving complex	9	26	0.00083
GO:0000786	C	nucleosome	9	26	0.00083
GO:0032993	C	protein-DNA complex	9	26	0.00083
GO:0043232	C	intracellular non-membrane-bounded organelle	63	506	0.001
GO:0043228	C	non-membrane-bounded organelle	63	506	0.001
GO:0009523	C	photosystem II	10	34	0.0012
GO:1990204	C	oxidoreductase complex	10	34	0.0012
GO:0009538	C	photosystem I reaction center	5	8	0.0018
GO:0071944	C	cell periphery	23	138	0.0019
GO:0005874	C	microtubule	7	19	0.0023
GO:0044422	C	organelle part	59	496	0.0036

Table S5.19 Gene ontology significant terms for cluster two in the flowers-only cluster analysis (continued).

GO:0044446	C	intracellular organelle part	59	496	0.0036
GO:0000785	C	chromatin	9	34	0.0038
GO:0019898	C	extrinsic component of membrane	8	28	0.0043
GO:0099512	C	supramolecular fiber	7	22	0.0046
GO:0099513	C	polymeric cytoskeletal fiber	7	22	0.0046
GO:0042555	C	MCM complex	3	3	0.007
GO:0000796	C	condensin complex	3	3	0.007
GO:0000775	C	chromosome, centromeric region	4	8	0.0093
GO:0098687	C	chromosomal region	4	9	0.013
GO:0000793	C	condensed chromosome	5	18	0.025

Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:0044710	P	single-organism metabolic process	56	2807	7.00E-07
GO:0044723	P	single-organism carbohydrate metabolic process	13	238	2.00E-06
GO:0016052	P	carbohydrate catabolic process	7	56	3.60E-06
GO:0044281	P	small molecule metabolic process	20	610	7.90E-06
GO:0005975	P	carbohydrate metabolic process	21	673	9.60E-06
GO:0009117	P	nucleotide metabolic process	9	149	3.70E-05
GO:0019637	P	organophosphate metabolic process	11	230	4.00E-05
GO:0006753	P	nucleoside phosphate metabolic process	9	151	4.10E-05
GO:0006732	P	coenzyme metabolic process	8	117	4.50E-05
GO:0006163	P	purine nucleotide metabolic process	8	120	5.30E-05
GO:0072521	P	purine-containing compound metabolic process	8	122	5.90E-05
GO:0005984	P	disaccharide metabolic process	5	39	8.40E-05
GO:0055086	P	nucleobase-containing small molecule metabolic process	9	181	0.00015
GO:0051186	P	cofactor metabolic process	8	143	0.00017
GO:0044699	P	single-organism process	69	4462	0.00017
GO:0009311	P	oligosaccharide metabolic process	5	46	0.00017
GO:0044724	P	single-organism carbohydrate catabolic process	5	47	0.00019
GO:0009150	P	purine ribonucleotide metabolic process	7	109	0.00019
GO:0009259	P	ribonucleotide metabolic process	7	109	0.00019
GO:0019752	P	carboxylic acid metabolic process	12	328	0.00021
GO:0044262	P	cellular carbohydrate metabolic process	8	151	0.00024
GO:0009719	P	response to endogenous stimulus	7	113	0.00024
GO:0009725	P	response to hormone	7	113	0.00024
GO:0043436	P	oxoacid metabolic process	12	337	0.00027
GO:0019693	P	ribose phosphate metabolic process	7	116	0.00028

Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis (continued).

GO:0010033	P	response to organic substance	7	119	0.00032
GO:0009167	P	purine ribonucleoside monophosphate metabolic process	6	87	0.00039
GO:0009161	P	ribonucleoside monophosphate metabolic process	6	87	0.00039
GO:0009126	P	purine nucleoside monophosphate metabolic process	6	87	0.00039
GO:0044763	P	single-organism cellular process	40	2241	0.00044
GO:0009123	P	nucleoside monophosphate metabolic process	6	90	0.00046
GO:0044712	P	single-organism catabolic process	6	91	0.00049
GO:0009733	P	response to auxin	6	93	0.00055
GO:0006082	P	organic acid metabolic process	12	366	0.00055
GO:0072524	P	pyridine-containing compound metabolic process	5	61	0.00058
GO:1901564	P	organonitrogen compound metabolic process	20	849	0.00063
GO:0009165	P	nucleotide biosynthetic process	5	69	0.00098
GO:1901293	P	nucleoside phosphate biosynthetic process	5	69	0.00098
GO:0046128	P	purine ribonucleoside metabolic process	6	105	0.001
GO:0042278	P	purine nucleoside metabolic process	6	105	0.001
GO:0090407	P	organophosphate biosynthetic process	6	108	0.0011
GO:0009119	P	ribonucleoside metabolic process	6	108	0.0011
GO:0046031	P	ADP metabolic process	4	41	0.0011
GO:0009179	P	purine ribonucleoside diphosphate metabolic process	4	41	0.0011
GO:0006757	P	ATP generation from ADP	4	41	0.0011
GO:0009135	P	purine nucleoside diphosphate metabolic process	4	41	0.0011
GO:0009185	P	ribonucleoside diphosphate metabolic process	4	41	0.0011
GO:0006096	P	glycolytic process	4	41	0.0011

Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis (continued).

GO:0044711	P	single-organism biosynthetic process	13	457	0.0012
GO:0046483	P	heterocycle metabolic process	33	1823	0.0012
GO:0046351	P	disaccharide biosynthetic process	3	19	0.0014
GO:0006165	P	nucleoside diphosphate phosphorylation	4	46	0.0017
GO:0046939	P	nucleotide phosphorylation	4	46	0.0017
GO:1901360	P	organic cyclic compound metabolic process	33	1866	0.0018
GO:0005985	P	sucrose metabolic process	3	21	0.0018
GO:0006090	P	pyruvate metabolic process	4	47	0.0018
GO:0009132	P	nucleoside diphosphate metabolic process	4	48	0.002
GO:0006164	P	purine nucleotide biosynthetic process	4	50	0.0023
GO:0009312	P	oligosaccharide biosynthetic process	3	23	0.0023
GO:0072522	P	purine-containing compound biosynthetic process	4	52	0.0026
GO:1901135	P	carbohydrate derivative metabolic process	9	274	0.0026
GO:1901575	P	organic substance catabolic process	8	224	0.0028
GO:0042221	P	response to chemical	8	226	0.0029
GO:0009116	P	nucleoside metabolic process	6	132	0.003
GO:1901657	P	glycosyl compound metabolic process	6	132	0.003
GO:0046496	P	nicotinamide nucleotide metabolic process	4	57	0.0035
GO:0019362	P	pyridine nucleotide metabolic process	4	57	0.0035
GO:0006733	P	oxidoreduction coenzyme metabolic process	4	58	0.0038
GO:0009108	P	coenzyme biosynthetic process	4	58	0.0038
GO:0032787	P	monocarboxylic acid metabolic process	6	143	0.0044
GO:0009056	P	catabolic process	8	244	0.0046
GO:0006725	P	cellular aromatic compound metabolic process	31	1829	0.0046

Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis (continued).

GO:0006807	P	nitrogen compound metabolic process	38	2383	0.0046
GO:0034641	P	cellular nitrogen compound metabolic process	35	2183	0.0061
GO:0042254	P	ribosome biogenesis	3	39	0.0091
GO:0009152	P	purine ribonucleotide biosynthetic process	3	39	0.0091
GO:0009260	P	ribonucleotide biosynthetic process	3	39	0.0091
GO:0046390	P	ribose phosphate biosynthetic process	3	39	0.0091
GO:0051188	P	cofactor biosynthetic process	4	80	0.011
GO:0034637	P	cellular carbohydrate biosynthetic process	4	80	0.011
GO:0046034	P	ATP metabolic process	4	82	0.012
GO:1901137	P	carbohydrate derivative biosynthetic process	5	128	0.012
GO:1901566	P	organonitrogen compound biosynthetic process	13	618	0.013
GO:0006139	P	nucleobase-containing compound metabolic process	28	1740	0.014
GO:0022613	P	ribonucleoprotein complex biogenesis	3	47	0.015
GO:0009199	P	ribonucleoside triphosphate metabolic process	4	88	0.015
GO:0009205	P	purine ribonucleoside triphosphate metabolic process	4	88	0.015
GO:0009144	P	purine nucleoside triphosphate metabolic process	4	88	0.015
GO:0016051	P	carbohydrate biosynthetic process	4	90	0.016
GO:0009141	P	nucleoside triphosphate metabolic process	4	90	0.016
GO:0006091	P	generation of precursor metabolites and energy	4	90	0.016
GO:0006520	P	cellular amino acid metabolic process	5	148	0.021
GO:0006396	P	RNA processing	6	211	0.025
GO:1901362	P	organic cyclic compound biosynthetic process	20	1207	0.027
GO:0055114	P	oxidation-reduction process	28	1849	0.028

Table S5.20 Gene ontology significant terms for cluster three in the flowers-only cluster analysis (continued).

GO:0071704	P	organic substance metabolic process	83	6726	0.029
GO:1901576	P	organic substance biosynthetic process	30	2025	0.03
GO:0018130	P	heterocycle biosynthetic process	19	1159	0.034
GO:0009058	P	biosynthetic process	31	2143	0.036
GO:0044249	P	cellular biosynthetic process	29	1984	0.037
GO:0065008	P	regulation of biological quality	5	176	0.039
GO:0008152	P	metabolic process	110	9388	0.04
GO:0055085	P	transmembrane transport	15	889	0.046
GO:0016874	F	ligase activity	7	137	0.00072
GO:0016879	F	ligase activity, forming carbon-nitrogen bonds	5	66	0.00081
GO:0016616	F	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	5	123	0.01
GO:0003824	F	catalytic activity	114	9441	0.012
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	5	140	0.017
GO:0016491	F	oxidoreductase activity	29	2023	0.046
GO:0035251	F	UDP-glucosyltransferase activity	3	77	0.049

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast.

GO term	Ont	Description	Number in input list	Number in BG/Ref	p-value
GO:1901564	P	organonitrogen compound metabolic process	367	849	1.30E-25
GO:0043603	P	cellular amide metabolic process	210	408	1.70E-21
GO:0006518	P	peptide metabolic process	207	403	3.90E-21
GO:0006412	P	translation	198	383	1.70E-20
GO:0043043	P	peptide biosynthetic process	199	391	5.10E-20
GO:0043604	P	amide biosynthetic process	199	391	5.10E-20
GO:1901566	P	organonitrogen compound biosynthetic process	269	618	2.90E-19
GO:0009058	P	biosynthetic process	683	2143	1.30E-17
GO:0044249	P	cellular biosynthetic process	635	1984	1.20E-16
GO:1901576	P	organic substance biosynthetic process	643	2025	2.90E-16
GO:0006807	P	nitrogen compound metabolic process	728	2383	2.60E-15
GO:0044271	P	cellular nitrogen compound biosynthetic process	512	1559	5.20E-15
GO:0034641	P	cellular nitrogen compound metabolic process	664	2183	1.30E-13
GO:0010467	P	gene expression	515	1634	1.30E-12
GO:0009059	P	macromolecule biosynthetic process	495	1581	8.70E-12
GO:0034645	P	cellular macromolecule biosynthetic process	494	1580	1.10E-11
GO:0044281	P	small molecule metabolic process	212	610	2.50E-08
GO:0051186	P	cofactor metabolic process	69	143	3.40E-07
GO:0051179	P	localization	442	1566	3.60E-06
GO:0044710	P	single-organism metabolic process	743	2807	3.90E-06
GO:0019637	P	organophosphate metabolic process	91	230	5.00E-06
GO:0051234	P	establishment of localization	438	1559	5.80E-06
GO:0019725	P	cellular homeostasis	58	124	5.90E-06
GO:0006810	P	transport	436	1552	6.20E-06
GO:0006732	P	coenzyme metabolic process	55	117	9.10E-06
GO:0042592	P	homeostatic process	59	130	1.00E-05
GO:0065008	P	regulation of biological quality	72	176	2.00E-05

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0045454	P	cell redox homeostasis	49	103	2.20E-05
GO:1901575	P	organic substance catabolic process	86	224	2.20E-05
GO:0051188	P	cofactor biosynthetic process	41	80	2.70E-05
GO:0009056	P	catabolic process	88	244	0.00011
GO:1901135	P	carbohydrate derivative metabolic process	95	274	0.00018
GO:0044711	P	single-organism biosynthetic process	144	457	0.00021
GO:0055086	P	nucleobase-containing small molecule metabolic process	68	181	0.00025
GO:1901362	P	organic cyclic compound biosynthetic process	331	1207	0.00036
GO:0006753	P	nucleoside phosphate metabolic process	58	151	0.00044
GO:0016192	P	vesicle-mediated transport	50	124	0.00044
GO:0008610	P	lipid biosynthetic process	66	179	0.00045
GO:0044255	P	cellular lipid metabolic process	74	208	0.0005
GO:0009108	P	coenzyme biosynthetic process	29	58	0.00052
GO:0009117	P	nucleotide metabolic process	57	149	0.00052
GO:0009057	P	macromolecule catabolic process	52	134	0.00069
GO:0070925	P	organelle assembly	14	18	0.00075
GO:0006629	P	lipid metabolic process	135	441	0.00083
GO:0006414	P	translational elongation	13	16	0.00087
GO:0022607	P	cellular component assembly	52	136	0.00091
GO:0019693	P	ribose phosphate metabolic process	46	116	0.00096
GO:0043241	P	protein complex disassembly	18	29	0.00096
GO:0043624	P	cellular protein complex disassembly	18	29	0.00096
GO:0072524	P	pyridine-containing compound metabolic process	29	61	0.00097
GO:0044248	P	cellular catabolic process	61	168	0.00099
GO:0018130	P	heterocycle biosynthetic process	313	1159	0.0012
GO:1901360	P	organic cyclic compound metabolic process	483	1866	0.0012
GO:0006082	P	organic acid metabolic process	114	366	0.0012
GO:0032787	P	monocarboxylic acid metabolic process	53	143	0.0014

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0022411	P	cellular component disassembly	18	31	0.0017
GO:0032984	P	macromolecular complex disassembly	18	31	0.0017
GO:0019752	P	carboxylic acid metabolic process	103	328	0.0017
GO:0006733	P	oxidoreduction coenzyme metabolic process	27	58	0.0018
GO:0016043	P	cellular component organization	114	372	0.002
GO:0051261	P	protein depolymerization	10	11	0.002
GO:0030042	P	actin filament depolymerization	10	11	0.002
GO:0009116	P	nucleoside metabolic process	49	132	0.002
GO:1901657	P	glycosyl compound metabolic process	49	132	0.002
GO:0090407	P	organophosphate biosynthetic process	42	108	0.0021
GO:0043436	P	oxoacid metabolic process	104	337	0.0025
GO:0016485	P	protein processing	11	14	0.0026
GO:0051604	P	protein maturation	11	14	0.0026
GO:0046496	P	nicotinamide nucleotide metabolic process	26	57	0.0027
GO:0019362	P	pyridine nucleotide metabolic process	26	57	0.0027
GO:0007034	P	vacuolar transport	14	22	0.0029
GO:0019438	P	aromatic compound biosynthetic process	306	1154	0.0031
GO:0006413	P	translational initiation	15	25	0.0032
GO:0071840	P	cellular component organization or biogenesis	122	412	0.0036
GO:0044257	P	cellular protein catabolic process	31	75	0.0037
GO:0051603	P	proteolysis involved in cellular protein catabolic process	31	75	0.0037
GO:0009259	P	ribonucleotide metabolic process	41	109	0.0037
GO:0009150	P	purine ribonucleotide metabolic process	41	109	0.0037
GO:0044085	P	cellular component biogenesis	60	177	0.0038
GO:0016052	P	carbohydrate catabolic process	25	56	0.004
GO:0006470	P	protein dephosphorylation	35	89	0.004
GO:0016311	P	dephosphorylation	35	89	0.004
GO:0006457	P	protein folding	39	103	0.0042

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0006662	P	glycerol ether metabolic process	18	35	0.0045
GO:0018904	P	ether metabolic process	18	35	0.0045
GO:0046483	P	heterocycle metabolic process	463	1823	0.0046
GO:0030163	P	protein catabolic process	32	80	0.0047
GO:0009132	P	nucleoside diphosphate metabolic process	22	48	0.0053
GO:0016236	P	macroautophagy	9	11	0.0053
GO:1905037	P	autophagosome organization	9	11	0.0053
GO:0000045	P	autophagosome assembly	9	11	0.0053
GO:0042278	P	purine nucleoside metabolic process	39	105	0.0054
GO:0046128	P	purine ribonucleoside metabolic process	39	105	0.0054
GO:0006163	P	purine nucleotide metabolic process	43	120	0.0061
GO:0006465	P	signal peptide processing	8	9	0.0061
GO:0007033	P	vacuole organization	10	14	0.0064
GO:0006165	P	nucleoside diphosphate phosphorylation	21	46	0.0065
GO:0046939	P	nucleotide phosphorylation	21	46	0.0065
GO:0006811	P	ion transport	134	469	0.0066
GO:0006725	P	cellular aromatic compound metabolic process	461	1829	0.0071
GO:0034654	P	nucleobase-containing compound biosynthetic process	286	1093	0.0074
GO:0009141	P	nucleoside triphosphate metabolic process	34	90	0.0074
GO:0072521	P	purine-containing compound metabolic process	43	122	0.0076
GO:0009119	P	ribonucleoside metabolic process	39	108	0.0078
GO:0015991	P	ATP hydrolysis coupled proton transport	15	29	0.0088
GO:0015988	P	energy coupled proton transmembrane transport, against electrochemical gradient	15	29	0.0088
GO:0090662	P	ATP hydrolysis coupled transmembrane transport	15	29	0.0088
GO:0009199	P	ribonucleoside triphosphate metabolic process	33	88	0.0089

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0009144	P	purine nucleoside triphosphate metabolic process	33	88	0.0089
GO:0009205	P	purine ribonucleoside triphosphate metabolic process	33	88	0.0089
GO:0031163	P	metallo-sulfur cluster assembly	8	10	0.0092
GO:0016226	P	iron-sulfur cluster assembly	8	10	0.0092
GO:0006914	P	autophagy	11	18	0.0099
GO:0006631	P	fatty acid metabolic process	33	89	0.01
GO:0044265	P	cellular macromolecule catabolic process	33	89	0.01
GO:0055114	P	oxidation-reduction process	462	1849	0.011
GO:0043244	P	regulation of protein complex disassembly	6	6	0.012
GO:0006417	P	regulation of translation	6	6	0.012
GO:1901070	P	guanosine-containing compound biosynthetic process	6	6	0.012
GO:0034248	P	regulation of cellular amide metabolic process	6	6	0.012
GO:0006448	P	regulation of translational elongation	6	6	0.012
GO:0071822	P	protein complex subunit organization	39	112	0.012
GO:1902600	P	hydrogen ion transmembrane transport	22	53	0.013
GO:1901565	P	organonitrogen compound catabolic process	21	50	0.013
GO:0044724	P	single-organism carbohydrate catabolic process	20	47	0.014
GO:0006366	P	transcription from RNA polymerase II promoter	16	35	0.016
GO:0006812	P	cation transport	103	363	0.018
GO:0009167	P	purine ribonucleoside monophosphate metabolic process	31	87	0.019
GO:0009161	P	ribonucleoside monophosphate metabolic process	31	87	0.019
GO:0009126	P	purine nucleoside monophosphate metabolic process	31	87	0.019
GO:1901068	P	guanosine-containing compound metabolic process	8	12	0.019

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0009147	P	pyrimidine nucleoside triphosphate metabolic process	6	7	0.019
GO:0010608	P	postranscriptional regulation of gene expression	6	7	0.019
GO:1901136	P	carbohydrate derivative catabolic process	16	36	0.02
GO:0097659	P	nucleic acid-templated transcription	261	1016	0.02
GO:0006351	P	transcription, DNA-templated	261	1016	0.02
GO:1902578	P	single-organism localization	106	377	0.02
GO:0032774	P	RNA biosynthetic process	261	1018	0.021
GO:0044262	P	cellular carbohydrate metabolic process	48	151	0.022
GO:0044282	P	small molecule catabolic process	12	24	0.022
GO:0006790	P	sulfur compound metabolic process	17	40	0.022
GO:0046132	P	pyrimidine ribonucleoside biosynthetic process	5	5	0.022
GO:0006241	P	CTP biosynthetic process	5	5	0.022
GO:0009220	P	pyrimidine ribonucleotide biosynthetic process	5	5	0.022
GO:0046036	P	CTP metabolic process	5	5	0.022
GO:0046051	P	UTP metabolic process	5	5	0.022
GO:0009148	P	pyrimidine nucleoside triphosphate biosynthetic process	5	5	0.022
GO:0009218	P	pyrimidine ribonucleotide metabolic process	5	5	0.022
GO:0046134	P	pyrimidine nucleoside biosynthetic process	5	5	0.022
GO:0006183	P	GTP biosynthetic process	5	5	0.022
GO:0009209	P	pyrimidine ribonucleoside triphosphate biosynthetic process	5	5	0.022
GO:0009208	P	pyrimidine ribonucleoside triphosphate metabolic process	5	5	0.022
GO:0006228	P	UTP biosynthetic process	5	5	0.022
GO:0006090	P	pyruvate metabolic process	19	47	0.024
GO:0006739	P	NADP metabolic process	8	13	0.026

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0006662	P	glycerol ether metabolic process	18	35	0.0045
GO:0018904	P	ether metabolic process	18	35	0.0045
GO:0046483	P	heterocycle metabolic process	463	1823	0.0046
GO:0030163	P	protein catabolic process	32	80	0.0047
GO:0009132	P	nucleoside diphosphate metabolic process	22	48	0.0053
GO:0016236	P	macroautophagy	9	11	0.0053
GO:1905037	P	autophagosome organization	9	11	0.0053
GO:0000045	P	autophagosome assembly	9	11	0.0053
GO:0042278	P	purine nucleoside metabolic process	39	105	0.0054
GO:0046128	P	purine ribonucleoside metabolic process	39	105	0.0054
GO:0006163	P	purine nucleotide metabolic process	43	120	0.0061
GO:0006465	P	signal peptide processing	8	9	0.0061
GO:0007033	P	vacuole organization	10	14	0.0064
GO:0006165	P	nucleoside diphosphate phosphorylation	21	46	0.0065
GO:0046939	P	nucleotide phosphorylation	21	46	0.0065
GO:0006811	P	ion transport	134	469	0.0066
GO:0006725	P	cellular aromatic compound metabolic process	461	1829	0.0071
GO:0034654	P	nucleobase-containing compound biosynthetic process	286	1093	0.0074
GO:0009141	P	nucleoside triphosphate metabolic process	34	90	0.0074
GO:0072521	P	purine-containing compound metabolic process	43	122	0.0076
GO:0009119	P	ribonucleoside metabolic process	39	108	0.0078
GO:0015991	P	ATP hydrolysis coupled proton transport	15	29	0.0088
GO:0015988	P	energy coupled proton transmembrane transport, against electrochemical gradient	15	29	0.0088
GO:0090662	P	ATP hydrolysis coupled transmembrane transport	15	29	0.0088
GO:0009199	P	ribonucleoside triphosphate metabolic process	33	88	0.0089

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0009144	P	purine nucleoside triphosphate metabolic process	33	88	0.0089
GO:0009205	P	purine ribonucleoside triphosphate metabolic process	33	88	0.0089
GO:0031163	P	metallo-sulfur cluster assembly	8	10	0.0092
GO:0016226	P	iron-sulfur cluster assembly	8	10	0.0092
GO:0006914	P	autophagy	11	18	0.0099
GO:0006631	P	fatty acid metabolic process	33	89	0.01
GO:0044265	P	cellular macromolecule catabolic process	33	89	0.01
GO:0055114	P	oxidation-reduction process	462	1849	0.011
GO:0043244	P	regulation of protein complex disassembly	6	6	0.012
GO:0006417	P	regulation of translation	6	6	0.012
GO:1901070	P	guanosine-containing compound biosynthetic process	6	6	0.012
GO:0034248	P	regulation of cellular amide metabolic process	6	6	0.012
GO:0006448	P	regulation of translational elongation	6	6	0.012
GO:0071822	P	protein complex subunit organization	39	112	0.012
GO:1902600	P	hydrogen ion transmembrane transport	22	53	0.013
GO:1901565	P	organonitrogen compound catabolic process	21	50	0.013
GO:0044724	P	single-organism carbohydrate catabolic process	20	47	0.014
GO:0006366	P	transcription from RNA polymerase II promoter	16	35	0.016
GO:0006812	P	cation transport	103	363	0.018
GO:0009167	P	purine ribonucleoside monophosphate metabolic process	31	87	0.019
GO:0009161	P	ribonucleoside monophosphate metabolic process	31	87	0.019
GO:0009126	P	purine nucleoside monophosphate metabolic process	31	87	0.019
GO:1901068	P	guanosine-containing compound metabolic process	8	12	0.019

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0009147	P	pyrimidine nucleoside triphosphate metabolic process	6	7	0.019
GO:0010608	P	postranscriptional regulation of gene expression	6	7	0.019
GO:1901136	P	carbohydrate derivative catabolic process	16	36	0.02
GO:0097659	P	nucleic acid-templated transcription	261	1016	0.02
GO:0006351	P	transcription, DNA-templated	261	1016	0.02
GO:1902578	P	single-organism localization	106	377	0.02
GO:0032774	P	RNA biosynthetic process	261	1018	0.021
GO:0044262	P	cellular carbohydrate metabolic process	48	151	0.022
GO:0044282	P	small molecule catabolic process	12	24	0.022
GO:0006790	P	sulfur compound metabolic process	17	40	0.022
GO:0046132	P	pyrimidine ribonucleoside biosynthetic process	5	5	0.022
GO:0006241	P	CTP biosynthetic process	5	5	0.022
GO:0009220	P	pyrimidine ribonucleotide biosynthetic process	5	5	0.022
GO:0046036	P	CTP metabolic process	5	5	0.022
GO:0046051	P	UTP metabolic process	5	5	0.022
GO:0009148	P	pyrimidine nucleoside triphosphate biosynthetic process	5	5	0.022
GO:0009218	P	pyrimidine ribonucleotide metabolic process	5	5	0.022
GO:0046134	P	pyrimidine nucleoside biosynthetic process	5	5	0.022
GO:0006183	P	GTP biosynthetic process	5	5	0.022
GO:0009209	P	pyrimidine ribonucleoside triphosphate biosynthetic process	5	5	0.022
GO:0009208	P	pyrimidine ribonucleoside triphosphate metabolic process	5	5	0.022
GO:0006228	P	UTP biosynthetic process	5	5	0.022
GO:0006090	P	pyruvate metabolic process	19	47	0.024
GO:0006739	P	NADP metabolic process	8	13	0.026

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0072525	P	pyridine-containing compound biosynthetic process	8	13	0.026
GO:0009123	P	nucleoside monophosphate metabolic process	31	90	0.026
GO:0006357	P	regulation of transcription from RNA polymerase II promoter	11	22	0.027
GO:0032268	P	regulation of cellular protein metabolic process	10	19	0.027
GO:0006139	P	nucleobase-containing compound metabolic process	428	1740	0.029
GO:0006694	P	steroid biosynthetic process	14	32	0.031
GO:0008154	P	actin polymerization or depolymerization	14	32	0.031
GO:0043632	P	modification-dependent macromolecule catabolic process	23	63	0.032
GO:0019941	P	modification-dependent protein catabolic process	23	63	0.032
GO:0006511	P	ubiquitin-dependent protein catabolic process	23	63	0.032
GO:0006633	P	fatty acid biosynthetic process	23	63	0.032
GO:0006767	P	water-soluble vitamin metabolic process	11	23	0.034
GO:0009110	P	vitamin biosynthetic process	11	23	0.034
GO:0006766	P	vitamin metabolic process	11	23	0.034
GO:0042364	P	water-soluble vitamin biosynthetic process	11	23	0.034
GO:0044283	P	small molecule biosynthetic process	52	172	0.034
GO:0006839	P	mitochondrial transport	8	14	0.034
GO:0006066	P	alcohol metabolic process	8	14	0.034
GO:0019751	P	polyol metabolic process	8	14	0.034
GO:0044765	P	single-organism transport	100	363	0.034
GO:0051246	P	regulation of protein metabolic process	10	20	0.035
GO:0046039	P	GTP metabolic process	5	6	0.035
GO:0046131	P	pyrimidine ribonucleoside metabolic process	5	6	0.035
GO:0006213	P	pyrimidine nucleoside metabolic process	5	6	0.035
GO:0046034	P	ATP metabolic process	28	82	0.036

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0008202	P	steroid metabolic process	14	33	0.036
GO:0010556	P	regulation of macromolecule biosynthetic process	231	911	0.04
GO:2000112	P	regulation of cellular macromolecule biosynthetic process	231	911	0.04
GO:0010468	P	regulation of gene expression	234	924	0.04
GO:0072511	P	divalent inorganic cation transport	6	9	0.041
GO:0070838	P	divalent metal ion transport	6	9	0.041
GO:0015693	P	magnesium ion transport	6	9	0.041
GO:0006452	P	translational frameshifting	4	4	0.041
GO:0042822	P	pyridoxal phosphate metabolic process	4	4	0.041
GO:0042823	P	pyridoxal phosphate biosynthetic process	4	4	0.041
GO:0034250	P	positive regulation of cellular amide metabolic process	4	4	0.041
GO:0043243	P	positive regulation of protein complex disassembly	4	4	0.041
GO:0045901	P	positive regulation of translational elongation	4	4	0.041
GO:0045727	P	positive regulation of translation	4	4	0.041
GO:0051259	P	protein oligomerization	4	4	0.041
GO:0006449	P	regulation of translational termination	4	4	0.041
GO:0045905	P	positive regulation of translational termination	4	4	0.041
GO:0006122	P	mitochondrial electron transport, ubiquinol to cytochrome c	4	4	0.041
GO:0042775	P	mitochondrial ATP synthesis coupled electron transport	4	4	0.041
GO:0009889	P	regulation of biosynthetic process	231	913	0.042
GO:0031326	P	regulation of cellular biosynthetic process	231	913	0.042
GO:0044712	P	single-organism catabolic process	30	91	0.043
GO:0009165	P	nucleotide biosynthetic process	24	69	0.043
GO:1901293	P	nucleoside phosphate biosynthetic process	24	69	0.043

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0070972	P	protein localization to endoplasmic reticulum	10	21	0.043
GO:0006613	P	cotranslational protein targeting to membrane	7	12	0.043
GO:0006614	P	SRP-dependent cotranslational protein targeting to membrane	7	12	0.043
GO:0043933	P	macromolecular complex subunit organization	47	156	0.044
GO:0006644	P	phospholipid metabolic process	22	62	0.044
GO:0016054	P	organic acid catabolic process	8	15	0.044
GO:0006757	P	ATP generation from ADP	16	41	0.045
GO:0006096	P	glycolytic process	16	41	0.045
GO:0046031	P	ADP metabolic process	16	41	0.045
GO:0009179	P	purine ribonucleoside diphosphate metabolic process	16	41	0.045
GO:0009135	P	purine nucleoside diphosphate metabolic process	16	41	4.50E-02
GO:0009185	P	ribonucleoside diphosphate metabolic process	16	41	4.50E-02
GO:0051171	P	regulation of nitrogen compound metabolic process	232	919	4.50E-02
GO:0044699	P	single-organism process	1046	4462	4.60E-02
GO:0046348	P	amino sugar catabolic process	13	31	4.60E-02
GO:0006026	P	aminoglycan catabolic process	13	31	0.046
GO:1901072	P	glucosamine-containing compound catabolic process	13	31	0.046
GO:0006030	P	chitin metabolic process	13	31	0.046
GO:0006032	P	chitin catabolic process	13	31	0.046
GO:0006040	P	amino sugar metabolic process	13	31	0.046
GO:1901071	P	glucosamine-containing compound metabolic process	13	31	0.046
GO:0061024	P	membrane organization	12	28	0.048
GO:0003735	F	structural constituent of ribosome	158	274	6.50E-20
GO:0005198	F	structural molecule activity	163	290	1.10E-19
GO:0008135	F	translation factor activity, RNA binding	38	64	4.30E-06
GO:0003723	F	RNA binding	113	321	2.90E-05

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0016616	F	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	53	123	8.10E-05
GO:0003743	F	translation initiation factor activity	23	36	0.00015
GO:0015036	F	disulfide oxidoreductase activity	39	83	0.00017
GO:0015035	F	protein disulfide oxidoreductase activity	39	83	0.00017
GO:0042578	F	phosphoric ester hydrolase activity	61	155	0.00019
GO:0015078	F	hydrogen ion transmembrane transporter activity	42	93	0.0002
GO:0019843	F	rRNA binding	18	25	0.00027
GO:0008324	F	cation transmembrane transporter activity	84	240	0.00033
GO:0016491	F	oxidoreductase activity	528	2023	0.00037
GO:0016667	F	oxidoreductase activity, acting on a sulfur group of donors	49	120	0.0004
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors	54	140	0.00062
GO:0005543	F	phospholipid binding	23	43	0.00096
GO:0022890	F	inorganic cation transmembrane transporter activity	67	189	0.00096
GO:0030234	F	enzyme regulator activity	51	135	0.0012
GO:0098772	F	molecular function regulator	55	150	0.0014
GO:0022892	F	substrate-specific transporter activity	142	478	0.0016
GO:0022891	F	substrate-specific transmembrane transporter activity	130	434	0.002
GO:0015077	F	monovalent inorganic cation transmembrane transporter activity	46	123	0.0024
GO:0048037	F	cofactor binding	138	469	0.0025
GO:0070003	F	threonine-type peptidase activity	13	19	0.0026
GO:0004298	F	threonine-type endopeptidase activity	13	19	0.0026
GO:0003746	F	translation elongation factor activity	11	14	0.0026
GO:0016791	F	phosphatase activity	47	127	0.0026
GO:0008289	F	lipid binding	32	77	0.003

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0051540	F	metal cluster binding	26	58	0.0032
GO:0051536	F	iron-sulfur cluster binding	26	58	0.0032
GO:0019205	F	nucleobase-containing compound kinase activity	14	23	0.0039
GO:0043021	F	ribonucleoprotein complex binding	7	7	0.0068
GO:0005544	F	calcium-dependent phospholipid binding	11	17	0.0073
GO:0004869	F	cysteine-type endopeptidase inhibitor activity	9	12	0.0078
GO:0015075	F	ion transmembrane transporter activity	113	390	0.0085
GO:0016701	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	20	45	0.0098
GO:0001076	F	transcription factor activity, RNA polymerase II transcription factor binding	11	18	0.0099
GO:0001104	F	RNA polymerase II transcription cofactor activity	11	18	0.0099
GO:0043168	F	anion binding	33	89	0.01
GO:0016702	F	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	17	36	0.011
GO:0004721	F	phosphoprotein phosphatase activity	34	93	0.011
GO:0005452	F	inorganic anion exchanger activity	6	6	0.012
GO:0016407	F	acetyltransferase activity	20	47	0.014
GO:0009055	F	electron carrier activity	49	150	0.014
GO:0016853	F	isomerase activity	46	139	0.014
GO:0008080	F	N-acetyltransferase activity	19	44	0.014
GO:0008270	F	zinc ion binding	206	780	0.016
GO:0016810	F	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	19	45	0.017
GO:0060589	F	nucleoside-triphosphatase regulator activity	13	26	0.017

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0008081	F	phosphoric diester hydrolase activity	12	23	0.017
GO:0051213	F	dioxygenase activity	17	39	0.019
GO:1901681	F	sulfur compound binding	6	7	0.019
GO:0050662	F	coenzyme binding	102	361	0.02
GO:0016811	F	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	9	15	0.021
GO:0019842	F	vitamin binding	13	27	0.021
GO:0008047	F	enzyme activator activity	10	18	0.021
GO:0043022	F	ribosome binding	5	5	0.022
GO:0008097	F	5S rRNA binding	5	5	0.022
GO:0004550	F	nucleoside diphosphate kinase activity	5	5	0.022
GO:0051287	F	NAD binding	26	71	0.023
GO:0016676	F	oxidoreductase activity, acting on a heme group of donors, oxygen as acceptor	7	10	0.023
GO:0016675	F	oxidoreductase activity, acting on a heme group of donors	7	10	0.023
GO:0015002	F	heme-copper terminal oxidase activity	7	10	0.023
GO:0004129	F	cytochrome-c oxidase activity	7	10	0.023
GO:0033218	F	amide binding	7	10	0.023
GO:0016410	F	N-acyltransferase activity	19	47	0.024
GO:0030170	F	pyridoxal phosphate binding	24	65	0.026
GO:0016679	F	oxidoreductase activity, acting on diphenols and related substances as donors	9	16	0.027
GO:0005215	F	transporter activity	255	1001	0.028
GO:0016776	F	phosphotransferase activity, phosphate group as acceptor	6	8	0.029
GO:0004602	F	glutathione peroxidase activity	6	8	0.029
GO:0032561	F	guanyl ribonucleotide binding	69	236	0.029
GO:0005525	F	GTP binding	69	236	0.029
GO:0061135	F	endopeptidase regulator activity	14	32	0.031
GO:0030414	F	peptidase inhibitor activity	14	32	0.031
GO:0004866	F	endopeptidase inhibitor activity	14	32	0.031
GO:0061134	F	peptidase regulator activity	14	32	0.031

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0004722	F	protein serine/threonine phosphatase activity	25	70	0.032
GO:0004857	F	enzyme inhibitor activity	29	85	0.034
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	15	36	0.035
GO:0030695	F	GTPase regulator activity	10	20	0.035
GO:0000062	F	fatty-acyl-CoA binding	5	6	0.035
GO:0008889	F	glycerophosphodiester phosphodiesterase activity	5	6	0.035
GO:0015291	F	secondary active transmembrane transporter activity	47	154	3.80E-02
GO:0015103	F	inorganic anion transmembrane transporter activity	16	40	3.80E-02
GO:0072509	F	divalent inorganic cation transmembrane transporter activity	6	9	4.10E-02
GO:0015095	F	magnesium ion transmembrane transporter activity	6	9	4.10E-02
GO:0008199	F	ferric iron binding	4	4	4.10E-02
GO:0008121	F	ubiquinol-cytochrome-c reductase activity	4	4	4.10E-02
GO:0016681	F	oxidoreductase activity, acting on diphenols and related substances as donors, cytochrome as acceptor	4	4	4.10E-02
GO:0022857	F	transmembrane transporter activity	200	783	4.30E-02
GO:0008061	F	chitin binding	7	12	4.30E-02
GO:0003899	F	DNA-directed RNA polymerase activity	18	48	4.50E-02
GO:0004568	F	chitinase activity	13	31	4.60E-02
GO:0003712	F	transcription cofactor activity	15	38	4.80E-02
GO:0033764	F	steroid dehydrogenase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	12	28	4.80E-02
GO:0003854	F	3-beta-hydroxy-delta5-steroid dehydrogenase activity	12	28	4.80E-02
GO:0016229	F	steroid dehydrogenase activity	12	28	4.80E-02
GO:0005737	C	cytoplasm	336	733	7.30E-27
GO:0044444	C	cytoplasmic part	287	596	1.50E-25

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0005623	C	cell	621	1831	9.30E-21
GO:0044464	C	cell part	621	1831	9.30E-21
GO:0030529	C	intracellular ribonucleoprotein complex	173	314	3.80E-20
GO:1990904	C	ribonucleoprotein complex	173	314	3.80E-20
GO:0005622	C	intracellular	588	1726	5.80E-20
GO:0005840	C	ribosome	157	273	1.00E-19
GO:0044424	C	intracellular part	550	1649	3.60E-17
GO:0032991	C	macromolecular complex	350	929	7.00E-17
GO:0043229	C	intracellular organelle	438	1295	1.70E-14
GO:0043226	C	organelle	438	1295	1.70E-14
GO:0043232	C	intracellular non-membrane-bounded organelle	207	506	5.30E-13
GO:0043228	C	non-membrane-bounded organelle	207	506	5.30E-13
GO:0016020	C	membrane	586	2184	1.40E-05
GO:0044425	C	membrane part	353	1243	2.40E-05
GO:0031224	C	intrinsic component of membrane	304	1053	3.30E-05
GO:0044429	C	mitochondrial part	28	46	5.50E-05
GO:0016021	C	integral component of membrane	298	1038	5.50E-05
GO:0005783	C	endoplasmic reticulum	39	80	9.50E-05
GO:0005740	C	mitochondrial envelope	24	38	0.00012
GO:0043231	C	intracellular membrane-bounded organelle	240	834	0.00026
GO:0043227	C	membrane-bounded organelle	240	834	0.00026
GO:0031966	C	mitochondrial membrane	21	35	0.00052
GO:0044422	C	organelle part	150	496	0.00068
GO:0044446	C	intracellular organelle part	150	496	0.00068
GO:0005739	C	mitochondrion	40	95	0.00081
GO:0012505	C	endomembrane system	51	133	0.00096
GO:0005743	C	mitochondrial inner membrane	16	24	0.001
GO:0019866	C	organelle inner membrane	16	25	0.0014
GO:0043234	C	protein complex	177	616	0.0017
GO:0031090	C	organelle membrane	34	82	0.0023
GO:0005839	C	proteasome core complex	13	19	0.0026
GO:0031975	C	envelope	25	55	0.0033
GO:0031967	C	organelle envelope	25	55	0.0033

Table S5.21 Gene ontology significant terms for the upregulated genes in the P vs L contrast (continued).

GO:0042175	C	nuclear outer membrane-endoplasmic reticulum membrane network	22	46	0.0036
GO:0005789	C	endoplasmic reticulum membrane	22	46	0.0036
GO:0044432	C	endoplasmic reticulum part	22	46	0.0036
GO:0016592	C	mediator complex	13	20	0.0036
GO:0044455	C	mitochondrial membrane part	13	20	0.0036
GO:0098798	C	mitochondrial protein complex	13	20	0.0036
GO:0098800	C	inner mitochondrial membrane protein complex	12	18	0.0044
GO:0033177	C	proton-transporting two-sector ATPase complex, proton-transporting domain	12	19	0.0059
GO:0000502	C	proteasome complex	15	29	0.0088
GO:0016469	C	proton-transporting two-sector ATPase complex	22	52	0.011
GO:0015934	C	large ribosomal subunit	9	14	0.015
GO:0005654	C	nucleoplasm	24	61	0.015
GO:0044451	C	nucleoplasm part	24	61	0.015
GO:0005801	C	cis-Golgi network	7	9	0.016
GO:0033176	C	proton-transporting V-type ATPase complex	11	20	0.017
GO:0005829	C	cytosol	6	7	0.019
GO:0005746	C	mitochondrial respiratory chain	5	5	0.022
GO:0098803	C	respiratory chain complex	5	5	0.022
GO:0070469	C	respiratory chain	5	5	0.022
GO:0016272	C	prefoldin complex	8	14	0.034
GO:0048046	C	apoplast	15	36	0.035
GO:0005787	C	signal peptidase complex	5	6	0.035
GO:0044445	C	cytosolic part	5	6	0.035
GO:0005576	C	extracellular region	20	55	0.045
GO:0071944	C	cell periphery	42	138	0.049