

The Pennsylvania State University

The Graduate School

Graduate Degree Program in Integrative Biosciences

FLOWER DEVELOPMENT IN *THEOBROMA CACAO* L.:

AN ASSESSMENT OF MORPHOLOGICAL AND MOLECULAR CONSERVATION
OF FLORAL DEVELOPMENT BETWEEN *ARABIDOPSIS THALIANA* AND
THEOBROMA CACAO

A Thesis in

Integrative Biosciences

by

John-David Swanson

© 2005 John-David Swanson

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2005

UMI Number: 3173830



UMI Microform 3173830

Copyright 2005 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

The thesis of John-David Swanson was reviewed and approved* by the following:

Mark J. Gultinan
Professor of Plant Molecular Biology
Thesis Co-Advisor
Co-Chair of Committee

John E. Carlson
Associate Professor of Molecular Genetics
Thesis Co-Advisor
Co-Chair of Committee

Hong Ma
Professor of Biology

David R. Hunter
Assistant Professor of Statistics

Nina V. Fedoroff
Acting Co-Director, Graduate Education
Integrative Biosciences Graduate Program
Huck Institutes of the Life Sciences

*Signatures are on file in the Graduate School

ABSTRACT

Comparative anatomical and developmental studies on *Theobroma cacao* L. flower development were conducted to provide insight into the general mechanisms that control floral development. These studies were also conducted to assess the level of conservation of these systems among plant species, especially with reference to *Arabidopsis thaliana*. To this end, we underwent a study to compare “normal” flowers of *T. cacao* to what is already known in model species. To provide complementary approaches to the same set of questions, we examined *T. cacao* floral development at the morphological, genetic and gene expression levels. Morphological comparisons were made through the analysis of time-lapse photography and light and electron microscopy to create mathematical models of flower development for *T. cacao*. At the genetic level, we compared EST sequence data from *Arabidopsis*, poplar, cotton and *T. cacao* through phylogenetic methods. To precisely localize gene expression patterns during *T. cacao* flower development, we performed *in situ* hybridizations using the floral integrator *LEAFY* as a probe, as well as several ABC genes. Comparison of the *T. cacao* floral developmental program with that of *Arabidopsis* revealed that although the final sizes and morphologies of flowers in the two species differ, their developmental programs are strikingly similar both morphologically and genetically. Consistent with this analysis, a cross-species analysis of the current knowledge in this field indicates a high degree of conservation kingdom wide.

TABLE OF CONTENTS

Chapter 1 The Setting	1
1.1 <i>Theobroma cacao</i>	2
1.2 Flower Development	4
1.2.1.1 Pathways that enable floral transition	8
1.2.1.2 Pathways that promote floral transition	10
1.2.2 The integrators.....	12
1.2.3 The floral identity genes and the ABC model.....	14
1.2.4 Flowering in other species.....	19
1.3 The Flowers of <i>Theobroma cacao</i> and their Importance.....	20
1.3.1 Morphology	20
1.3.2 Growth and development	21
1.3.3 The effect of Witches' Broom on <i>T. cacao</i> flower development.....	23
1.4 Relevance.....	25
1.5 References.....	27
Chapter 2 Morphology of <i>T. cacao</i> Flowers and Comparison to Other Species	35
2.1 Introduction.....	36
2.2 Methods	39
2.2.1 Plant materials and growth conditions	39
2.2.2 Time-lapse photography and morphometric measurements.....	40
2.2.3 Light microscopy	40
2.2.4 Scanning electron microscopy.....	41
2.2.5 Measurement of floral tissues.....	42
2.3 Results.....	42
2.3.1 Analysis of floral development using time-lapse, macro-imaging.....	42
2.3.2 Microscopic analysis of floral development.....	46
2.3.3 Description of developmental events	53
2.3.4 Stages 1-6	54
2.3.5 Stages 7-12	55
2.3.6 Flower opening and abscission.....	57
2.3.7 Organs novel to <i>T. cacao</i>	58
2.4 Discussion.....	59
2.4.1 Floral morphogenesis	59
2.4.2 Comparison to the temporal growth of <i>Arabidopsis thaliana</i> flowers.....	61
2.4.3 Stage 1-6.....	61
2.4.4 Stage 7-12.....	63
2.4.5 Comparison of timing between <i>Arabidopsis</i> and <i>T. cacao</i>	64
2.4.6 Comparison to well characterized flowers of other species	66
2.5 References.....	71

Chapter 3 Design and Testing of Generalized PCR Primers for <i>Theobroma cacao</i> and other Rosidae Plant Species	73
3.1 Introduction.....	74
3.2 Methods	76
3.2.1 Preliminary data.....	76
3.2.2 RNA extraction and cDNA construction from <i>T. cacao</i>	76
3.2.3 Selection of genes, primer design, cloning, and sequencing of <i>T. cacao</i> cDNAs	77
3.2.4 Data analysis.....	80
3.2.5 Use of degenerate primers in other species	81
3.3 Results.....	82
3.3.1 Preliminary studies	82
3.3.2 <i>T. cacao</i> cDNA isolation	83
3.3.4 Data analysis.....	84
3.3.4.1 SNPs	84
3.3.4.1 Phylogenetic analysis	86
3.3.5 Fragment amplification in other flowering plant species	88
3.4 Discussion.....	90
3.4.1 Relationships of <i>T. cacao</i> , cotton and <i>Arabidopsis</i> and gene families	90
3.4.2 Use of amplicons as DNA markers in <i>T. cacao</i> and in other species...	93
3.5 References.....	95
Chapter 4 Spatial and Temporal Expression Patterns of Several Key Floral Regulatory Factors during Flower Development in <i>T. cacao</i>	97
4.1 Introduction.....	98
4.1.1 <i>LEAFY</i> and the ABC model	98
4.1.2 Methods of analyzing gene expression.....	100
4.1.3 <i>In situ</i> hybridization and the ABCs of <i>T. cacao</i>	101
4.2 Methods	102
4.2.1 Gene selection and isolation.....	102
4.2.2 Southern analysis of the <i>T. cacao LEAFY</i> gene	102
4.2.3 Dot blot hybridization of heterologous probes.....	104
4.2.4 <i>In situ</i> hybridization probe preparation	105
4.2.5 Tissue selection and fixation	107
4.2.6 Slide preparation.....	108
4.2.7 Hybridization and signal development.....	109
4.2.8 Slide analysis	110
4.3 Results.....	111
4.3.1 Gene selection and analysis of <i>in situ</i> hybridization probes.....	111
4.3.1.1 <i>CELLULOSE SYNTHASE</i> and <i>LEAFY</i> gene	111
4.3.1.2 <i>Arabidopsis</i> genes <i>APETALA1</i> , <i>APETALA3</i> , <i>PISTILLATA</i> , and <i>AGAMOUS</i>	112

4.3.2 <i>In situ</i> analysis	114
4.3.2.1 <i>CELLULOSE SYNTHASE (CS)</i>	115
4.3.2.2 <i>LEAFY (LFY)</i>	115
4.3.2.3 <i>APETALA1 (API)</i>	117
4.3.2.4 <i>APETALA3 (AP3)</i>	118
4.3.2.5 <i>PISTILLATA (PI)</i>	121
4.3.2.6 <i>AGAMOUS (AG)</i>	121
4.4 Discussion.....	125
4.4.1 Determination of <i>in situ</i> hybridization probes.....	125
4.4.2 <i>LEAFY</i>	126
4.4.3 <i>APETALA1</i>	128
4.4.4 <i>APETALA3</i>	128
4.4.5 <i>PISTILLATA</i>	129
4.4.6 <i>AGAMOUS</i>	130
4.4.7 Proposal of a <i>T. cacao</i> model of ABC flower development.....	131
4.5 References.....	134
Chapter 5 Conclusion	138
5.1 Flower Development of <i>T. cacao</i>	139
5.2 Is <i>T. cacao</i> Really Like <i>Arabidopsis</i> ?.....	140
5.3 Flower Development: Implications and Reflections	141
5.4 What is Next?	143
5.5 References.....	148
Appendix A Degenerate deoxyribonucleotide primers created from consensus sequences of a highly conserved set of <i>Arabidopsis</i> and <i>cotton</i> genes.....	150
i. Clones used to create degenerate primers.....	151
ii. Degenerate deoxyribonucleotide primers.....	154
Appendix B Sequences of <i>T. cacao</i> PCR fragments amplified with the degenerate primer set	155
Appendix C Phylogenetic trees created by Mega 2 to compare genes in <i>Arabidopsis</i> , <i>cotton</i> , <i>Populus</i> and <i>T. cacao</i>	161
i. Phylogenetic trees of <i>T. cacao</i> , <i>cotton</i> , <i>Arabidopsis</i> and <i>Populus</i>	162
ii. Phylogenetic trees of <i>T. cacao</i> , <i>cotton</i> , <i>Populus</i> , and gene family members of <i>Arabidopsis</i> and <i>O. sativa</i>	174
iii. Phylogenetic trees of <i>T. cacao</i> , <i>cotton</i> , <i>Populus</i> , and gene family members of <i>Arabidopsis</i>	178
Appendix D Analysis of the gene expression of <i>LEAFY</i> , <i>API</i> , <i>AP3</i> , <i>PI</i> , <i>AG</i> , and <i>CELLULOSE SYNTHASE</i> in <i>Arabidopsis</i> using the GENEVESTIGATOR database.....	182

LIST OF FIGURES

Figure 1-1: Pathways that promote or enable the floral transition.....	7
Figure 1-2: Major floral promotion pathways.	8
Figure 1-3: The floral integrators.....	9
Figure 1-4: The ABC model of flower development.....	15
Figure 1-5: The pathways, post integrators, of the ABC model leading to floral organ initiation and differentiation	18
Figure 1-6: A cacao flower (Wood and Lass, 1985).....	22
Figure 1-7: <i>T. cacao</i> flowers infected with <i>C. pernicioso</i>	24
Figure 2-1: Floral anatomy of <i>T. cacao</i> . Images of transversely and longitudinally sectioned 22 day old flower buds	38
Figure 2-2: The growth, velocity, and opening angles of growing <i>T. cacao</i> flower buds over time.....	45
Figure 2-3: Micrographs showing important features of the <i>T. cacao</i> pedicle, and features of the petal.....	46
Figure 2-4 A-G: SEM photographs showing the ontogeny of the <i>T. cacao</i> flower and a photograph of a flower cushion.....	49
Figure 2-5: Comparison of <i>T. cacao</i> and <i>Arabidopsis</i> flower development.....	52
Figure 2-6: Photographs of a <i>T. cacao</i> flower and an <i>Arabidopsis</i> flower to allow comparison of final floral structure.	62
Figure 2-7: Comparison of the duration of each stage in <i>Arabidopsis</i> vs. <i>T. cacao</i> . <i>Arabidopsis</i> (horizontal axis) vs. <i>T. cacao</i> (vertical axis) from stage 2-12 as defined in Smyth et al. (1990)	66
Figure 2-8: Comparison of stages of flower growth of <i>T. cacao</i> , tobacco, and tomato.	68
Figure 3-1: Example of a typical PCR amplification using heterologous primers.	79

Figure 3-2: Phylogenetic tree of <i>T. cacao</i> (Tca), cotton (Ghi) <i>Arabidopsis</i> (Ath), and <i>Populus</i> (Ptr) based on chloroplast <i>atpB</i> sequence.	83
Figure 3-3: Plots depicting p-distance among <i>T. cacao</i> , cotton and <i>Arabidopsis</i> based on the combined data from all forty-three extracted cDNAs	86
Figure 3-4: Generalized dendrogram of expected phylogenetic relationships among <i>T. cacao</i> , cotton, <i>Arabidopsis</i> and <i>Populus</i>	87
Figure 3-5: Dendrogram showing cluster analysis of PCR using <i>T. cacao</i> degenerate primers in <i>Citrus limon</i> , <i>Carica papaya</i> , <i>Cucumis sativus</i> , <i>T. cacao</i> , and <i>Aesculus glabra</i>	90
Figure 4-1: Genomic Southern blot of putative <i>T. cacao</i> <i>LEAFY</i> gene.	113
Figure 4-2: Dot Blot showing cross-hybridization of <i>Arabidopsis</i> and <i>T. cacao</i> DNA to <i>Arabidopsis</i> and <i>T. cacao</i> cDNA.	114
Figure 4-3: <i>In situ</i> hybridization of <i>CELLULOSE SYNTHASE</i> in PSU-SCA6 plants.....	116
Figure 4-4: <i>In situ</i> hybridization of <i>LEAFY</i> probe in PSU-SCA6 plants.....	117
Figure 4-5: <i>In situ</i> hybridization of <i>Arabidopsis AP1</i> in PSU-SCA6 plants.	119
Figure 4-6: <i>In situ</i> hybridization of <i>Arabidopsis AP3</i> in PSU-SCA6 plants	120
Figure 4-7: <i>In situ</i> hybridization of <i>Arabidopsis PI</i> in PSU-SCA6 plants.....	122
Figure 4-8: <i>In situ</i> hybridization of <i>Arabidopsis PI</i> in stage 6 PSU-SCA6 flower buds.....	123
Figure 4-9: <i>In situ</i> hybridization of <i>Arabidopsis AG</i> in PSU-SCA6 plants.	124
Figure 4-10: ABC model for <i>T. cacao</i>	133

LIST OF TABLES

Table 2-1 : Definitions and abbreviations of flower measurements made in <i>T. cacao</i> flowers.....	44
Table 2-2 : Size and age of <i>T. cacao</i> flower buds sampled for microscopy.	48
Table 2-3 : Growth models of various organs during the development of <i>T. cacao</i> flower buds.	50
Table 2-4 : Growth models of percentage growths of various organs during the development of <i>T. cacao</i> flower buds.....	51
Table 2-5 : Comparative stages in <i>T. cacao</i> and <i>Arabidopsis</i> flower development	53
Table 3-1 : Genes extracted from <i>T. cacao</i> using degenerate primers.....	85
Table 3-2 : PCR results using <i>T. cacao</i> degenerate primers in reactions containing DNA template isolated from <i>Citrus limon</i> , <i>Carica papaya</i> , <i>Cucumis sativus</i> , <i>T. cacao</i> , and <i>Aesculus glabra</i>	89
Table 4-1 : List of restriction enzymes and RNA Polymerases used for <i>in situ</i> probe preparation.....	105

ACKNOWLEDGEMENTS

This thesis was in no way the work of a single person, it was truly “integrative”.

First and foremost, I would like to thank my parents for supporting me throughout the entire degree process. Thank you Mum for being there, and Dad your spirit, and love of life has never left me; I have felt your hand on my shoulder even though you left us suddenly.

I would like to thank my family here in the USA especially Holly, and my grandmother at home. I love you both.

To people who have helped me at Penn State: Joe Verica, Francesca Chiaromonte, Sharon Pishak, Ann Young, Siela Maximova, Jill Davidson, Alex Lee, Sam St Clair, Claude dePamphilis, the kind staff at the Electron microscopy facility, DNA sequencing facility, Cell flow lab, Carlson lab, dePamphilis lab, and Ma lab. Your advice and help has been appreciated.

Chen Chang-Bin for his beautiful calligraphy.

John Pfau and Bruce McPherson for helping me out during that last semester.

My friends, peers and students, your banter, advice and friendship pulled me through some very difficult times.

My committee, Mark Gultinan, John Carlson, Hong Ma, and David Hunter, thank you all for helping me from beginning, to end.

Chapter 1

The Setting



Earth

1.1 *Theobroma cacao*

Theobroma cacao L. is an understory tree endemic to the Amazon rainforest. The tree produces the cocoa bean which when processed, provides cocoa, the main ingredient for chocolate. *T. cacao* is thought to have been discovered by the Olmecs in Central America as far back as 600 BC (Hurst et al., 2002). The Mayans and Aztecs are recorded drinking liquid concoctions made from the cocoa bean in ceremonies as well as using the bean for currency (Hurst et al., 2002). Cocoa was brought to Europe in 1519 and spread throughout Europe during the 16th and 17th centuries where it was refined to the form we know today (Lopez, 2002). Currently, the United States is one of the largest consumers of chocolate and the trade generated from dried cacao beans worldwide is estimated at 58 billion dollars per year (Morais, 2005).

T. cacao is grown exclusively within 20 degrees of the equator in many tropical countries, with the largest producing countries being Cote d'Ivoire, Ghana, Indonesia, Nigeria, Brazil, Cameroon, Ecuador, and Malaysia (Bartolome, 1951; Wood and Lass, 1985). The majority of the five million hectare cacao crop is found on small, 1-2 ha farms (Duguma et al., 2001; Ramirez et al., 2001). Unfortunately, *T. cacao* is a very hard crop to grow profitably due to several endemic problems including: high yield variability, cocoa price volatility, and an annual average loss of 40% of the potential crop due to pests and diseases (Wood and Lass, 1985).

Fungal pathogens are the most devastating of these problems. The major pathogens are: Black Pod Rot (*Phytophthora palmivora*), Monilla Pod Rot (*Moniliophthora roreri*), and Witches Broom (*Crinipellis perniciososa*) (Wood and Lass, 1985). Black Pod Rot has been a major problem in Africa and the West Indies. Monilla Pod Rot and Witches Broom are problems in Central and South America. Witches Broom alone has been responsible for decreasing yields by 60% in Bahia, Brazil's major cocoa producing region (Purseglove, 1968; Wood and Lass, 1985; Smith, 1992). In an

effort to stabilize cocoa production, there has been an increasing amount of research on the molecular and physiological mechanisms that lead to these symptoms.

Theobroma cacao is a member of the family Malveaceae that includes the commercially important crop *Gossypium hirsutum* (cotton). The family Malvaceae is a member of the order Eurosids II which includes the model plant *Arabidopsis thaliana* (Soltis et al., 2000). Due to the abundance of data, these relatively closely related species form an invaluable tool to be able to apply to *T. cacao* research.

The genome of *T. cacao* is relatively small, being roughly 445 Mbp (Arumuganathan, Carlson, Gultinan, unpublished data). Its genetic material is located on ten chromosomes (Flament et al., 2001). Genomic work in *T. cacao* is generally very recent, involving sequencing of 4,455 EST sequences and developing a unigene set of 1,380 members (Jones et al., 2002; Bennett, 2003). Verica et al., (2004) identified a set of 1,256 unigenes from leaves treated to elicit a defense response. This set of unigenes was used to construct a microarray (M. J. Gultinan pers. comm.). Recently, two genomic libraries were constructed that are becoming a useful tool for *T. cacao* molecular biologists (Clement et al., 2004). High density molecular maps have been constructed comprised of 465 markers with an average spacing of 1.7cM (Pugh et al., 2004). These map data have allowed analyses of the geographic origin of *T. cacao* with the results indicating that *T. cacao* originated from a single region of the upper Amazon (Bennett, 2003). These data have also been used extensively to elucidate quantitative trait loci (QTL) associated with resistance to *Phytophthora palmivora* and *Crinipellis perniciosia*. There has been one QTL identified in two separate populations that account for 47% of the resistance to *P. palmivora*, as well as a single QTL that explains ~35% of resistance to *C. perniciosia* (Bennett, 2003). Finally, there has been a DNA fingerprinting study in an attempt to assess the validity of the labeling of the global collection of germplasm (Saunders et al., 2004). A transformation system has also been developed for *T. cacao* and has produced stable transgenic plants (Maximova et al., 2003). This work highlights that further molecular research into this crop is required that we might better understand

its growth and response to its environment and find mechanisms to decrease disease susceptibility and thus increase yield.

1.2 Flower Development

Flowers are the reproductive organs of the plant. It is believed that the floral organs evolved from modified leaves and formed many different morphologies through millions of years of selective pressure and co-evolution with pollinators. Flowers originate from vegetative meristems that convert to floral meristems, and will form several whorls, which in perfect flowers include the sepals, petals, stamens, and carpels. The growth and differentiation of the meristematic cells into the highly specialized organs of the four whorls requires a complex program of floral developmental cascades. Over the last 15 years, there has been a rapid increase in knowledge of the mechanisms that control floral induction, floral patterning, and floral organ identity. This information results from intensive studies in primarily three species: *Antirrhinum majus*, *Arabidopsis thaliana*, and *Petunia hybrida*. Of the three species, *Arabidopsis* is the best studied. This is fortuitous for this thesis as, of the three model species, *Arabidopsis* is most closely related to *T. cacao*. Due to this fact, the studies relating to *Arabidopsis* will be the primary focus of this literature review.

It is also important to note at this point that there are several unifying principles for flower development that seem to hold true for most flowers studied. These were discussed in a review by Jack, (2004) and will be briefly discussed here. The first unifying principle is known as the ABC model. This model was developed primarily through the identification of developmental mutants and the subsequent characterization of genes associated with these mutations. The results of these experiments have indicated that flowers arise from a well-controlled and ordered cascade of genes being turned on and off in specific tissues at specific times. It has been shown to be applicable to a wide range of monocot and eudicot species (Bowman et al., 1991; Coen and Meyerowitz,

1991; Meyerowitz et al., 1998; Ambrose et al., 2000; Theissen et al., 2000; Fornara et al., 2003).

The ABC model refers to a combination of three classes of genes that effect organogenesis of the four whorls of the flower (Jack, 2004; Ma 2005). Simplified, A class gene products are required for the identity of sepals, the combination of the A and B gene products are required for the determination of petals, the combination of B and C gene products are required for determination of stamen identity, and expression of the C gene products are required for the identity of carpels. Interestingly, A and C genes are mutually repressive. This model has been extended to include two further classes of genes, aptly named D and E (Theissen et al., 2000; Ma, 2005). The D function genes were identified initially in *Petunia* and were associated with ovule identity (Angenent and Colombo, 1996). The E function genes are thought to act in concert with the A, AB, BC, C and D complexes for sepal, petal, stamen, and carpel identity (Theissen et al., 2000; Ma et al. 2005). The specific genes associated with the ABC model will be discussed later in more detail.

The second unifying principal according to Jack (2004) involves the central role of the *LEAFY (LFY)* gene. *LFY* is both an integrator of the floral pathways (Blazquez et al., 1997; Nilsson et al., 1998) and an activator of the floral organ identity genes (ABC genes) (Weigel and Meyerowitz, 1993; Parcy et al., 1998; Lenhard et al., 2001; Lohmann et al., 2001). Orthologs of *LFY* have been identified in many flowering and non-flowering plant species including: Tomato (*FLORICAULA*), *N. tabacum (NFLI)*, *Populus (PTLF)*, *Eucalyptus (ELF1, and 2)*, *Pinus radiata (NEEDLY)*, and *Vitis vinifera (VFL)* (Gocal et al., 2001).

When simplified, flower production can be broken down into several distinct stages (Jack, 2004). These stages are: 1) In response to environmental and internal signals the plant switches growth from vegetative to reproductive growth, 2) signals from these pathways are integrated and lead to the activation of a few meristem

identity genes, 3) the meristem identity genes in turn activate floral organ identity genes in distinct regions of the floral primordium, and 4) once organ identity has been established, genes are up regulated that allow the formation of the floral organs. These stages can be illustrated in the following diagram (Figure 1-1), and are discussed in the following sections.

1.2.1 Control of flowering time

In general, for plants to produce flowers they require input from five major promotion pathways: long-day/photoperiod, light quality, gibberellins (GA), autonomous (maturity), and the vernalization pathways. Simply put, the plant must integrate signals from all of these inputs to be able to transition from a vegetative state to a flowering state. In *Arabidopsis*, all five promotion pathways need to be integrated for the plant to produce flowers; however other species are able to produce flowers after receiving input from only a single or few pathways. These major pathways are summarized in Figure 1-2.

In a recent review by Boss et al. (2004), the floral pathways were divided into two groups. The first group are pathways that enable floral transition and include vernalization and developmental age (Figure 1-2, lower half). These were defined as repressors that function to antagonize the activation of the floral pathway integrators (Boss et al., 2004). The best defined of these genes is *FLOWERING LOCUS C (FLC)* which functions to integrate signals from the vernalization pathway, the developmental age pathway, as well as various developmental promoters.

The second group are pathways that promote the floral transition and include photoperiod, light quality and hormones (Figure 1-2, top half). These pathways are defined as providing mechanisms to promote the floral transition integrators.

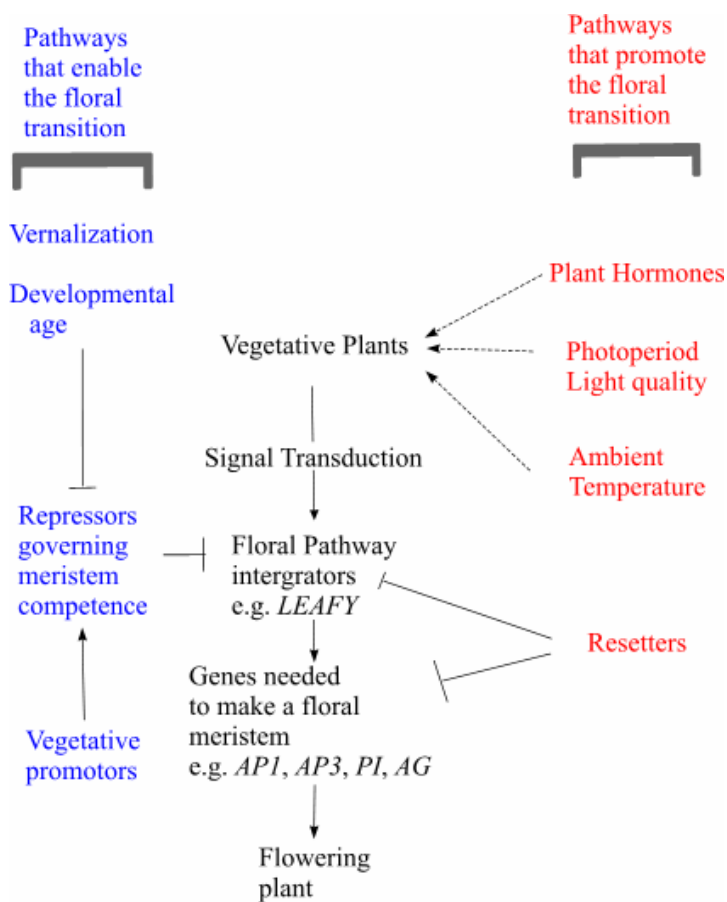


Figure 1-1: Pathways that promote or enable the floral transition. This view of the vegetative to floral meristem transition clearly shows the integration of signals both internal and external to the plant, which leads to the expression of genes required to produce a flowering plant (adapted from Boss et al., 2004).

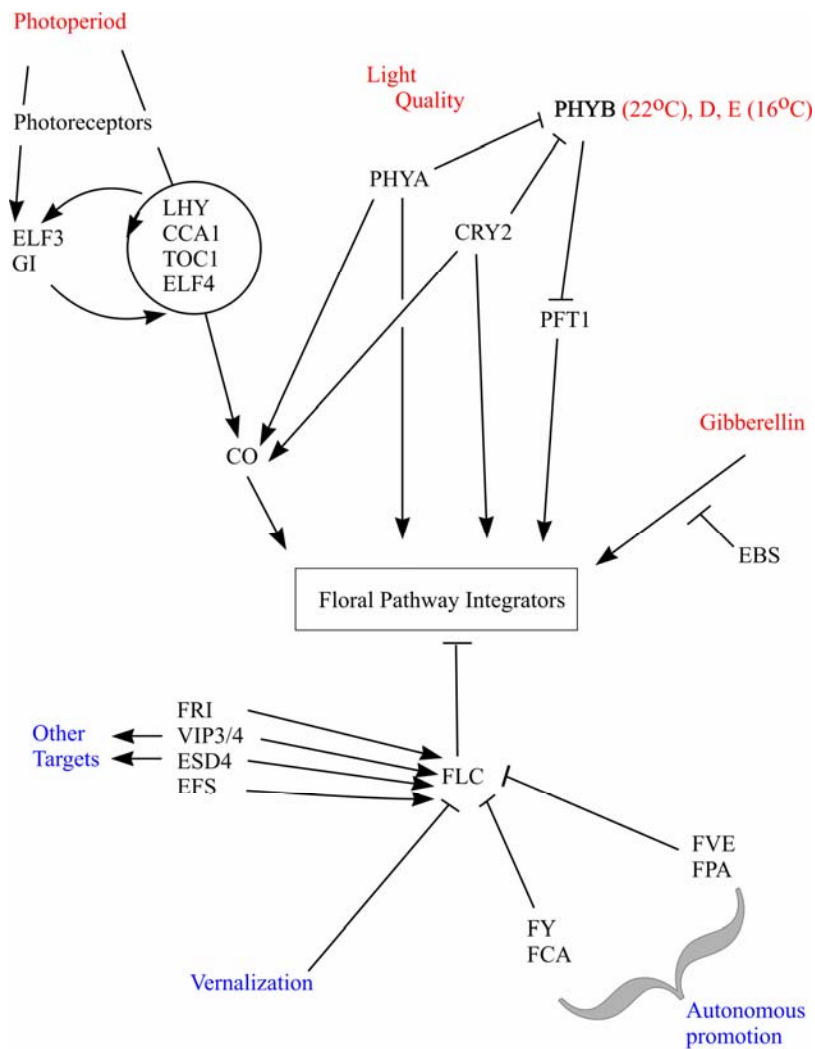


Figure 1-2: Major floral promotion pathways. (adapted from Boss et al., (2004))

1.2.1.1 Pathways that enable floral transition

As stated above, the MADS box gene *FLC* is the most well understood gene in this pathway. It has been shown that mutants deficient in *FLC* flower early, while mutants with autonomous levels of *FLC* flower late (Michaels and Amasino, 1999;

Sheldon et al., 1999; Michaels and Amasino, 2001). It also was shown that flowering time was directly related to the amount of *FLC* RNA/protein present. As a result of these experiments, it was postulated that *FLC* repressed the floral pathway integrators *FLOWERING TIME (FT)*, and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* (Kardailsky et al., 1999; Blazquez and Weigel, 2000) (Figure 1-3).

Michaels and Amasino, (2001) showed that vernalization decreased the amount of *FLC* and allowed flowering to continue. Also, through mutant analysis, the autonomous pathway gene *FCA* showed that *FCA* by itself had little effect on flowering time whereas it did have an effect on levels of the *FLC* gene (Reeves and Coupland, 2001).

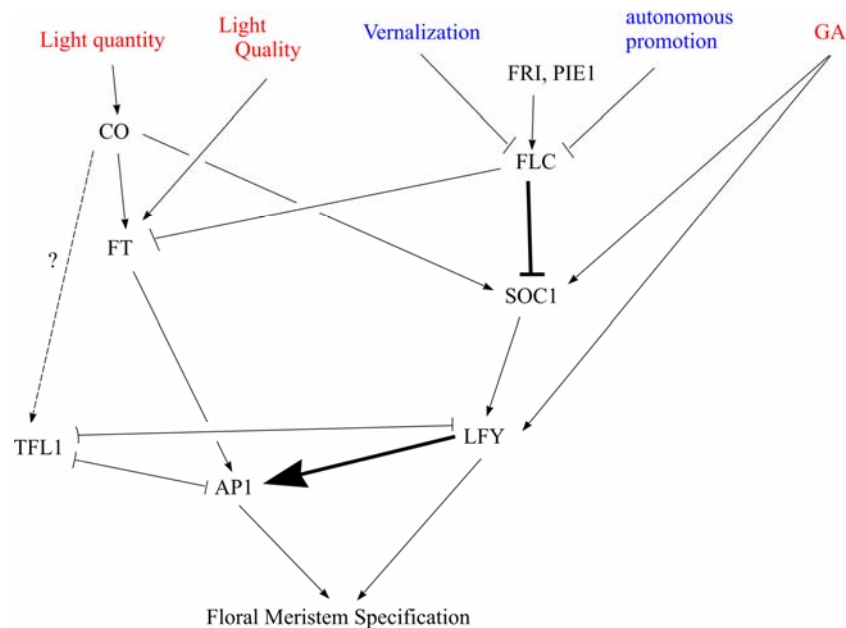


Figure 1-3: The floral integrators (adapted from Jack, 2004). Signals from the five major inductive floral pathways integrated by the genes; *FLC*, *SOC1*, *FT*, and *LFY*. Demonstrated direct interactions are bolded.

The *FRIGIDA (FRI)* gene causes an annual habit in *Arabidopsis*. That is, the *FRI* gene strongly promotes *FLC* levels thus keeping *Arabidopsis* vegetative over

winter. Once vernalization is complete *FLC* levels drop to allow flowering in spring (Napp-Zinn, 1961). *FRI* also appears to have many alleles that alter flowering time variation in *Arabidopsis* (Johanson et al., 2000). Other than *FRI*, genes that promote *FLC* have been identified through analysis of early flowering mutants (Boss et al., 2004). They include genes such as *EARLY FLOWERING SHORT DAYS (EFS)* (Soppe et al. 1999), *EARLY IN SHORT DAYS (ESD4)* (Reeves et al., 2002), the chromatin remodeling protein *PHOTOPERIOD INDEPENDANT EARLY FLOWERING1 (PIE1)* (Noh and Amasino, 2003), and *VERNALIZATION INDEPENDANCE (VIP)* (Zhang and van Nocker, 2002).

Other than vernalization, there are several other repressors of *FLC*. These have been identified in the *fca*, *fld*, *fpa*, *fve*, *fy*, *ld*, and *fld* mutants. These genes are termed as being in the autonomous pathway and are shown to reduce *FLC* mRNA accumulation (Michaels and Amasino, 1999; Sheldon et al., 1999; Sheldon et al., 2000; Michaels and Amasino, 2001).

1.2.1.2 Pathways that promote floral transition

The floral promotion pathways have been defined by the use of environmental and endogenous cues to promote the activation of the floral pathway integrators (Boss et al., 2004). Of these, photoperiod is the most well understood, and recent work has shown that there is a separate pathway that links light quality with ambient temperature (Blazquez et al., 2003; Cerdan and Chory, 2003).

The photoperiod pathway uses both red/far red light and UV/blue absorbing phytochromes to detect light (Lin, 2000; Quail, 2002). These phytochromes act on a number of entrainment genes (Boss et al, 2004) that keep the plant in synchronicity with its environment. These genes include *PHYTOCHROME-INTERACTING FACTOR 3 (EIF3)* that is bound by light activated *PHYTOCHROME (PHY) B* and in turn up regulates *CIRCADIAN CLOCK ASSOCIATED1 (CCA1)* and *LATE ELONGATED HYPOCOTYL (LHY)*. *CCA1*, *LEY*, *TIMING OF CAB*

EXPRESSION1 (TOC1), and *EARLY FLOWERING4 (ELF4)* are considered to be constituents of a central oscillator that forms the basis of the circadian system (Martinez-Garcia et al., 2000; Boss et al., 2004). The genes *EARLY FLOWERING3 (ELF3)* and *GIGANTEA (GI)* are thought to function between the photoreceptors and the oscillator and are thought to maintain the viability of the circadian system so that it generates correct amplitude and period (Boss et al., 2004).

It has been shown in a set of experiments that altered levels of circadian related genes produce changes in levels of the *CONSTANS (CO)* gene that is a key promoter of several floral pathway integrating genes (Putterill et al., 1995; Suarez-Lopez et al., 2001). Suarez-Lopez (2001) also showed that *CO* mRNA levels oscillated in long days with changing light levels (leading to an increase of *FT*).

It has been suggested by Simpson and Dean, (2002) that light quality also affects flowering time. This has been expressed in a ratio of red to far-red light, which can be reduced by shading. It has been shown that the photoreceptor *PHYB* down regulates the gene *PHYTOCHROME AND FLOWERING TIME1 (PFT1)* that activates *FT* (Cerdan and Chory, 2003). *PHYD* and *E* also have an effect but only if the effect of *PHYB* is absent (Halliday et al., 1994). It has also been shown that far-red and blue light are able to promote flowering via two independent pathways involving either *CRYPTOCHROME2 (CRY2)* or *PHYA* (Reed et al., 1994; Lin, 2000; Mockler et al., 2003). These two pathways form an antagonistic relationship with the *PFT* pathway and allow the plant to react to varying light qualities (Boss et al. 2004).

Gibberelic acid (GA) has been shown to have a dramatic effect on flowering as plants deficient in the hormone often have delays in flowering (Wilson et al., 1992). Currently no genes have been isolated that are on a GA pathway specific for flowering control (Jack, 2004), however among other suggestions, it seems most feasible that GA has an effect on *FT*. This has been shown by studies relating to the gene *EARLY BOLTING IN SHORT DAYS (EBS)* (Gomez-Mena et al., 2001; Pineiro et al., 2003).

Recent evidence also has shown that GA is able to up regulate the ABC genes *AP3*, *PI*, and *AG*, two fold (Yu et al., 2004).

1.2.2 The integrators

As indicated above, once all signals from the pathways that promote the floral transition and the pathways that enable floral transition are perceived by the plant they interact with the products of a few major genes that act to integrate these signals and allow the floral transition to commence. These major integrators have been identified as *LEAFY (LFY)*, *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (SOC1)*, *FLOWERING LOCUS C (FLC)*, and *FLOWERING LOCUS T (FT)*. Their integration of the flowering pathway is shown in Figure 1-3.

The MADS activator *FLOWERING LOCUS C (FLC)* is the main integrator for the pathways that enable the floral transition, e.g. the vernalization and autonomous pathways (Michaels and Amasino, 1999; Sheldon et al., 2000). It also integrates the positive signals from the *FRI* and *PIE1* genes (Johanson et al., 2000; Noh and Amasino, 2003). *FLC* functions to repress the floral activator *SOC1* (Lee et al, 2000; Hepworth et al., 2002). *SOC1* is in-turn activated by the long day photoperiod pathway (via *CO*) (Samach et al. 2000) and the GA pathway (Borner et al., 2000; Moon et al., 2003). *FLC* also has been shown to repress the floral integrator *FT*. It is important to note here that in short days GA is the only pathway that can activate *SOC1* (Moon et al, 2003).

FLOWERING TIME (FT) is down regulated by *FLC* until repressor signals from vernalization and signals from the autonomous promotion pathway are perceived and down regulate *FLC* (Sheldon et al., 2000; Michaels and Amasino, 2001). *FT* is also down regulated by *EBS* (Pineiro et al., 2003). *CO*, down regulated by long days, stimulates both *SOC1* and *FT* (Hepworth et al. 2002). There also has been some suggestion that *FT* is up regulated by *PFT1* which is stimulated by light quality (Cerdan

and Chory, 2003). Currently the nature of how *FT* integrates signals from *FLC*, *CO* and *EBS* is unknown. Thus, there could well be genes that occur mid-way through these pathways that are yet to be elucidated. *FT* functions to up regulate *APETALA1* (*API*), the first floral organogenesis gene.

LEAFY (*LFY*) further integrates all of the flower transition pathways by being up regulated by both the *SOC1* and GA pathways (Nilsson et al., 1998; Blazquez and Weigel, 2000). Of the transition pathways, only the GA pathway is proposed to have a direct effect on *LFY* (Blazquez and Weigel, 2000). Evidence for this has been through the discovery of a MYB protein (AtMYB33) that binds in vitro to an 8 bp sequence found in the *LFY* promoter region. This 8 bp region is conserved in the *LFY* promoter, and it occurs in *Populus* species as well (Brunner and Nilsson, 2004). However, it is currently unknown if AtMYB33 is required for the activation of *LFY* through GA. *LFY* is shown to interact in a positive fashion with the downstream gene *API*. Evidence for this has been through mutants either deficient or over expressing these genes. For example, plants deficient in *API* but transgenic for 35S:*LFY* constitutive expression lose the *LFY* floral promotion effect, while plants deficient in *LFY* transformed with the same construct do not. Moreover with 35S:*API* constructs floral organ identity is not correctly specified. This indicates that the spatial/terminal expression pattern of *LFY* is important for correct specification of organs and that the activity of *LFY* is separate from *API* (Weigel and Nilsson, 1995; Mandel and Yanofsky, 1995; Jack 2004).

The regulation of the gene *TFL1* is poorly understood (Jack, 2004) although it has been proposed to be somewhat positively regulated by the *CO* gene (Soltis et al., 2002). *TFL1* antagonizes *LFY* and *API* expression and can be found at low levels in the vegetative meristem and seems to play a role in preventing premature flowering (Bradley et al., 1997). The other notable thing about *TFL1* is that it is in the same gene family as *FT* with which it shares 50% homology (Kardailsky et al., 1999). Interestingly, the two genes have opposite effects on flowering timing with *ft* mutants being late flowering and *tfl1* mutants being early flowering (Mimida et al., 2001).

1.2.3 The floral identity genes and the ABC model

One of the important functions of the floral meristem identity genes (*LFY*, *SOCI*, *FLC*, and *FLT*) is to promote the expression of the ABC floral organ identity genes. In general, these genes are broken up into five major groups according to the organs that they help identify. These groups are called the A, B, C, and E groups and are summarized in Figure 1-4. Definitive evidence for these genes as being important for flower organ identity has been through the use of mutants that are deficient in one or more of the ABCE genes or *LFY*, (summarized in Thiessen et al., 2000; Jack, 2004; Ma, 2005). Through these mutant studies, it has been shown that A+E gene products are required for sepal identity, expression of A+B+E gene products are required for petal identity, B+C+E gene products are required for stamen identity, and C+E gene products are required for carpel identity (Ma, 2005).

The A group consists of the genes *APETALA1* (*API*) and *APETALA2* (*AP2*), and along with *SEPALATA1,2,3,4* and *LFY* are responsible for the identity of sepals in whorl 1 and petals in whorl 2 (petals arise from a combination of A and B class genes). Expression of *API* is initially found throughout the meristem but in later floral stages it is only found in whorls 1 and 2 (Mandel et al., 1992; Gustafson-Brown et al., 1994). *API* expression is promoted by the genes *LFY* and *FT* (Theissen et al., 2000; Jack, 2004). In *ap1* mutants there are partial conversions of flowers to shoots, and in *lfy+ap1* there is nearly complete reversion of flowers into shoots (Irish and Sussex, 1990; Mandel et al., 1992; Weigel et al., 1992; Weigel and Meyerowitz, 1993). The expression of *AP2* is found throughout the meristem during organogenesis although it only functions in whorls 1 and 2 and it is repressed in whorls 3 and 4 at the translation level by microRNA (miRNA) (Chen, 2004). In *ap2* mutant plants, sepals and petals are replaced by carpel-like structures due to ectopic expression of *AG* (Weigel and Meyerowitz, 1994).

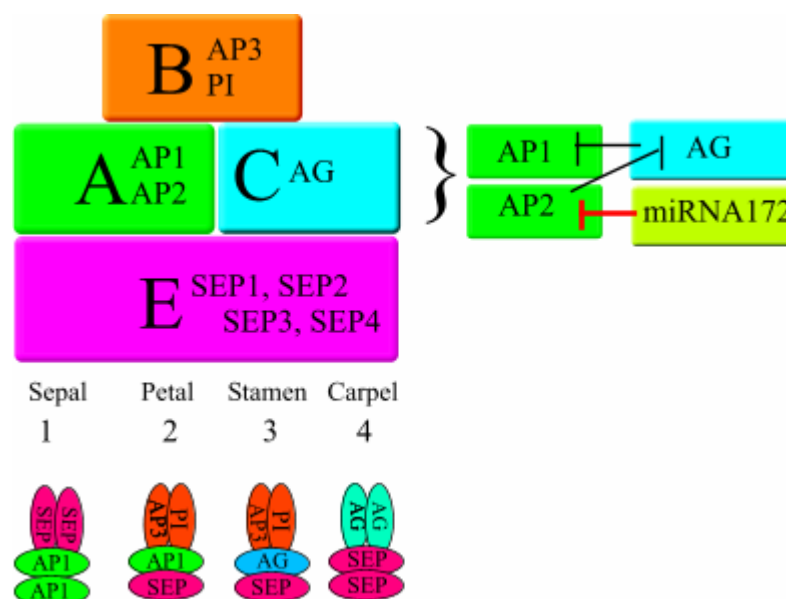


Figure 1-4: The ABC model of flower development (Adapted from Jack 2004, and Thiessen 2000). The top diagram represents the current ABCE model of flower development which states A genes promote pathways for sepal identity, A+B+E genes promote pathways for petal identity, B+C+E genes promote pathways for the identity of stamens, and C+E genes promote pathways for carpel identity. To the right of the model the mechanisms by which A and C genes are mutually repressive are depicted. The putative tetrads from the “quartet model” are also represented.

The B group consists of the genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) and is partially responsible for the identity of petals in whorl 2 in combination with the A class genes and the stamens and anthers in whorl 3 in combination with the C class genes. Both *ap3* and *pi* mutants share the same phenotype, that is, the petals are converted to sepals and the stamens are converted to carpels (Jack et al., 1992; Goto and Meyerowitz, 1994). Their expression is found only in whorls 2 and 3 (Weigel and Meyerowitz, 1993) and for proper expression the genes *LFY* and the F-box protein *UNUSUAL FLORAL ORGANS* (*UFO*) and *ARABIDOPSIS SKP1-LIKE1* (*ASK1*) are required. The absence of *LFY* reduces the expression of *AP3*, and *ufo* mutants tend to exhibit similar characteristics to *ap3* mutants in that they have a reduced number of petals

and stamens (Weigel and Meyerowitz, 1993; Levin and Meyerowitz, 1995; Wilkinson and Haughn, 1995). Finally, the *SUPERMAN (SUP)* genes negatively regulate *PI* and *AP3*. Mutants deficient in *SUP* exhibit expansion of *AP3* and *PI* expression into whorl 4 indicating its repressive role (Bowman et al., 1992).

UFO is expressed in different patterns over three stages of floral development suggesting that this temporal expression is due to discrete functions. The first is in the precursor cells of whorls 3 and 4 just before the meristem begins to form discrete organs (Ingram et al., 1995; Lee et al., 1997). This is thought to be important for the correct arrangement of floral organs. The second is in the precursors for petals and stamens as the sepal primordia emerge and is thought to be responsible for the activation of *AP3* (Levin and Meyerowitz, 1995; Wilkinson and Haughn, 1995). The third is as the petal and stamen primordia arise just below the petals and seems to be responsible for the development of these organs (Durfee et al., 2003). Furthermore it has been suggested that *UFO*, and *ASK1* (both AKP1-Cullin/CDC53-F-box-Protein (SCF) cooperate with *LFY* to activate *PI* and *AP3* expression (Lee et al., 1997; Zhao et al., 2001; Durfee et al., 2003).

The C class gene *AGAMOUS (AG)* is responsible for carpel identity and in combination with the B class genes, for anthers. *AG* also has the function of inhibiting the A class gene *API*, and *AP2* in turn represses *AG* in whorls 1 and 2 in *Arabidopsis* (Chen, 2004; Jack, 2004). The gene *LFY* has been shown to positively regulate *AG* by binding to a *cis* element of an *AG* intron. Further proof of this interaction is shown by high *LFY* expression being sufficient to activate *AG* in vegetative tissues (Busch et al., 1999). However, the presence of genes such as *CURLY LEAF (CLF)*, *LEUNIG (LUG)* and others that also function to repress *AG* in whorls 1 and 2 seem to indicate that this is a very complex system and requires further investigation. There are also some lines of investigation that indicate the genes *HEN* and *HUA* are involved in the processing of *AG* mRNA and thus lead to regulation and maintenance of B and C genes (Jack, 2004).

The genes *LFY* and *WUSHL* (*WUS*) activate *AG*. *WUS* binds to the same *AG* intron as *LFY* and is thought to act as a region specific activator and in combination with *LFY* initiates *AG* (Mayer et al., 1998; Parcy et al., 1998). This is comparable to the *LFY/UFO* activation of *AP3*. The authors denoted that after activation of *AG*, *AG* would in-turn down regulate *WUS* in whorls 3 and 4 to prevent loss in floral determinacy since *WUS* can cause the meristem to produce abundantly and sustain growth. It is unknown if this is direct or not (Lenhard et al., 2001; Lohmann et al., 2001).

The D class genes were initially identified in petunia and were attributed to determining placenta and ovule identity (Colombo et al., 1995; Theissen et al., 2000). The Arabidopsis ortholog of the genes found in petunia was originally called AG11 but has been renamed *SEEDSTICK* (*STK*) (Pinyopich et al., 2003). Along with the *SHATERPROOF* (*SHP*) 1 and 2 genes and *AG*, *STK* helps specify ovule development.

The E class genes were also first identified in petunia and are necessary for petal, stamen and carpel identity (Pelaz et al., 2000; Pelaz et al., 2001). In Arabidopsis the E class genes consist of four family members: *SEPALLATA* (*SEP*) 1, 2, 3, and 4 and triple mutations involving *SEP* 1-3 genes lead to all organs in the flowers representing sepal like organs (Pelaz et al., 2000). The original nomenclature for these genes was *AGL2* (*SEP1*), *AGL4* (*SEP2*), *AGL9* (*SEP3*), *AGL3* (*SEP4*). The important result of the discovery of these *SEP* genes is that the ABC model was revised to the “ABCE” model that states that sepal identity is by A and E, petals by A, B and E, stamens by B, C, and E, and carpels by C and E genes.

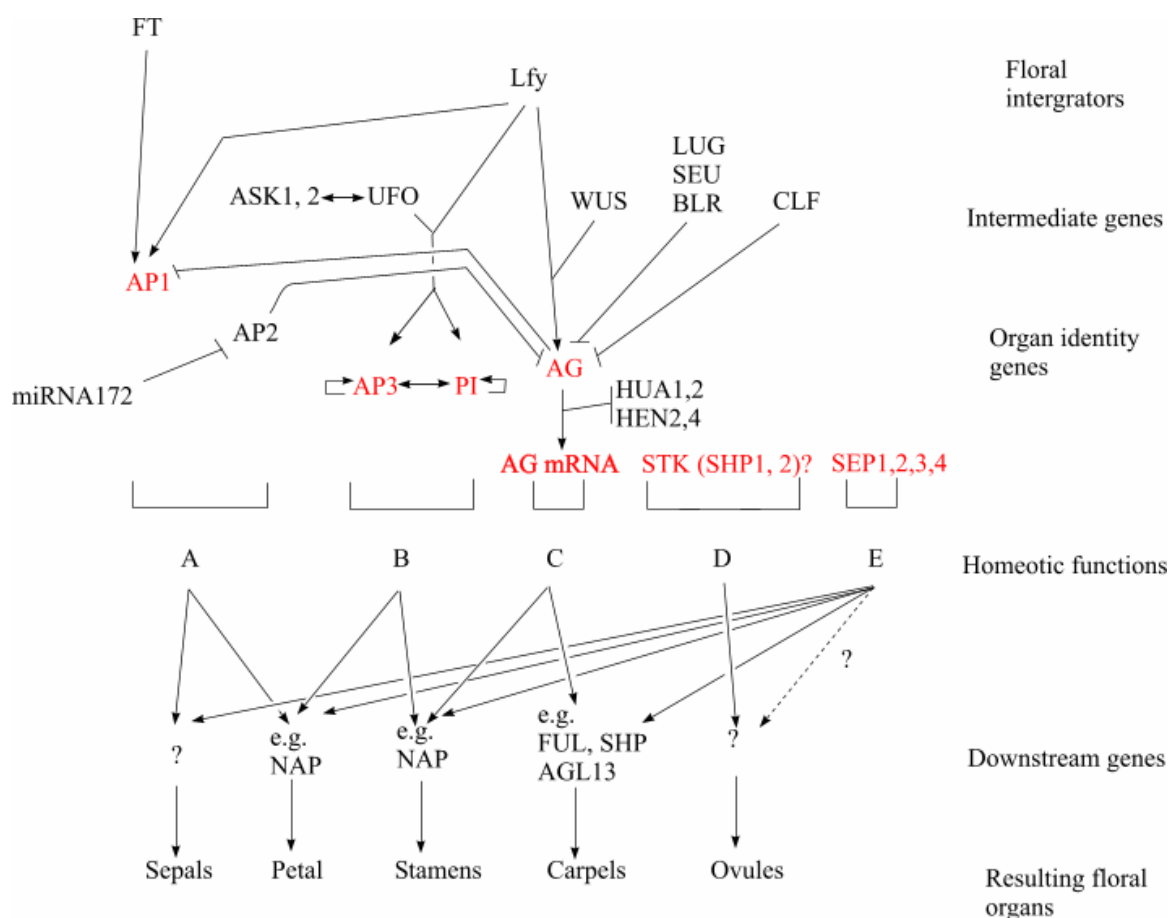


Figure 1-5: The pathways, post integrators, of the ABC model leading to floral organ initiation and differentiation. (MADS box genes are shown in red). Adapted from Theissen, (2001).

Most of the genes that are involved in the ABC model are known as MADS transcription factors. These include the genes: *AP1*, *AP3*, *PI*, *AG*, *STK*, *SHP1* and 2, and *SEP 1*, 2, 3, and 4. Some of the flowering time proteins and integrators are also MADS including: *FLC*, *SOC1*, *SVP*, *CAL*, and *FUL* (Jack, 2004). It has been estimated that there are more than 100 MADS genes in *Arabidopsis* and in general they are divided into two groups (Alvarez-Buylla et al., 2000; De Bodt et al., 2003). Most floral MADS proteins have the characteristic MIKC structure that contains (from 5' to 3'): a N-terminus, the MADS DNA binding/nuclear localization/dimerization domain, intervening

region, a K domain involved in mediating protein-protein interactions and dimerization functions, and the C domain for forming transcriptional activation and the formation of higher order complexes (McGonigle et al., 1996; Fan et al., 1997; Riechmann and Meyerowitz, 1997; Egea-Cortines et al., 1999; Moon et al., 1999; Honma and Goto, 2001; Immink et al., 2002; Yang et al., 2003). Type 1 MADS genes are similar but lack the K domain (Kohler et al., 2003).

It has been postulated that MADS gene products form tetramer complexes that control floral organ identity (Theissen and Saedler, 2001). These tetramers consist of two MADS dimers binding to a single MADS site with the tetramers forming by protein-protein interactions between MADS dimers (Theissen, 2001; Theissen and Saedler, 2001). Unfortunately, evidence for this model has come from *in vitro* experiments such as Yeast 2, 3 and 4 hybrid experiments and *in planta* evidence is required to verify the model *in vivo*. However, based on the results of these and other lines of investigation, predictions have been made as to the complexes that may occur. In short, sepals in whorl 1 may be specified by the tetramer of SEP/SEP-AP1/AP1, petals in whorl 2 may be specified by the tetramer of AP3/PI-SEP/AP1, the stamens of whorl 3 would be specified by AP3/PI-SEP/AG, and the carpels of whorl 4 would be specified by AG/AG-SEP/SEP (Theissen and Saedler, 2001; Jack, 2004; Ma, 2005).

1.2.4 Flowering in other species

As stated previously, the majority of the work in the molecular mechanisms of flowering has taken place in the species *Arabidopsis*, petunia, and snapdragon (Buzgo et al., 2004; Jack, 2004). However there is the continued integration of other species, for example tomato and maize, to create a clearer picture of a “universal” model of flower development if such exists. There is also continued work on the floral genome project “FGP” which aims to elucidate and characterize floral genes in many “missing-link” taxa (Soltis et al., 2002).

With differing signals required to initiate flower transition not to mention differing numbers of floral organs and morphologies it can seem that there will be little similarity among different species. However, some unifying principles such as the integration of signals through *LFY* and the basis of the ABC model should remain robust and thus allow us to use the information learned in species such as *Arabidopsis*, snapdragon and petunia as the starting point for studies into other species.

1.3 The Flowers of *Theobroma cacao* and their Importance

Considering that flowers are an essential component for the production of chocolate there is surprisingly very little literature pertaining to their growth in *T. cacao* (Bayer and Hoppe, 1990). Furthermore, there is no literature with any molecular information on flowering in *T. cacao*. The following few sections will describe what is currently known about the flowers of *T. cacao*.

1.3.1 Morphology

The flowers of *T. cacao* form on the trunk or main branches of the tree (cauliflorous) (Figure 1-6). The tissue that flowers develop from is of a certain minimum physiological age (usually 2-3 years), and eventually forms flower cushions that originate from leaf axils. The flower cushions thicken and consist of many (depending on the age of the flower pad) compressed cincinnal cymes. The flowers take a month to mature (Rajamony and Mohanakumarn, 1991; Smith, 1992) and develop from long pedicels usually 12-13 mm in length. The flowers have 5 free sepals, 5 free petals, 10 stamens and an ovary of five united carpels (Smith, 1992). The petals are narrow at the base but extend into cup-shaped pouches and are usually pink. The stamens are made of two whorls; the outer consists of five non-fertile staminodes and the inner of five fertile stamens. The stamens have two anthers that lie in the pouch of the corresponding petal. The style is twice as long as the ovary and consists of five parts around an axis (Wood

and Lass, 1985). The style has an appearance of being single but is divided at the tip into five stigmas (Wood and Lass, 1985; Smith, 1992) (Figure 1-6). Most of what has been published describes the fully formed, open *T. cacao* flower, however some research on flower development in *T. cacao* has been published.

1.3.2 Growth and development

The flowering of *T. cacao* has been used as the focus of several ecophysiology studies looking at soil moisture, humidity, temperature, and the energy cost of flowering (Sale, 1969, 1970, 1970; Valle et al., 1990). These studies have shown that *T. cacao* will flush with flowers when the environment changes from one with low air humidity to one with high air humidity (Sale, 1970). This is comparable to the observation that *T. cacao* will experience large increases in flower numbers (flushing) when soil conditions change from dry to wet (Sale, 1970). Similarly, continuous pod removal causes flowering rate to increase (Vale et al., 1990). Finally, it has been reported that most flowering occurs when the average night temperatures is about 80 °F (Sale, 1969).

Other research involving *T. cacao* flowers includes genetic studies of self-incompatibility, and hormone levels and their role in senescence. The self incompatibility phenomenon in *T. cacao* has been studied using genetics (Cope, 1962, 1962), that showed that a “s” locus is at least partly responsible for self incompatibility in *T. cacao* (Cope, 1962). The hormones abscisic acid and ethylene that are involved in the senescence of flowers have also been described (Aneja et al., 1999). Aneja et al. (1999) found that abscisic acid was critical for flower senescence in *T. cacao* and ethylene only played a supportive role; this is in contrast to other species where ethylene plays the major role (Aneja et al., 1999).

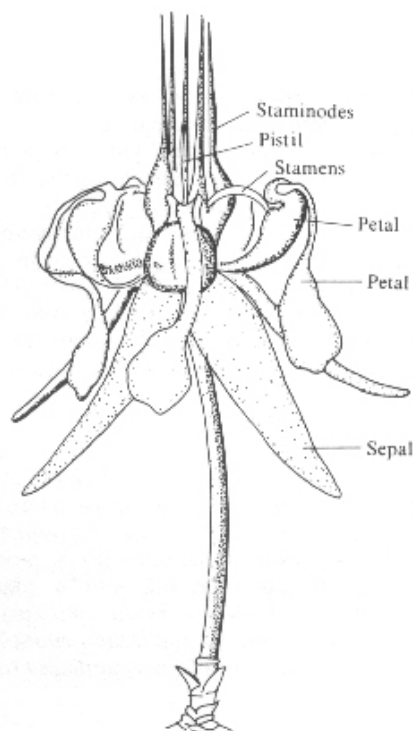


Figure 1-6: A cacao flower (Wood and Lass, 1985). Reprinted with kind permission from Blackwell Publishing, Oxford, United Kingdom.

In the study most relevant to this thesis, Bayer and Hoppe, (1990) described the order of events for flower development in *T. cacao* at a morphological level. They described floral growth in *T. cacao* by looking at a series of electron micrographs; however, no quantitative data were collected. These results are discussed in detail in relation to our data in Chapter 2.

Although it has been estimated that flower development in *T. cacao* takes between 21 to 30 days (Bayer and Hoppe, 1990; Rajamony and Mohanakumarn, 1991; Smith, 1992), there have been no studies carried out to characterize the stages of *T. cacao* flower development in the detail that they have been characterized for model systems such as *Arabidopsis* (Smyth et al., 1990), nor have there been studies on the molecular

mechanisms of flower development in *T. cacao*. One intention of the work contained herein is to characterize “normal” flower developmental morphology.

1.3.3 The effect of Witches’ Broom on *T. cacao* flower development

Crinipellis pernicioso (Witches’ Broom) infects flower cushions, as well as vegetative meristems. Witches’ broom has been a major pathogen in Brazil accounting for up to 60% of crop losses over the past 10 years (Purseglove, 1968; Smith, 1992; Pereira et al., 1997). The fungus primarily affects the meristematic tissues of the plant causing proliferation of branches and hypertrophied shoots, producing a “broom-like” effect. It also causes meristematic flower cushions to produce vegetative brooms (Purseglove, 1968). First reported by Van Hall and Drost, (1909), *C. pernicioso* can infect single flowers to entire cushions forming what are called “starblooms” where the pedicles of the flower often become abnormally thick, large, and stand erect due to added rigidity (Stahel, 1932; Baker and McKee, 1943) (Figure 1-7). The fungus can cause the cushion to form vegetative stems that can range in length from 1 cm to 54 cm, and can in turn produce auxiliary flowers. Infected flowers tend not to fall off the plant if they remain unfertilized. Occasionally fertilized infected flowers produce small “indirectly infected pods” which tend to be distorted into a spherical or “strawberry” shape (Baker and McKee, 1943). Infected parts of cushions and flowers can be removed from the tree and the cushion can remain healthy and continue to produce flowers throughout the life of the tree (Baker and McKee, 1943). The disease eventually infects the fruit making them hard and woody (Purseglove, 1968).

Most molecular research relating to the response of *T. cacao* to this pathogen has been based on the leaf, an organ that is not really involved with the disease (Verica et al., 2004). There has been no published literature to date exploring the mechanisms with which *C. pernicioso* is able to infect and modify the plant floral meristems causing reversion to vegetative meristems.



Figure 1-7: *T. cacao* flowers infected with *C. perniciosia* (photograph taken by M. J. Gultinan).

Floral meristem reversion to vegetative development has been described in other plant species such as *Arabidopsis*, *Triticum aestivum* (wheat), *Impatiens balsamina*, and *Anagallis* (Battey and Lyndon, 1990). These reversions can be induced in *Arabidopsis* by manipulating photoperiod in *ag* mutants and *lfy* heterozygote mutants (Okamuro et al., 1993; Okamuro et al., 1996; Mizukami and Ma, 1997). The authors also demonstrated in these studies that floral reversion could be prevented in *spindly* mutants, (*spindly* mutants show an activation of gibberellin signal transduction pathways), or by the addition of exogenous gibberellins, or *hyl* mutants (*hyl* causes a blockage in phytochrome activity). Therefore, it was proposed that *LFY* and *AG* play an important role in the maintenance of flower meristem identity and that floral meristem reversion in *ag* and *lfy* mutants is controlled by a phytochrome and gibberellin signal cascade. These

findings provide interesting lines of research to investigate the mechanism that *C. perniciosus* may use to cause reversion of flowers to shoots in *T. cacao*.

1.4 Relevance

Much of our understanding about flowering mechanisms has been derived from model species such as *Arabidopsis*, petunia, and snapdragon. However, the understanding of flowering in many other species ranges from detailed to nothing. Prior to this work, there was little developmental information and no molecular information on the flowering mechanisms of *T. cacao*. We have accurately described floral growth in *T. cacao* on three levels. The first was the morphological level where we accurately described the formation, growth, and senescence of *T. cacao* flowers using time-lapse photography and microscopy techniques. At the genomic level, we described the DNA sequence of fifty genes including the *T. cacao LFY* gene and compared them to their *Arabidopsis* ortholog. At the level of expression, we examined the flowering genes *LFY*, *API*, *AP3*, *PI*, and *AG*. The combination of data sets affords a good picture of “normal” floral development in *T. cacao* and sets the scene for future research, such as into the reversion to the vegetative floral state in *T. cacao* flowers infected with *C. perniciosus*.

The second aim of this thesis was to compare *T. cacao* floral development to the substantial amount of data available for *Arabidopsis* to assess the degree of evolutionary conservation between the two species. In short, we tested the hypothesis stating that due to the evolutionary relatedness of the two species, the mechanisms and developmental pathways controlling flower development in *Arabidopsis* and *T. cacao* may be highly conserved. If this hypothesis is correct, at the molecular level, many of the genes and their expression patterns should be very similar and functionally equivalent.

In order to assess if it was possible to compare any molecular data between *T. cacao* to *Arabidopsis* it was decided to compare at the DNA level first. One

other study had been carried out comparing the chloroplast data of the Byttnerioideae (Whitlock et al., 2001). Using these data, we compared some chloroplast genes of *Arabidopsis* to *T. cacao* and a high homology was observed (Chapter 3). This gave some support to the hypothesis that *Arabidopsis* may be a good model for comparison.

If this hypothesis is supported, research in *T. cacao* can be greatly facilitated using discoveries in the model system *Arabidopsis*. Thus, by being able to apply directly *Arabidopsis* research we will concurrently accelerate the understanding of *T. cacao* flower development. Furthermore, it can be postulated that systems other than floral development could also be conserved between the two species. This evolution-development (“evo-devo”) type approach is invaluable to lend insight to the lesser-studied species *T. cacao*. However, it could stand to reason that if the pathways are conserved between *T. cacao* and *Arabidopsis* that they could even be extended to other more distantly related species. This allows the approach taken in this thesis to be adopted as a model by other researchers as a method to compare their organisms to their respective model species.

1.5 References

- Alvarez-Buylla ER, Pelaz S, Liljegren SJ, Gold SE, Burgeff C, Ditta GS, de Poupiana LR, Martinez-Castilla L, Yanofsky MF** (2000) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 5328-5333
- Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ** (2000) Molecular and genetic analyses of the *SILKY1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell* **5**: 569-579
- Aneja M, Gianfagna T, Ng E** (1999) The roles of abscisic acid and ethylene in the abscission and senescence of cocoa flowers. *Plant Growth Regulation* **27**: 149-155
- Angenot GC, Colombo L** (1996) Molecular control of ovule development. *Trends in Plant Sciences* **1**: 228-232
- Baker RED, McKee RK** (1943) Witches' broom disease investigations: IV. The infection of flower cushions and pods of cacao by *Marasmius pernicius* Stahel. *Tropical Agriculture* **20**: 188-194
- Bartolome R** (1951) Cacao. *The Philippine Journal of Agriculture* **16**: 1-53
- Batley NH, Lyndon RF** (1990) Reversion of flowering. *Botanical Review* **56**: 162-189
- Bayer C, Hoppe JR** (1990) Die Blütenentwicklung von *Theobroma cacao* L. (Sterculiaceae). *Beitr. Biol. Pflanzen* **65**: 301-312
- Bennett AB** (2003) Out of the Amazon: *Theobroma cacao* enters the genomic era. *Trends in Plant Science* **18**: 561-563
- Blazquez MA, Ahn JH, Weigel D** (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nature Genetics* **33**: 168-171
- Blazquez MA, Soowal LN, Lee I, Weigel D** (1997) *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**: 3835-3844
- Blazquez MA, Weigel D** (2000) Integration of floral inductive signals in *Arabidopsis*. *Nature* **404**: 889-892
- Boss PK, Bastow RM, Mylne JS, Dean C** (2004) Multiple pathways in the decision to flower: Enabling, promoting, and resetting. *Plant Cell* **16**: S18-S31
- Bowman JL, Sakai H, Jack T, Weigel D, Mayer U, Meyerowitz EM** (1992) *SUPERMAN*, a regulator of floral homeotic genes in *Arabidopsis*. *Development* **114**: 599-615
- Bowman JL, Smyth DR, Meyerowitz EM** (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1-20
- Bradley D, Ratcliffe O, Vincent C, Carpenter R, Coen E** (1997) Inflorescence commitment and architecture in *Arabidopsis*. *Science* **275**: 80-83

- Brunner AM, Nilsson O** (2004) Revisiting tree maturation and floral initiation in the poplar functional genomics era. *New Phytologist* **164**: 43-51
- Busch MA, Bomblies K, Weigel D** (1999) Activation of a floral homeotic gene in *Arabidopsis*. *Science* **285**: 585-587
- Buzgo M, Soltis DE, Soltis PS, Ma H** (2004) Towards a comprehensive integration of morphological and genetic studies of floral development. *Trends in Plant Science* **9**: 164-173
- Cerdan PD, Chory J** (2003) Regulation of flowering time by light quality. *Nature* **423**: 881-885
- Chen XM** (2004) A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* **303**: 2022-2025
- Clement D, Lanaud C, Sabau X, Fouet O, Le Cunff L, Ruiz E, Risterucci AM, Glaszmann JC, Piffanelli P** (2004) Creation of BAC genomic resources for cocoa (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones. *Theoretical and Applied Genetics* **108**: 1627-1634
- Coen ES, Meyerowitz EM** (1991) The war of the whorls - Genetic interactions controlling flower development. *Nature* **353**: 31-37
- Colombo L, Franken J, Koetje E, Vanwent J, Dons HJM, Angenent GC, Vantunen AJ** (1995) The Petunia Mads Box gene Fbp11 determines ovule identity. *Plant Cell* **7**: 1859-1868
- Cope FW** (1962) The effects of incompatibility and compatibility on genotype proportions in populations of *Theobroma cacao* L. *Heredity* **17**: 183-195
- Cope FW** (1962) The mechanism of pollen incompatibility in *Theobroma cacao* L. *Heredity* **17**: 157-182
- De Bodt S, Raes J, Florquin K, Rombauts S, Rouze P, Theissen G, Van de Peer Y** (2003) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *Journal Of Molecular Evolution* **56**: 573-586
- Duguma B, Gockowshi J, Bakala J** (2001) Smallholder Cacao (*Theobroma cacao* L.) clutivation in agroforestry systems of West and Central Africa: Challenges and oppertunities. *Agroforestry Systems* **51**: 177-188
- Durfee T, Roe JL, Sessions RA, Inouye C, Serikawa K, Feldmann KA, Weigel D, Zambryski PC** (2003) The F-box-containing protein UFO and AGAMOUS participate in antagonistic pathways governing early petal development in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 8571-8576
- Egea-Cortines M, Saedler H, Sommer H** (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *Embo Journal* **18**: 5370-5379
- Fan HY, Hu Y, Tudor M, Ma H** (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant Journal* **12**: 999-1010
- Flament MH, Kebe I, Clement D, Pieretti I, Risterucci AM, N'Goran JA, Cilas C, Despreaux D, Lanaud C** (2001) Genetic mapping of resistance factors to *Phytophthora palmivora* in cocoa. *Genome* **44**: 79-85

- Fornara F, Marziani G, Mizzi L, Kater M, Colombo L** (2003) MADS-box genes controlling flower development in rice. *Plant Biology* **5**: 16-22
- Gocal GFW, King RW, Blundell CA, Schwartz OM, Andersen CH, Weigel D** (2001) Evolution of floral meristem identity genes. Analysis of *Lolium temulentum* genes related to *APETALA1* and *LEAFY* of *Arabidopsis*. *Plant Physiology* **125**: 1788-1801
- Gomez-Mena C, Pineiro M, Franco-Zorrilla JM, Salinas J, Coupland G, Martinez-Zapater JM** (2001) *early bolting in short days*: An *Arabidopsis* mutation that causes early flowering and partially suppresses the floral phenotype of *leafy*. *Plant Cell* **13**: 1011-1024
- Goto K, Meyerowitz EM** (1994) Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Development* **8**: 1548-1560
- Gustafson-Brown C, Savidge B, Yanofsky MF** (1994) Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* **76**: 131-143
- Halliday KJ, Koornneef M, Whitelam GC** (1994) Phytochrome B and at least one other phytochrome mediate the accelerated flowering response of *Arabidopsis thaliana* L to low red/far-red ratio. *Plant Physiology* **104**: 1311-1315
- Honma T, Goto K** (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**: 525-529
- Hurst WJ, Tarka SM, Jr., Powis TG, Valdez F, Jr., Hester TR** (2002) Cacao usage by the earliest Maya civilization. *Nature* **418**: 289-290
- Immink RGH, Gadella TWJ, Ferrario S, Busscher M, Angenent GC** (2002) Analysis of MADS box protein-protein interactions in living plant cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2416-2421
- Ingram GC, Goodrich J, Wilkinson MD, Simon R, Haughn GW, Coen ES** (1995) Parallels between *UNUSUAL FLORAL ORGANS* and *FIMBRIATA*, genes-controlling flower development in *Arabidopsis* and *Antirrhinum*. *Plant Cell* **7**: 1501-1510
- Irish VF, Sussex IM** (1990) Function of the *APETALA-1* gene during *Arabidopsis* floral development. *Plant Cell* **2**: 741-753
- Jack T** (2004) Molecular and genetic mechanisms of floral control. *Plant Cell* **16**: S1-S17
- Jack T, Brockman LL, Meyerowitz EM** (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* **68**: 683-697
- Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C** (2000) Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**: 344-347
- Jones PG, Allaway D, Gilmore DM, Harris C, Rankin D, Retzel ER, Jones CA** (2002) Gene discovery and microarray analysis of cacao (*Theobroma cacao* L.) varieties. *Planta* **216**: 255-264
- Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D** (1999) Activation tagging of the floral inducer FT. *Science* **286**: 1962-1965

- Kohler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U** (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes & Development* **17**: 1540-1553
- Lee I, Wolfe DS, Nilsson O, Weigel D** (1997) A *LEAFY* co-regulator encoded by *UNUSUAL FLORAL ORGANS*. *Current Biology* **7**: 95-104
- Lenhard M, Bohnert A, Jurgens G, Laux T** (2001) Termination of stem cell maintenance in Arabidopsis floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell* **105**: 805-814
- Levin JZ, Meyerowitz EM** (1995) *UFO* - An *Arabidopsis* gene involved in both floral meristem and floral organ development. *Plant Cell* **7**: 529-548
- Lin CT** (2000) Photoreceptors and regulation of flowering time. *Plant Physiology* **123**: 39-50
- Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D** (2001) A molecular link between stem cell regulation and floral patterning in Arabidopsis. *Cell* **105**: 793-803
- Lopez R** (2002) *Chocolate: The nature of indulgence*, Ed 1. Harry N. Abrams, Inc., New York
- Ma H** (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual Reviews in Plant Biology* **56**: 393-434
- Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF** (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**: 273-277
- Martinez-Garcia JF, Huq E, Quail PH** (2000) Direct targeting of light signals to a promoter element-bound transcription factor. *Science* **288**: 859-863
- Maximova S, Miller C, Antunez de Mayolo G, Pishak S, Young A, Guiltinan MJ** (2003) Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. *Plant Cell Reports* **21**: 872-883
- Mayer KFX, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T** (1998) Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. *Cell* **95**: 805-815
- McGonigle B, Bouhidel K, Irish VF** (1996) Nuclear localization of the *Arabidopsis* *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes & Development* **10**: 1812-1821
- Meyerowitz EM, Running MP, Sakai H, Williams RW** (1998) Multiple modes of cell division control in *Arabidopsis* flower development. *Symp Soc Exp Biol* **51**: 19-26
- Michaels SD, Amasino RM** (1999) *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**: 949-956
- Michaels SD, Amasino RM** (2001) Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* **13**: 935-941
- Mimida N, Goto K, Kobayashi Y, Araki T, Ahn JH, Weigel D, Murata M, Motoyoshi F, Sakamoto W** (2001) Functional divergence of the *TFL1*-like gene

- family in *Arabidopsis* revealed by characterization of a novel homologue. *Genes to Cells* **6**: 327-336
- Mizukami Y, Ma H** (1997) Determination of *Arabidopsis* Floral Meristem Identity by AGAMOUS. *The Plant Cell* **9**: 393-408
- Mockler T, Yang HY, Yu XH, Parikh D, Cheng YC, Dolan S, Lin CT** (2003) Regulation of photoperiodic flowering by *Arabidopsis* photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 2140-2145
- Moon YH, Kang HG, Jung JY, Jeon JS, Sung SK, An G** (1999) Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. *Plant Physiology* **120**: 1193-1203
- Morais RC** (2005) The gnomes of cocoa. *Forbes Magazine* (11 April)
- Napp-Zinn K** (1961) Über die Bedeutung genetischer Untersuchungen an kaltebedürftigen Pflanzen für die Aufklärung von Vernalisationserscheinungen. *Züchter* **31**: 128-135
- Nilsson O, Lee I, Blazquez MA, Weigel D** (1998) Flowering-time genes modulate the response to *LEAFY* activity. *Genetics* **150**: 403-410
- Noh YS, Amasino RM** (2003) *PIE1*, an ISWI family gene, is required for *FLC* activation and floral repression in *Arabidopsis*. *Plant Cell* **15**: 1671-1682
- Okamuro JK, Den Boer BGW, Jofuku KD** (1993) Regulation of *Arabidopsis* flower development. *The Plant Cell* **5**: 1183-1193
- Okamuro JK, den Boer BGW, Lotys-Prass C, Szeto W, Jofuku KD** (1996) Flowers into shoots: Photo and hormonal control of a meristem identity switch in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 13831-13836
- Parcy F, Nilsson O, Busch MA, Lee I, Weigel D** (1998) A genetic framework for floral patterning. *Nature* **395**: 561-566
- 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
- Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF** (2001) APETALA1 and SEPALLATA3 interact to promote flower development. *Plant Journal* **26**: 385-394
- Pereira JL, Almeida LCC, Santos SM** (1997) Witches' broom disease of cocoa in Bahia, Brazil: Strategy in tentative eradication and containment. *In Proceedings of the first international cocoa pests and diseases seminar*, Cocoa Research Institute of Ghana: Tafo, Ghana, pp 139-166
- Pineiro M, Gomez-Mena C, Schaffer R, Martinez-Zapater JM, Coupland G** (2003) EARLY BOLTING IN SHORT DAYS is related to chromatin remodeling factors and regulates flowering in *Arabidopsis* by repressing FT. *Plant Cell* **15**: 1552-1562
- 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

- Pugh T, Fouet O, Risterucci AM, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N'Goran JA, Lanaud C** (2004) A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. *Theoretical and Applied Genetics* **108**: 1151-1161
- Purseglove JW** (1968) *Tropical crops. Dicotyledons*. John Wiley and Sons Inc., New York
- Putterill J, Robson F, Lee K, Simon R, Coupland G** (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc-finger transcription factors. *Cell* **80**: 847-857
- Quail PH** (2002) Phytochrome photosensory signalling networks. *Nature Reviews Molecular Cell Biology* **3**: 85-93
- Rajamony L, Mohanakumarn N** (1991) A note on the floral biology of cacao (*Theobroma cacao L.*). *South Indian Horticulture* **39**: 168
- Ramirez OA, Somarriba E, Ludewigs T** (2001) Financial returns, stability and risk of cacao-plantain-timber agroforestry systems in central America. *Agroforest Systems* **51**: 141-151
- Reed JW, Nagatani A, Elich TD, Fagan M, Chory J** (1994) Phytochrome-A and Phytochrome-B have overlapping but distinct functions in *Arabidopsis* development. *Plant Physiology* **104**: 1139-1149
- Reeves PH, Coupland G** (2001) Analysis of flowering time control in *Arabidopsis* by comparison of double and triple mutants. *Plant Physiology* **126**: 1085-1091
- Reeves PH, Murtas G, Dash S, Coupland G** (2002) early in short days 4, a mutation in *Arabidopsis* that causes early flowering and reduces the mRNA abundance of the floral repressor FLC. *Development* **129**: 5349-5361
- Riechmann JL, Meyerowitz EM** (1997) MADS domain proteins in plant development. *Biological Chemistry* **378**: 1079-1101
- Sale PJM** (1969) Flowering of cacao under controlled temperature conditions. *Journal of Horticulture Science* **44**
- Sale PJM** (1970) Growth and flowering of cacao under controlled relative humidities. *Journal of Horticulture Science* **45**: 119-132
- Sale PJM** (1970) Growth, flowering and fruiting of cacao under controlled soil moisture conditions. *Journal of Horticulture Science* **45**: 99-118
- Saunders JA, Mischke S, Leamy EA, Hemeida AA** (2004) Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theoretical And Applied Genetics* **110**: 41-47
- Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES** (1999) The *FLF* MADS box gene: A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**: 445-458
- Sheldon CC, Rouse DT, Finnegan EJ, Peacock WJ, Dennis ES** (2000) The molecular basis of vernalization: The central role of *FLOWERING LOCUS C (FLC)*. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 3753-3758
- Simpson GG, Dean C** (2002) Flowering - *Arabidopsis*, the rosetta stone of flowering time? *Science* **296**: 285-289

- Smith NJH** (1992) Tropical forests and their crops. Comstock Pub. Associates, Ithaca, N.Y.
- Soltis DE, Soltis PS, Albert VA, Oppenheimer DG, dePamphilis CW, Ma H, Frohlich MW, Theissen G** (2002) Missing links: the genetic architecture of flower and floral diversification. *Trends in Plant Science* **7**: 22-31
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS** (2000) Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Botanical Journal of the Linnean Society* **133**: 381-461
- Stahel G** (1932) Contribution to the knowledge of Witchbroom disease. *Tropical Agriculture* **IX**: 167-176
- Suarez-Lopez P, Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G** (2001) CONSTANS mediates between the circadian clock and the control of flowering in *Arabidopsis*. *Nature* **410**: 1116-1120
- Theissen G** (2001) Development of floral organ identity: stories from the MADS house. *Current Opinion In Plant Biology* **4**: 75-85
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H** (2000) A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149
- Theissen G, Saedler H** (2001) Plant biology - Floral quartets. *Nature* **409**: 469-471
- Valle RR, De Almedia A-AF, De RM, Leite O** (1990) Energy costs of flowering, fruiting, and cherelle wilt in cacao. *Tree Physiology* **6**: 329-336
- Van Hall CJJ, Drost AW** (1909) Heksenbezemziekte can cacao. Surinam Dep. of Argiculture, Bull. **16**
- Verica JA, Maximova SN, Strem MD, Carlson JE, Bailey BA, Gultinan MJ** (2004) Isolation of ESTs from cacao (*Theobroma cacao* L.) leaves treated with inducers of the defense response. *Plant Cell Reports* **23**: 404-413
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM** (1992) LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843-859
- Weigel D, Meyerowitz EM** (1993) Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**: 1723-1726
- Weigel D, Meyerowitz EM** (1994) The ABCs of floral homeotic genes. *Cell* **78**: 203-209
- Whitlock BA, Bayer C, Baum DA** (2001) Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) based on sequences of the chloroplast gene, ndhF. *Systematic Botany* **26**: 420-437
- Wilkinson MD, Haughn GW** (1995) UNUSUAL FLORAL ORGANS controls meristem identity and organ primordia fate in *Arabidopsis*. *Plant Cell* **7**: 1485-1499
- Wilson RN, Heckman JW, Somerville CR** (1992) Gibberellin is required for flowering in *Arabidopsis thaliana* under short days. *Plant Physiology* **100**: 403-408
- Wood GAR, Lass RA** (1985) Cocoa, Ed 4. Longman Scientific and Technical, New York
- Yang YZ, Xiang HJ, Jack T** (2003) *pistillata-5*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. *Plant Journal* **33**: 177-188

- Yu H, Ito T, Zhao Y, Peng J, Kumar P, Meyerowitz EM** (2004) Floral homeotic genes are targets of gibberellin signaling in flower development. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 7827-7832
- Zhang H, van Nocker S** (2002) The *VERNALIZATION INDEPENDENCE 4* gene encodes a novel regulator of FLOWERING LOCUS C. *Plant Journal* **31**: 663-673
- Zhao D, Yu Q, Chen XM, Ma H** (2001) The ASK1 gene regulates B function gene expression in cooperation with UFO and LEAFY in *Arabidopsis* anther. *Development* **128**: 2735-2746

Chapter 2

Morphology of *T. cacao* Flowers and Comparison to Other Species



Water

2.1 Introduction

The understanding of molecular mechanisms associated with the floral biology of angiosperms is expanding rapidly (for reviews please see Jack, (2001); Boss et al., (2004); Jack, (2004); Ma, (2005)). This has been primarily achieved through the identification of developmental mutants, the subsequent characterization of genes associated with these mutations, and the pathways they control in several model plant species (reviewed in Chapter 1). It is now clear that flowers arise from a spatially and temporally controlled cascade of gene regulation, the products of which interact to determine the fate of cellular differentiation in various floral organs. This cascade has been identified as the “ABC” model of flower development (Meyerowitz et al., 1998; Theissen et al., 2000; Jack, 2001; Kanno et al., 2003; Jack, 2004). Deciphering such developmental control mechanisms has implications not only for our basic understanding of development in general, but may also contribute to applications of plant biotechnology in the future.

The use of molecular tools to investigate the complex pathways of floral development first requires a detailed characterization of the morphological development of the species in question. For any yet unstudied species, this will involve building a base of morphological and cytological definitions of distinct key morphological events during the development and growth of the flower. Some key morphological landmarks and common terminology have been suggested by Buzgo et al., (2004). They studied three species occupying key phylogenetic positions, and defined the following morphological landmarks: inflorescence formation, initiation of outermost perianth organs, initiation of inner perianth organs, first stamens, carpel initiation, microsporangia, ovule initiation 1 and 2, male and female meiosis and flower opening. Morphologically and temporally defining these landmarks in a given plant species will greatly facilitate the comparison to better-characterized model species and facilitate the transference of mechanistic knowledge gained in model plant systems.

Theobroma cacao L. is an understory tree endemic to the Amazon rainforest. The seeds of *T. cacao*, cocoa beans, are processed to create the main ingredient for chocolate (Bartolome 1951; Wood and Lass, 1985). The flowers of *T. cacao* are cauliflorous, forming directly from the trunk of the mature tree. The switch in development from juvenile to adult stages (phase change), occurs suddenly in *T. cacao* when the tree reaches approximately 1.5 – 2 meters in height (~1.5 years), and is marked by the formation of a four or five branched jorquette of near horizontally growing branches (Bartolome, 1951). Shortly after this change, flowers form on the trunk and main branches of the plant. The initial flowers appear to arise above leaf scars, in the axils of abscised leaf. Inflorescences subsequently arise continuously from the same spot and eventually flower cushions form consisting of many compressed cincinnal cymes. The flowers take approximately one month to mature (Rajamony and Mohanakumarn, 1991; Smith, 1992) and are subtended by long pedicles. The flowers have 5 free sepals, 5 free petals, 10 stamens (5 fertile and 5 infertile staminodes) and an ovary of five united carpels (Smith, 1992). The petals are narrow at the base but extend into cup-shaped pouches and are usually pink and white, the color indicative of a given genotype. The stamens are arranged in two whorls; the outer consists of five non-fertile staminodes and the inner of five fertile stamens. The stamens have two anthers that lie in the pouch of the corresponding petal. The style is twice as long as the ovary and consists of five parts around an axis (Wood and Lass, 1985). The style, being fused, has the appearance of being single but is divided at the tip into five stigmas (Wood and Lass, 1985; Smith, 1992). Much of this structural information is summarized in both Figures 1-6 and 2-1.

Bayer and Hoppe, (1990) described the order of events for flower development in *T. cacao* by looking at a series of electron micrographs, however no quantitative data were collected. Although it has been estimated that flower development in *T. cacao* takes between 21 to 30 days (Rajamony and Mohanakumarn, 1991) there have been no studies carried out to characterize the stages of *T. cacao* flower development in the detail that they have been characterized for model systems such as *Arabidopsis* (Smyth et al., 1990),

nor have there been any studies regarding the molecular mechanisms of flower development in *T. cacao*.

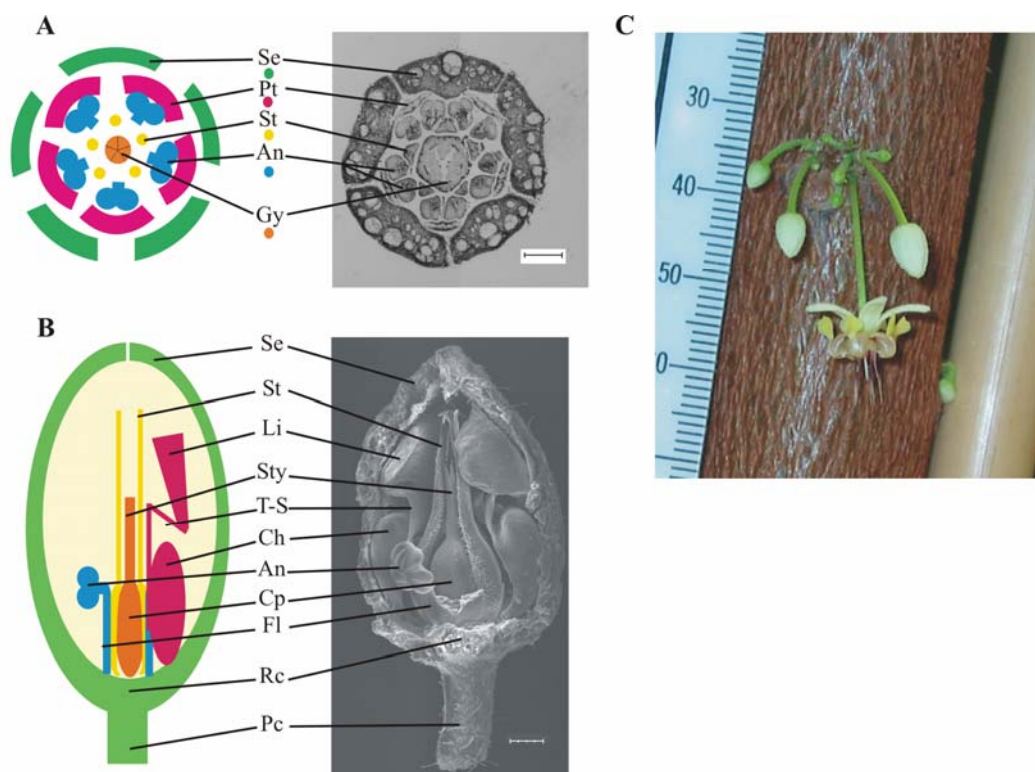


Figure 2-1: Floral anatomy of *T. cacao*. Images of transversely (A) and longitudinally (B) sectioned 22 day old flower buds: A, Transverse view approximately two thirds of the way to the receptacle of the *T. cacao* flower clearly showing four whorls. The calyx consists of sepals (Se), the corolla consists of petals (Pt), the androecium (Ad) consists of anthers (An) and filaments (stamens), and non-fertile staminodes (St), the gynoecium (Gy) consists of the carpel containing stigma, style and ovary. B, A longitudinal view of the *T. cacao* flower showing the following organs: Sepal (Se), staminode (St), petal; consisting of the ligula (Li) taenial-segment (T-S), and concha (Ch). Style (Sty), carpel (Cp), Anther (An), filament (Fl), receptacle (Rc), and the Pedicel (Pc). The electron micrograph is a 22 day old bud. C, Photograph of a *T. cacao* cyme with several flowers in differing stages of development (ruler shown with 1mm divisions). Bars=200 μ m in A and B.

The research described in this chapter had two distinct aims: the first was to determine precisely the timing of key morphological events in the development of *T.*

cacao flowers. The second was to compare floral development in *T. cacao* to that of other model systems, especially *Arabidopsis*, the species most closely related to *T. cacao* in which flower development has been studied in depth. Such comparisons should reveal the similarities and differences between the developmental programs of *T. cacao* and other plants and contribute to the understanding of the mechanisms of floral evolution. This information can subsequently be used in further molecular studies on the mechanisms which regulate the development of *T. cacao* flowers.

2.2 Methods

2.2.1 Plant materials and growth conditions

Ramets of the *Theobroma cacao* PSU-“SCA6” clone were used for all aspects of these experiments. The clone was sent to the Pennsylvania State University from the USDA Miami Tropical Research Laboratory. DNA fingerprinting studies have revealed it is not identical to the authentic SCA-6 accession in the International germplasm collection at Trinidad. The plants were grown in greenhouses at the Pennsylvania State University in deep pots (6.35 cm diameter, 25.4 cm deep) with silica sand for adequate drainage. Humidity was maintained at 60%, and the photoperiod was set to 12 h light/29°C and 12 h dark/26°C. Natural light was supplemented with 430-W high pressure sodium lamps as needed to maintain a minimum light level of 250 $\mu\text{mol m}^{-2}\text{s}^{-1}$ PAR, while automatically retractable shading limited light levels to a maximum of 1000 $\mu\text{mol m}^{-2}\text{s}^{-1}$ PAR. Irrigation with 1/10-strength Hoagland’s nutrient solution (160 ppm N) was applied daily at multiple times to maintain adequate moisture.

2.2.2 Time-lapse photography and morphometric measurements

Time-lapse photography was performed using a Canon S40 digital camera controlled by an iMAC computer and Canon Remote Capture (v.2.7.3) software. Photos were taken every hour over two separate time periods: the first was from Feb 3 - Feb 18, 2003, consisting of 16 days, the second was from Feb 24 – Mar 31, 2003, consisting of 34 days. These photographs were interlaced using the Apple computer iMovie® software and can be viewed at <http://guiltinanlab.cas.psu.edu/Research/Cocoa/flowers.htm>. Growth measurements were taken from the images obtained at 2 pm each day. Three measurements were taken for each flower; length of the flower bud (distal tip of the sepals to the base of the receptacle), width of the flower at its widest point, and the length of the pedicle (from the base of the flower bud to the tree). In total, six separate flower clusters were observed containing a total of 56 flowers. Measurements were calibrated from the images using a ruler that was physically attached to the tree and included in each photograph. Using the Adobe Photoshop® ruler, flower buds were measured and converted to millimeters (Figure 2-1, c).

To determine the rate of flower opening (the rate for the fully developed flower bud to open), sequential one-hour images were collected of eight individual flowers undergoing flower opening. These eight flowers were from four different flower cushions and from five different times throughout the time-lapse range. Photographs were used to measure the angle of flower opening by measuring the angle of the inside edges of the sepals with a protractor. Zero degrees indicated a closed flower while 180 – 220 degrees indicated a fully open flower.

2.2.3 Light microscopy

Flower samples were collected at various stages of growth and were fixed in FAA (50% ethanol, 5% acetic acid, 3.7% formaldehyde, and 41.3% water) overnight

(<16 hours) under vacuum. The tissue was then put through a series of ethanol washes of 50%, 50%, 60%, and 70% of at least 30 min each. Tissues with large airspaces were slit on one side using a razorblade to allow easier access to the tissues before embedding in paraffin. Tissues were embedded using the Shannon Citadel 2000 (Shandon, Pittsburgh, PA) under the following 22 hour program: 70% ethanol for 15 min, 80% ethanol for 30 min, 95% ethanol for 30 min, 95% ethanol for 8 hours, 100% ethanol for 1 hour, 100% ethanol for 30 min, three changes of HistoSolve (ThermoShandon, Pittsburgh, PA) for one hour each, and two changes of paraffin for 4 hours each. The tissues were embedded in paraffin blocks, hardened for 24 hours at room temperature and sectioned to 8 μm using a Shandon Finesse Microtome (Thermo Electron Corp., Pittsburgh, PA). The sections were placed on positively charged glass slides (Cat. number 6776214, Thermo Electron Corp., Pittsburgh, PA) and incubated overnight on a slide warmer at 42°C. The sections were stained using a solution of toluidine blue (0.1%) for 15 sec. then rinsed in water for at least one hour, after which they were dipped for 10 seconds each through the following ethanol and histosolve series: 25% ethanol, 50% ethanol, 75% ethanol, 100% ethanol, and histosolve. The sections were covered with a glass coverslip. Sections were photographed at 500x magnification using an Olympus BX-60 compound microscope with an Automated Prior Stage (Prior, Rockland, MA) controlled by the ImagePro Plus (v. 4.1) (Media Cybernetics, Silver Spring, MD) software. In total, 12 flowers were photographed encompassing all stages of flower development.

2.2.4 Scanning electron microscopy

Scanning electron microscopy (SEM) was conducted by collecting eighteen flower samples at various stages of growth. Samples were fixed in FAA (50% ethanol, 5% acetic acid, 3.7% formaldehyde, and 41.3% water) overnight (>16 hours) under vacuum. The tissue was put through a series of ethanol washes of 50%, 50%, 60%, 70%, 85%, 95%, and three washes of 100% of at least 1 hour each. The samples were then critical point dried using bone dry carbon dioxide with exchange conditions of 10°C

and 3.4 atm, at 31°C and 4.76 atm in the Bal-Tec CPD030 Critical Point Dryer (Technotrade, Manchester, MD). Tissues were then dissected, sputter coated with gold, and viewed using a Jeol JSM-S400 Scanning Electron Microscope (Jeol, Peabody, MA). The images were calibrated using PGT Imix-PC V.10 software to obtain accurate size measurements.

2.2.5 Measurement of floral tissues

Using all three sources of data, a large number of morphometric parameters were recorded (Table 2-1). In total, the internal organs of thirty-two flower buds were measured. To correct measurements for the level of tissue shrinkage during the fixation process, photographs of representative samples were taken before fixing and the length of flower buds were measured (FrL) using Adobe Photoshop®. These same tissues were compared to the resulting fixed tissues used in light microscopy and SEM (FxL). Using the two measurements FrL (length before fixation), and FxL (Length after fixation), size measurements were corrected for tissue shrinkage using the formula $Fr(L)=Fx(L)+(Fx(L)\times 0.14)$.

2.3 Results

2.3.1 Analysis of floral development using time-lapse, macro-imaging

To precisely measure the rate of organ growth during floral development, time-lapse imaging was used to capture hourly images of flowers growing on trees in greenhouse conditions. The compiled images can be viewed as time-lapse movies and are available online (see Methods). The measurements from the 2 pm time point images were initially grouped by flowers within cushions and variation among floral cushions and between the time periods compared. No significant effect of time or flower cushion

position on the individual growth rates of the flowers (Figure 2-2, a and b) was observed. In other words, in this experiment there was no positional or temporal effect on floral growth rate.

Using these data, we were able to combine the measurements for each parameter separately into a single Lowess curve encompassing all of the data. This curve was compared by correlation to the raw data and produced an r^2 of 0.96. (Figure 2-2, c). We were able to relate time with the length of the flower bud and it was determined that the duration of flower development in genotype PSU SCA-6 was 30 days. While the length and width of the flower bud increased for the entire 30 days, the pedicle grew only between days 10 to 30. At the time of flower opening, the average length of the flower bud is 5.7 mm (s.d. 0.4 mm), the average width of the flower bud is 3.6 mm (s.d. 0.2 mm), and the average length of the pedicle is 13.2 mm (s.d. 1.4 mm). Minor variation in mature flower size is a well-known genotypic indicator and has been observed in different genotypes in the field (M. Guiltinan pers. comm.), so these sizes are representative of this genotype only.

By calculating the rate of growth (size increment in mm / time in days) from the original data, then fitting Lowess curves to the rate data set, we were able to describe patterns in the relative rates of growth in the bud length, bud width and pedicle length. From these calculations we estimated that the pedicle had the highest terminal rate of growth (growth rate just before flower opening) of 0.54 mm d^{-1} , followed by bud length growth of 0.26 mm d^{-1} , and finally bud width growth of 0.16 mm d^{-1} (Figure 2-2, d). The rate of growth in pedicle length generally increased throughout the development of the flower bud. However, the greatest increase in growth rate was observed from day 17 to 30. The rate of growth of flower bud length and width actually decreased to a minimum growth rate (0.09 mm d^{-1}) at day 16, after which the growth rate increased for the remaining duration of growth (Figure 2-2, d). When the flower initially forms, it is enclosed by two leafy bracts. By day 15, the flower bud emerges from the bracts and the pedicle continues to elongate. At flower opening the bracts have abscised leaving a scar

that later was observed to form just below (200 μm in a 19 day old flower bud) the site of abscission of unfertilized flowers (Figure 2-3, a).

Table 2-1: Definitions and abbreviations of flower measurements made in *T. cacao* flowers

Name	Symbol	Measurement
Anther Diameter	AntherD	Diameter of the anther from the face side
Filament Length	Fill	Length of the filament from the connection to the anther to the receptacle
Staminode Length	StamL	Length from the distal tip of the staminode to the receptacle
Ovary Diameter	OvryD	Width of the ovary at its widest point
Ovary Height	OvryH	Length of the ovary from the base of the style to the receptacle
Style Length	StyleL	Length of the style from the top of the ovary
Sepal Wall thickness	SepWITh	Width of the sepal wall at it thickest point
Concha Length	CL	Length from the receptacle to the beginning of the connective “bar” (taenial-segment)
Taenial-segment Length	TS	Length of the constrictive “bar” between the Concha and Ligula
Ligula Lenth	LL	Length of the blade of the petal from the distal end of the taenial- segment to the distal end of the petal
Inside	In	Inside of the flower bud from receptacle to the inside surface of the sepal
Outside	Out	The outside of the of the flower bud from the receptacle to the outside surface of the sepal
Pollen Diameter	PolDia	Diameter of pollen
Female Gametophyte	FG	Diameter of female gametophytes
Fixed Length	FxL	Fixed total length from the beginning of the pedicle to the outside surface of the sepal
Fresh Length	FrL	Fresh total length from the beginning of the pedicle to the outside surface of the sepal
Fresh Width	FrW	Fresh total width of the flower bud at it widest point
Pedicle Length	PedL	Length of the pedicle from the peduncle to the trunk

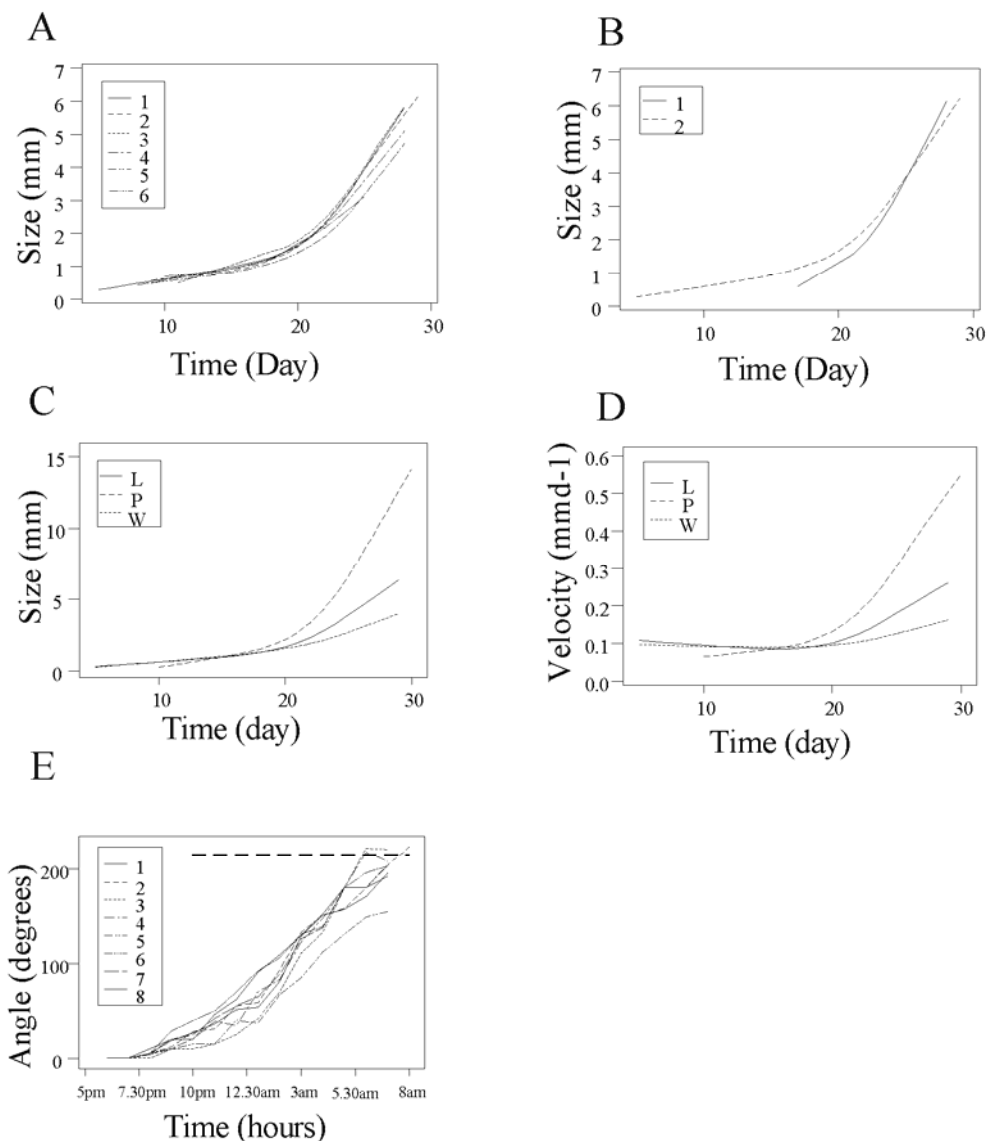


Figure 2-2: The growth, velocity, and opening angles of growing *T. cacao* flower buds over time: **A**, Lowess curve showing the length of the bud (Size in mm) vs. time (day of growth) comparing flowers of six separate flowers. **B**, Lowess curves describing the length of the bud (size in mm) vs. time (day of growth) comparing flowers of two separate times. **C**, Lowess curves describing size (length, width, and length of pedicle in mm) vs. time (day of growth) using combined data from all flower cushions and both time periods. **D**, Lowess curves describing velocity (length, width, and length of pedicle in mmd^{-1}) using data from combined flower cushions and both time periods. **E**, Opening angles (degrees) of eight flower buds over the time (hours) required for the flowers to open (dashed line shows where the flowers are fully open).

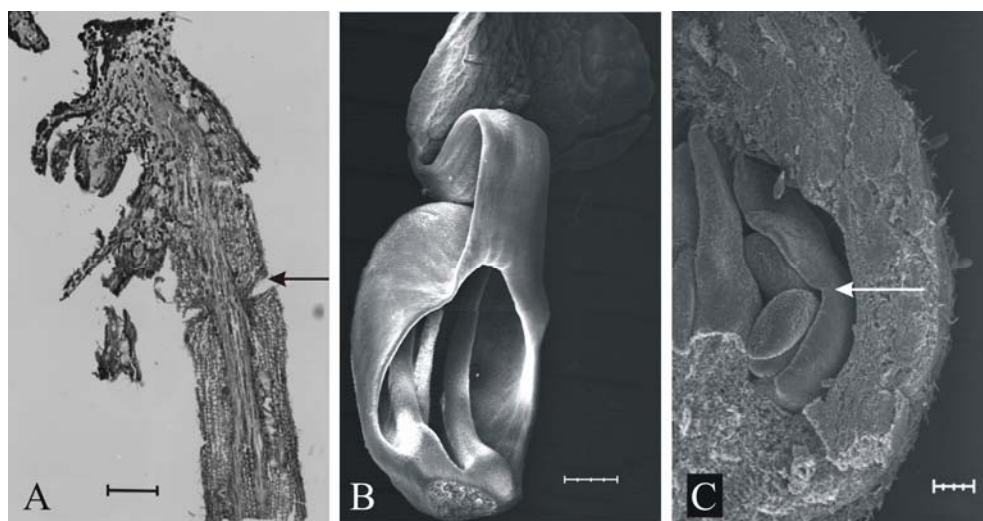


Figure 2-3: Micrographs showing important features of the *T. cacao* pedicel **A**, and features of the petal (**B-C**): **A**, Micrograph showing the pedicel of a 19 day old flower bud. The arrow shows the restriction almost to the vasculature of the pedicel. This becomes the site of abscission for the flower. **B**, An electron micrograph of a 29 day old petal of *T. cacao* showing the three sections; ligula, taenial-segment and concha, as well as the two “running tracks” used to guide the filament in the latter stages of elongation. **C**, Scanning electron micrograph of a size one flower bud, 16 days old. The arrow indicates the restriction appearing in the petal that will eventually form the taenial-segment. Bars=200 μm in **A**, 100 μm in **B**, and 500 μm in **C**.

2.3.2 Microscopic analysis of floral development

To assess in more detail the internal development of cacao flowers, thirty samples encompassing all stages of flower development were fixed and dissected (see methods for details). To allow us to compare size data from images taken of fresh flowers with those of fixed tissues, photographs were taken both before and after fixation to measure the level of shrinkage due to the fixation process. It was found that between 7-20% shrinkage occurred. There was no correlation between size of tissue fixed and percentage shrinkage. There was also no difference in shrinkage between the tissues

fixed for SEM vs. light microscopy. Thus, an average increase of 14% was used to adjust all data from fixed samples to the size of fresh samples.

Flowers of six size categories were arbitrarily selected for observation by both electron and light microscopy. These size classes are described in Table 2-2 and are shown in Figure 2-4. Size 0 buds were extracted by dissecting the flower cushion from the tree. Size 1-5 buds could be extracted without disturbing the flower cushion (sizes are as indicated on the top half of Figure 2-5). The morphometric data were analyzed using linear regression to create models that described the particular mode of growth of each organ relative to the length of the total flower bud (e.g. linear, log, exp). These growth models can be found in Table 2-3. These models accurately described the growth of the organs contained in the four whorls of the *T. cacao* flower.

To validate our models relating organ growth to bud length, we re-analyzed the electron micrographs from Bayer and Hoppe (1990) in the same way as we did our own electron micrographs (data not shown). There was little difference observed between our data vs. the Bayer and Hoppe (1990) data in terms of the actual size data and the models produced. This provides further validation as to the universality of the models presented here.

It is important to note that it has been observed in both *Arabidopsis* and *T. cacao* that the size of the flower buds can vary by up to 40% in *Arabidopsis* and 20% in *T. cacao* depending upon genotype and growth conditions (Ma and Gultinan pers. comm.). This indicates that models reflecting growth may not be quantitatively accurate when applied to other genotypes and growth conditions of *T. cacao*. To create more universal models, growth was expressed as a percentage of final size for each organ measured and related to time of development (Table 2-4).

Table 2-2: Size and age of *T. cacao* flower buds sampled for microscopy.

Name	Average Length (FrL)	Std. dev of size (mm)	Age (d) (from Lowess)	% age Length*	Description
Size 0	0.9	NA	<14	NA	Buds less than 1mm in length
Size 1	1.2	0.2	15	12	Smallest bud size that could be extracted without disturbing the flower cushion
Size 2	2.3	0.09	22	20	Green flower bud with a pedicle
Size 3	3.3	0.7	25	38	Flower bud turning from green to white
Size 4	5.4	0.3	27	74	Flower bud midway in size after it has turned white
Size 5	6.5	0.4	29	100	Flower bud the day before it is about to open
Size 6	N/A	N/A	30	N/A	Open flower

* %age Length = (Length at time X / Final Length) *100

N/A = Could not be measured

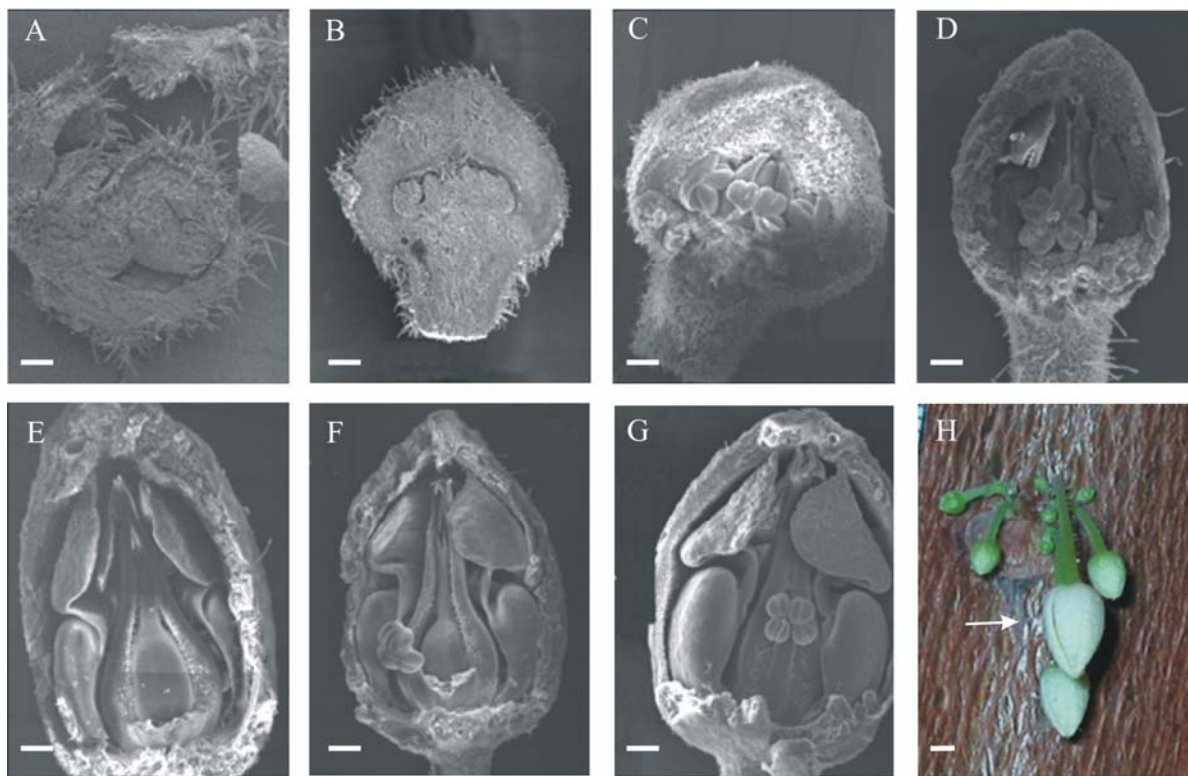


Figure 2-4 **A-G**: SEM photographs showing the ontogeny of the *T. cacao* flower and **H**: a photograph of a flower cushion. **A**, A four day old meristem emerged from the cyme and protected by leafy bracts. **B**, A 12 day old flower bud showing the early formation of floral organs. The flower buds at this time have relatively thick sepals relative to the thickness of the meristematic tissue. **C**, A 15 day old size one flower bud showing that most organs are fully developed and are ready to expand. **D**, A 19 day old size two flower bud showing the constriction of the petal that will eventually form from the taenial segment. **E**, A 22 day old size three flower bud. The concha has completely enclosed the anther whose filament has not yet begun to expand. **F**, A 27 day old size four flower bud in which all organs are expanding. The style is not yet at the same length as the staminodes. **G**, A 29 day old fully expanded flower bud just prior to flower opening. At this stage all organs are fully developed and at their final relative lengths. The anther is housed in the concha, the style is the equivalent length of the staminodes and “bumps” have appeared on the carpel. **H**, A small flower cushion of *T. cacao*. The white arrow shows the abscission zone where the sepals are beginning to separate in an opening flower. Bars 100 μm in **A**, 200 μm in **B**, 250 μm in **C**, 300 μm in **D**, 500 μm in **E**, **F**, and **G**, and 2 mm in **H**.

Table 2-3: Growth models of various organs during the development of *T. cacao* flower buds.

Bud Length (L) Vs.	Model	R ²	Type	Growth Begins at... (Bud length mm, day)
Fresh Width	$L=1.60 \times FrW - 0.5$	0.97	Lin	N/A
Pedicle Length	$L=1.91 \times PedL - 1.14$	0.99	Lin	0.56, 10
Anther Dia.	$L=0.3 \ln(\text{AntherD}) + 0.31$	0.88	Log	0.35, 6
Filament Leng.	$L=0.11e^{0.36 \times FilL}$	0.87	Exp	0.61, 15
Staminode Leng.	$L=0.74 \times StamL - 0.48$	0.94	Lin	0.32, 6
Ovary Ht.	$L=0.23 \times OvryH + 0.07$	0.89	Lin	0.35, 6
Ovary Dia.	$L=0.14 \times OvryD + 0.11$	0.82	Lin	0.33, 6
Style Leng.	$L=0.05 \text{StyleL}^2 - 0.02 \text{StyleL} + 0.02$	0.92	Poly	1.17, 17
Concha Leng.	$L=0.01 \text{CL}^3 + 0.16 \text{CL}^2 - 0.18 \text{CL} + 0.04$	0.96	Poly	0.88, 14
Taenial Segment	$L=0.05 \text{TS}^2 - 0.06 \text{TS} + 0.14$	0.96	Poly	0.95, 14.9
Ligula length	$L=0.35 \text{LL} - 0.1$	0.96	Lin	0.29, 6
Female Gametophyte	$L=0.25 \times FG - 0.17$	0.99	Lin	1.21, 18
Pollen	N/A	N/A	N/A	1.87, 21

Table 2-4: Growth models of percentage growths of various organs during the development of *T. cacao* flower buds.

Bud Length Vs.	Model	Type	R ²	Notes
% Anther Diameter	% AntherD=0.0895Day ² +0.5545Day-7.1103	Lin	0.92	
%Staminode Length	% StamL=0.1047e ^{0.2292xDay}	Exp	0.85	
%Style Length	% StyleL=0.6858Day ² -24.748Day+221.55	Lin	0.89	After d 14
%Petal Length	% PetalL=3x10 ⁻⁰⁶ Day ^{5.104}	Pow	0.93	
% Ovary Height	% OvryH=1.7335e ^{0.1327xDay}	Exp	0.76	
% Ovary Diameter	% OvryD=4.5929Day-39.247	Lin	0.86	
%Filament Length	% FilL=0.6594Day ² -25.225Day+242.7	Lin	0.82	After d 14
%Female Gametophyte	% FG=8.3335Day-151.61	Lin	0.96	After d 18
% Fresh length	% FrL=1.7783e ^{0.1355xDay}	Exp	1	
% Fresh width	% Width= 2.6138e ^{0.1254xDay}	Exp	0.96	
% Pedicle Length	% PedL=0.4077e ^{0.1865xDay}	Exp	0.98	
Length	y = 0.1223e ^{0.1355xDay}	Exp	1	

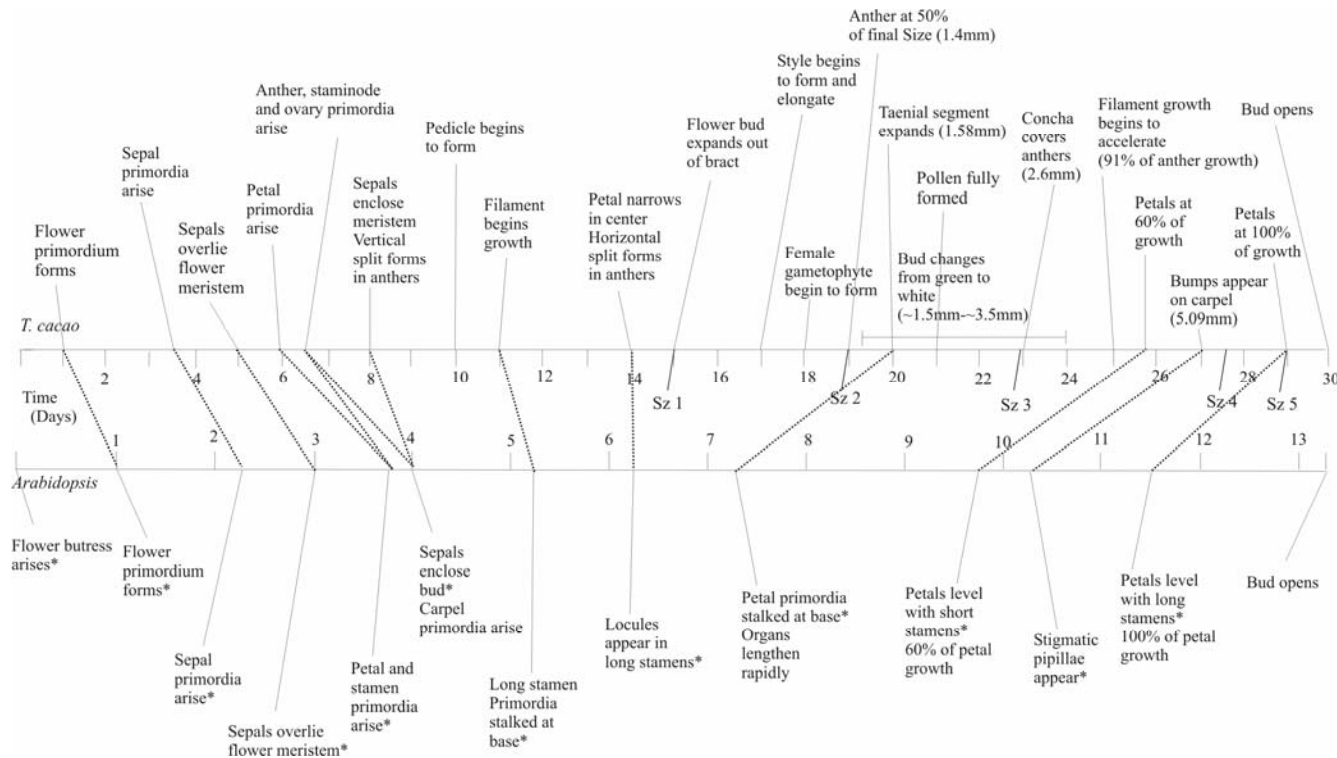


Figure 2-5: Comparison of *T. cacao* and *Arabidopsis* flower development. The major morphological events in the growth of *T. cacao* (top) and *Arabidopsis* (Bottom) plotted along an axis of thirty and fourteen days, respectively. Bud sizes (Sz 1-5) refer to the average sizes of flower buds used in this experiment and are indicated along the axis. The dotted lines connecting the two diagrams represent stages as defined by Smyth et al. (1990).

Table 2-5: Comparative stages in *T. cacao* and *Arabidopsis* flower development

<i>Arabidopsis</i> Stage	<i>T. cacao</i> Stage	Description	Time in <i>Arabidopsis</i> (Beginning of stage (day))	Time in <i>T. cacao</i> (beginning of stage (day))	Ratio Cacao/Arab
1	1*	Flower butress arises	0	n/a	n/a
2	2	Flower primordium Forms	1	1	1
3	3	Sepal primordia arise	2.25	3.5	1.55
4	4	Sepals overlie flower meristem	3	5	1.66
5	5	Petal and stamen primordial arise	3.75	6	1.6
6	6	Sepals enclose bud	4	8	2
7	7	Long stamen primordial stalked at base	5.25	11	2.09
8	8	Locules appear in long stamens	6.25	14	2.24
9	9	Petal primordial stalked at base	7.25	20	2.75
10	10	Petals level with short stamens (60% of growth)	9.75	26	2.66
11	11	Stigmatic papillae appear	10.25	27	2.63
12	12	Petals level with short stamens (100% of Growth)	11.5	29	2.52
		Flower Opening	13.25	30	2.26

* this stage has not been observed in cacao due to the cyme structure, but is thought to exist on the first flower produced in the flower cushion.

2.3.3 Description of developmental events

To facilitate comparison of the morphological development of *T. cacao* and *Arabidopsis*, twelve morphological stages correlating to those found in *Arabidopsis* development were defined for *T. cacao* (Smyth et al., 1990) (Table 2-5). Stages 1-6 primarily were involved with the meristem development and the organogenesis of the floral organs and were complete by day 10. Stages 7-12 were involved with elongation

and further differentiation within individual organs to produce the fully developed flower by day 30.

2.3.4 Stages 1-6

Stage 1 is defined by the emergence of the floral buttress arising from the trunk in the initial flower of a specific flower cushion. This stage was not observed as this research only involved developed flower cushions and no floral buttresses were observed directly in the cyme structure. We suggest that the floral buttress would occur only in the very first flower bud at each leaf axil. However, we did observe in young plants that primary floral meristems arose from the location of a leaf scar, most likely from auxiliary meristems that formed during vegetative development.

Stage 2 begins with the formation of the flower primordium. This occurs near the base of the pedicle of a maturing flower that is usually of stage nine or above. The emergence of primordia was first observed as a dome of cells that was 140 μm wide.

The formation of sepal primordia is apparent when the floral meristem is 0.18 mm in width (at about 3.5 days, 3% of FrL) and mark the beginning of stage 3. They initiate in a quincundaler sequence, meaning that if the sepals were numbered clockwise from 1 to 5, the sepals would initiate in the order 1, 3, 5, 2, and 4 (Bayer and Hoppe 1990). The sepals continue to grow and overlie the floral meristem by day five. This demarks stage 4. Stage 5 is defined by the emergence of petal primordia when the bud is 0.29 mm in length (6 days, 4% FrL) (Figure 2-4, b). Shortly after, the anther (0.35 mm, 6 days) (Figure 2-4, b), the staminodes (sterile stamen-like structures that also form part of this whorl) (0.32 mm, 6 days) (Figure 2-4, b), and five carpel primordia (0.35 mm in length, 6 days) (Figure 2-4, b) form on the meristem. The carpel primordia soon grow together and eventually fuse to form a tube.

Stage 6 occurs as the sepals continue to grow and completely enclose the meristem (day eight) (Figure 2-4, a). At this stage the sepals are relatively thick, half of the total thickness of the developing flower (Figure 2-4, b and c). Also at stage 6, the anthers form vertical splits down their centers resulting in two lobed structures (0.44 mm, 8 days, 5.2% FrL). Concurrently, the staminodes begin to become pointed as the anthers form this vertical split. Finally the pedicle begins to form at day 10 (6.9% FrL).

2.3.5 Stages 7-12

Stage 7 is marked by the formation of the anther filaments when the bud is 0.61 mm in length (11 days old, 7.9% FrL). Meanwhile the petals grow and form a triangular shape as they elongate from day 6 to day 14 (4%-11.9% FrL) (Figure 2-4, c). When the bud is 0.88 mm in length (14 days, 11.9% FrL) a constriction appears near the base of the petals (Figure 2-4, d, and Figure 2-3, c). The sections above the constriction form the ligula, while the sections below form the concha. This constriction is indicative of stage 8. At this time, the anther primordia have separated from the meristem and a horizontal split begins to form, (0.85 mm, 14 days), giving anthers their characteristic four-lobed morphology (Figure 2-4, d, f, and g). The staminodes are taller than the anthers at this stage. Moreover, staminodes initially grow at the same rate as petals eventuating in a longer final length than petals. The staminodes overtake the petals in length as the taenial-segments form (14 days).

Stage 8 continues with the flower buds expanding out of the bracts at day 15 (13.6% FrL). The carpels close in at the top and the style forms when the bud reaches a length of 1.17 mm, at 17 days (17.8% FrL). Bayer and Hoppe (1990) reported that closing of the carpel coincided with the formation of pollen, however we observed the first pollen formation to occur a few days later and was actually a part of Stage 9. Another landmark of stage eight is the initial formation of female gametes when the bud is 1.21 mm in length (18 days, 20.4% FrL). The individual gametes grow at a linear rate

over the time it took for the flower to reach flower opening (Table 2-3). The anthers follow a logarithmic pattern of growth more closely than any other organ measured. Due to this growth pattern, anthers are one of the fastest forming organs in the flower, completing 50% of their growth by day 19 (Figure 2-4, c).

During stage 9, the petals form three distinct sections (Figure 2-1 and Figure 2-3, b). The concha is proximal to the receptacle and the ligula is most distal. The taenial-segment of the petal connects the ligula and concha. The constriction mentioned in stage 8 begins to elongate when the bud reaches 1.58 mm in length (20 days, 26.7% FrL) and forms the taenial-segments. Pollen is fully formed soon after the horizontal split in the anther forms (1.87 mm, 21 days, 30.6% FrL) (Bayer and Hoppe, 1990). The average diameter of the pollen is 0.018 mm. The outside of the flower bud begins to change in color from green to white as the flower bud expands (1.5 mm, day 19.5 to 3.5 mm, day 24, 46% FrL). Once filaments form in stage 7, they show little elongation until the conchas cover over the anthers at which time the filaments begin to accelerate in growth (4.47 mm, 25 days, 52.6% FrL) (Figure 2-4, e, f, and g). The anthers are at 91% of final length when the filaments begin to expand. This filament growth is guided into the conchas by the two track-like grooves that ran on the adaxial surface and resemble a pair of tracks (Figure 2-3, b). These “tracks” appear to guide the filaments to position the anther in the center of the concha.

Stage 10 is represented when the petals obtain 60% of their final length (at day 26, 60.2% FrL). Stage 11 is observed as the final morphological structures appear as “bump-like” lobes on the surface of the carpel and hair-like trichomes on the staminodes (Figure 2-4, f, and g). These trichomes completely enclose the style of the flower (Figure 2-4, f, and g). Stage 12 is represented by the petals obtaining 100% of their final length (day 29). At this point, the taenial-segments are curved and represent ribbon-like structures (Figure 2-4, e, f, and g). It is also notable that as the flower continues to grow and lengthen, the sepal walls do not grow any thicker. They lengthen but remain at this initial thickness (0.25-0.35 mm) until flower opening (Figure 2-4, d, e, f, and g).

2.3.6 Flower opening and abscission

Marking the initiation of flower opening, the sepal tissues, which initially develop as one contiguous whorl, form a series of five longitudinal abscission zones, from the tip to the base of the floral bud. These longitudinal abscission zones separate (Figure 2-4, h), forming the five individual sepals which then expand and open outward. As the sepals open outwards, the petals open at the same time. Our measurements reveal that flower opening takes place over a twelve hour period beginning with the evening of day 29. We observe that under our conditions, on average, the first observable splits (average of 4.8 degrees, s.d. 2.6) in petals occur between 6-7 pm of the 29th day and concluded with the flower being completely open (an average of 197 degrees, s.d., 20.7 degrees) by 6 am on the 30th day, shortly before sunrise.

It is also interesting to note that flower opening is very well synchronized between the cohort of mature flowers opening each night. The flowers open at almost exactly the same time and rate, irrespective of the position on the trunk (Figure 2-2, e), a fact dramatically observable in the online video images (URL above). Eight individual flowers located in different flower cushions were observed and seven of the eight flowers began opening during the same hourly interval (6 pm). It was also observed that the flowers all opened at similar rates, with an average standard deviation among the opening angle of eight flowers over the entire period of flower opening of only 12.8 degrees. The velocity associated with the opening of the flowers is also fairly constant among the flowers studied (average of 14.7 degrees hour⁻¹) with the fastest rates occurring between 1 am and 3 am (23-33 degrees hour⁻¹, s.d.= 11 degrees hour⁻¹), and at 4 am (27 degrees hour⁻¹, s.d.=15 degrees hour⁻¹). This high degree of synchronicity is likely the result of millions of years of selective pressure to ensure the optimal receptivity of flowers upon opening at the time of maximal insect activity in the early morning hours.

Unfertilized flowers are observed to abscise from the trunk approximately one day after flower opening. This is in accordance with previous descriptions

(Cheeseman, 1932). By observing the time between flower opening and when the flowers fell from the tree, abscission of the unfertilized flower is estimated to occur at the earliest between 3-5 am the day after flower opening. This is completed just prior to opening of the next set of flowers, and results in increased visitation of pollinators to the newly opened flowers by the reduction in competition of the day old flowers. Normally, a small physical disturbance is required to shake the flower loose (e.g. wind, physical contact) and as a result, in absence of contact, some flowers fall several days post flower opening. We also notice that the site for abscission is at the bract scars as we have described above. This can be most readily appreciated when viewing the time-lapse movie.

2.3.7 Organs novel to *T. cacao*

The flowers of *T. cacao* have several features that are quite different from *Arabidopsis*. The most obvious feature is the formation of sterile stamen-like structures called staminodes. Staminodes do occur in other plant species and have been used as a morphological tool to understand plant evolution (Walker-Larsen and Harder, 2000). These are interdigitated with the stamens in whorl three and begin development at the same time as the stamens. The function of the staminodes is unknown, however it seems reasonable to speculate that the staminode barrier between the anthers and the stigmatic surfaces reduces the incidence of self-pollination by midges and other small insects, and thus contributes to increase out-crossing in self-compatible genotypes of *T. cacao*, but this hypothesis remains to be tested.

2.4 Discussion

2.4.1 Floral morphogenesis

Under the conditions used in our studies, flowers of *T. cacao* take 30 days to form from an initial floral meristem containing a few cells, to a fully developed receptive flower. This is consistent with what has been reported by others (Bayer and Hoppe 1990; Smith 1992; Wood and Lass 1985), but longer than the time of 21-24 days reported by Rajamony and Mohanakumaran (1991). This latter discrepancy could have resulted from: 1) Rajamony and Mohanakumaran (1991) not recording the early stages of development, or 2) differences in genotype, or 3) environmental effects, none of which could be determined from their report.

The flower forms from a compressed cincinnal cyme structure and is cauliflorous in nature, growing from the main trunk and branches of the plant. The specific site of inflorescence has been observed by others to be from the periderm of old stems and axils of abscised leaves in lateral branches (Bartolome, 1951; Zamora et al., 1960). In young flowering plants we observed that the primary floral meristem arises from the location of a leaf scar, most likely from auxiliary meristems formed during vegetative development. The flower bud grows to an average length of 5.7 mm and an average width of 4 mm just before flower opening. The pedicle grows to an average of 14.1 mm.

The flowers of *T. cacao* undergo opening over a twelve hour period. In our study, this took place at night between ~7 pm and ~7 am (under greenhouse conditions). It is also important to note that the flowers, irrespective of their position on the tree, opened in a synchronous fashion at similar rates over this period. This implies that there is a very strong regulation controlling the initiation and the rate of flower opening. We observed that the opening of the flower was not a smooth motion but rather

discontinuous with the rates of opening constantly increasing and decreasing. These changes in opening rates were reproducible from flower to flower, suggesting that the plant has mechanisms in place to control the rates that flowers open. The mechanism and its physiological nature remains to be investigated.

The flowers normally remain open on the tree for one day, after which they abscise. A constriction close to the base of the pedicle, distal from the flower, was observed first by Cheeseman, (1932). He did not describe the function of the restriction but did note that it extended almost to the vascular strands of the pedicle. We have also observed the restriction (Figure 2-3, a) and postulate that this constriction forms the site for abscission of the flower.

Our observations of the development of the internal organs of the *T. cacao* flower bud generally agree with those made by Bayer and Hoppe (1990) and Zamora (1960). However, there is some disagreement in the literature on sequence. Zamora (1960) infers that the internal whorls (androecium and gynoecium) both begin to form together, whereas Bayer and Hoppe (1990) state that, the flower forms in an acropetal sequence, in other words sequentially initiating in the order of sepals, petals, androecium and gynoecium. In our studies, we also observed that the flower forms in an acropetal sequence, however the internal two whorls initiate very close together, less than 24 hours, very near the limit of the temporal resolution of the models presented of this study.

Our time-lapse data and the resulting Lowess curves allowed us to relate temporal data to the size of the observed flower buds. This provides predictive power to relate subsequent size data to age of the bud. By sampling over the entire spectrum of floral development, we are able to use models describing the size/shape and growth rate of different organs. The models allow experimental manipulations to be performed on specific stages of development, without having to dissect flowers each time to determine their stage of development.

2.4.2 Comparison to the temporal growth of *Arabidopsis thaliana* flowers

Arabidopsis thaliana is the most well studied model system for investigating the mechanisms regulating flower development (Meinke et al., 1998; Simpson and Dean, 2002). Its flowers are less than half the size of *T. cacao*, with a bud length just before flower opening of approximately 2.8 mm. The structure of the *Arabidopsis* flower includes a calyx of four sepals, a corolla of four petals, an androecium with four medial (long) stamens and two lateral (short) stamens. The gynoecium is comprised of two carpels (Smyth et al. 1990). The *T. cacao* flower and *Arabidopsis* flower have very different morphologies as illustrated in Figure 2-6. Smyth et al. (1990) divided *Arabidopsis* flower development into twelve temporal stages. They reported that *Arabidopsis* buds took 13.25 days to form and develop. Evolutionarily, *T. cacao* and *Arabidopsis* are fairly closely related with both species belonging to Eurosids II. Using the data from this study, we compared the relative rates and duration of *Arabidopsis* and *T. cacao* flower development in relation to developmental events. The total duration was defined as the time that elapsed between meristem formation and flower opening, against which we compared timing of the developmental stages of flowers of both species (Figure 2-5).

2.4.3 Stage 1-6

Stage 1 is defined in *Arabidopsis* as the flower buttress arising. This was not observed in our observations of *T. cacao* flower development due to our observations being exclusively on previously formed flower cushions. The buttress is hypothesized to exist in *T. cacao* during the formation of the first flower of a flower cushion. For this reason, we have hypothesized that this stage exists but is not included in our analysis because of the difficulty in capturing this very rare event. Stage 2 is defined as the formation of the flower primordium. For *T. cacao*, this occurs at day 1 and lasts for 3.5 days, in *Arabidopsis* this stage also begins at day 1 but lasts for 2.25 days.

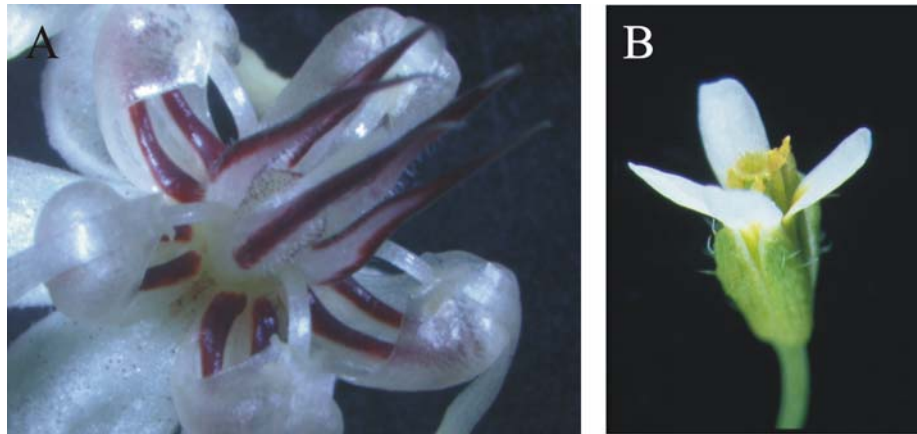


Figure 2-6: Photographs of a *T. cacao* flower and an *Arabidopsis* flower to allow comparison of final floral structure. **A**, An open and fully developed *T. cacao* flower (photograph taken by M. J. Gultinan). **B**, A flower of *Arabidopsis* (Photograph from ucsdnews.ucsd.edu) **A** and **B** are not to relative scale.

Stage 3 is defined in *Arabidopsis* as when the sepal primordia arise. This begins at day 2.25 and lasts for 18 hours. In *Arabidopsis* the abaxial sepal arises first followed by the adaxial and two lateral sepals (Smyth et al. 1990). In *T. cacao*, this stage begins at day 3.5 and lasts for 1.5 days. The sepal primordia arise in a quincundaler sequence as presented in the results section. Sepals overlying the meristem demarcate the beginning of stage 4. This occurs in *Arabidopsis* at day 3 and lasts for 18 hours compared to *T. cacao* where it occurs at day 5 and lasts for one day.

Stage 5 in *Arabidopsis* is defined as the formation of petal and stamen primordia, and is the first major deviation in development between the two species. In *Arabidopsis* both petal and stamen primordia form simultaneously at day 3.75 and the stage lasts for 6 hours, however in *T. cacao* petal primordia form initially at day 6 followed 12 hours later by the formation of the anther, staminode and carpel primordia sequentially. Stage 5 lasts for 2 days in *T. cacao*. Stage 6 is marked in *Arabidopsis* by the sepals completely enclosing the bud. This occurs at day 4 and lasts for 1.25 days. Another important event at stage 6 for *Arabidopsis* is the initial formation of the carpel

primordia. This is another difference between the two species, as *T. cacao* has already formed carpel primordia by the end of stage 5. Stage 6 in cacao begins at day 8 and lasts for 3 days. Important events that occur in *T. cacao* stage 6 that do not in *Arabidopsis* are the vertical splits that form down the length of the anthers and the initial formation of the pedicle.

2.4.4 Stage 7-12

The initial filament formation defines the start of stage 7 in *Arabidopsis*, occurs at day 5.25, and lasts for one day. In *T. cacao* stage 7 occurs at day 11 and lasts for three days. Stage 8 is marked by the formation of the locules in *Arabidopsis*, occurs at day 6.25, and lasts for one day. This can be identified by the long stamens forming visible convex protrusions on the inner surface. In *T. cacao* the locules occur once the horizontal split forms on the anther giving its four lobed structure. Stage 8 in *T. cacao* begins at day 14 and lasts for six days. Another important difference between the two species is that *Arabidopsis* has four stamens, two long and two short. *T. cacao* has five fertile stamens that all grow at identical rates to one another and grow to the same length. In *T. cacao* other landmark events includes the flower expanding out of the bract, the style forming and beginning elongation, and the female gametophytes beginning to form. It is also important to note that at 1.4 mm (19 days) that the anther has already formed 50% of its total final diameter.

Stage 9 is reported in *Arabidopsis* as having the petal primordia stalked at their base. It also marks a period where all organs lengthen rapidly. Stage 9 begins in *Arabidopsis* at day 7.25 and lasts for 2.5 days. *T. cacao* also undergoes a period of rapid expansion during stage 9 (day 20 and lasts for 6 days). The initiating event is the lengthening of the taenial-segment, which is followed by the completion of pollen formation. The chonchas cover the anthers in stage 9 and the filaments elongate pushing the anthers into the chonchas using the previously described “running-track” like

structures. Stage 10 is defined as the petals in *Arabidopsis* being level with the short stamens; it begins at day 9.75 and lasts for 12 hours. Due to the *T. cacao* flower petals not growing straight, we measured petal length from micrographs in Smyth et al. (1990) and determined that when the petals are level (same height) with the short stamens they are at 60% of their final length. We therefore redefined stage 10 as when the petals are at 60% of their final length. In *T. cacao* this occurs at day 26 and lasts for one day.

Stage 11 is defined in *Arabidopsis* with the formation of stigmatic papillae at grow from the end of the stigma. This begins at day 10.25 and lasts for 1.25 days. Stigmatic papillae are the final morphological characteristic that appears on the *Arabidopsis* flower before opening. *T. cacao* does not form stigmatic papillae; however, “bump-like” lobes appear on the surface of the ovary and are the last observed morphological characteristic before flower opening. This occurs at day 27 in *T. cacao* and lasts for 2 days. Stage 12 is defined as the petals being level with the long stamens and is the last morphological event before flower opening. In *Arabidopsis* it occurs at day 11.5 and lasts for 1.75 days at which time the flower opens. For the same reasons as described for stage 10 we have defined the equivalent stage in *T. cacao* as when the petals reach 100% of their length. In *T. cacao* this occurs at day 29 and concludes with flower opening after one day.

2.4.5 Comparison of timing between *Arabidopsis* and *T. cacao*

The developmental stages of flowering in *T. cacao* and *Arabidopsis* were compared in a time corrected scale as presented in Figure 2-5 . The *T. cacao* flower developmental process takes roughly 2.26 times longer than *Arabidopsis*. When related to overall bud length obtained, it can be seen that the overall relative rate of growth is similar in the two species. This indicates that although *Arabidopsis* flowers may develop faster than *T. cacao* in absolute terms, the relative growth rates of various organs are in fact similar. For example, the ratio of final bud length in *T. cacao* (5.7 mm) to final bud

length in *Arabidopsis* (2.8 mm) and the ratio of flower bud development time in *T. cacao* (30 days) and *Arabidopsis* (13.25 days) are similar, at 2.03. So especially for early bud development, *Arabidopsis* and *T. cacao* flower buds develop at the same rate if measured in unit length per time (0.211 mm per day for both). Unfortunately, we have only compared *Arabidopsis* and *T. cacao* so it is unknown if this relationship exists between other species. Therefore, this relationship between *Arabidopsis* and *T. cacao* may reflect a commonality, or just be serendipitous.

When the durations of each stage was plotted for both *T. cacao* and *Arabidopsis*, a co-linear relationship was observed (regression slope p-value = 0.11) (Figure 2-7). It is important to note that the regression line has a positive slope that is significant, however the regression model fitted does violate the assumption that these data are independent. This is due to these data having to reach a certain maximal value. However, there is clearly a relationship between the duration of the stages between the two plants. This indicates that although the two species are morphologically different and they develop at different rates, their relative morphological programs are similar. Therefore, it is possible to postulate that the expression of genes underlying floral development are likely to follow similar developmental programs between the two species. Finally, it is important to note that all stages are in the same order between the two plants and will sum to their respective maximal growths. However, when comparing other species in this manner this may not be the case as equivalent stages may not occur in the same order as *Arabidopsis* and *T. cacao*. Therefore, the durations of equivalent stages in other species could appear very different and possibly even negative. However, we would expect all of the flower buds to be growing, regardless of the order of stages, and eventually the flower buds will reach their respective maximal sizes before opening, making different species' floral developmental programs comparable to *T. cacao* and *Arabidopsis*.

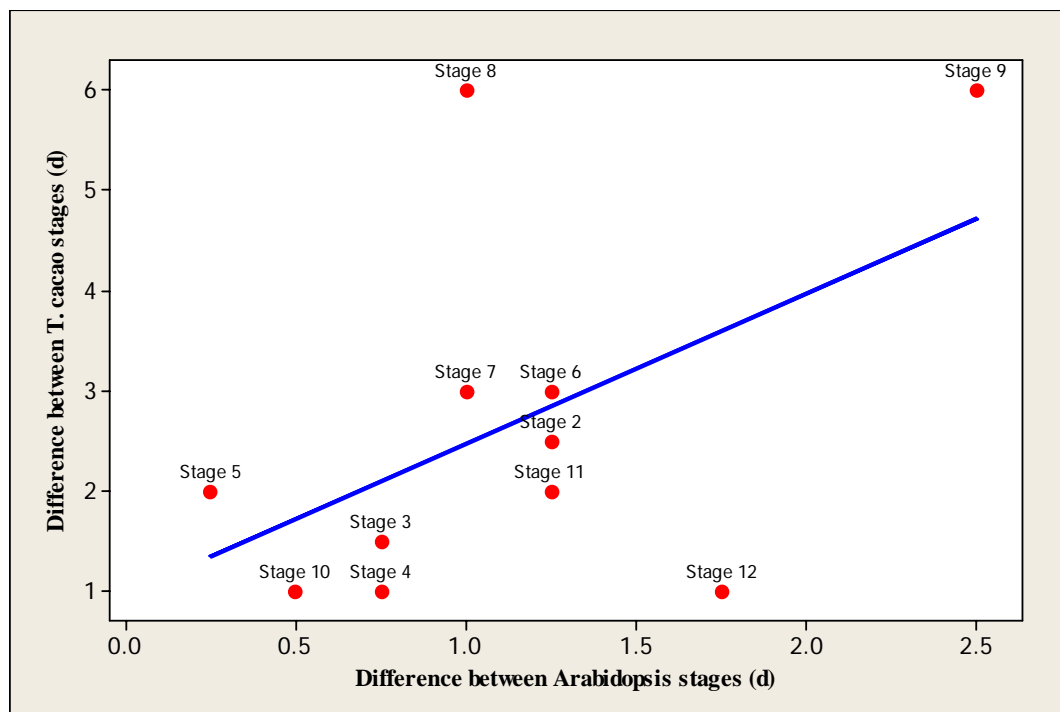


Figure 2-7: Comparison of the duration of each stage in *Arabidopsis* vs. *T. cacao*. *Arabidopsis* (horizontal axis) vs. *T. cacao* (vertical axis) from stage 2-12 as defined in Smyth et al. (1990). A weak linear relationship is observed ($p=0.11$).

2.4.6 Comparison to well characterized flowers of other species

Comparative anatomy and development can provide evidence regarding the evolution and conservation of developmental pathways. Therefore, it is informative to compare flower growth in *T. cacao* to other well characterized systems, such as: *Medicago truncatula* (barrel medic), *Brassica napus* (rape), *Antirrhinum majus* (snapdragon), *Petunia spp.*, *Lycopersicon esculentum* (tomato), *Nicotiana tabacum* (tobacco) and, *Silene latifolia*. Although these species' landmark morphological events have been characterized, they have not yet been quantified relative to time as has been done with *T. cacao* and *Arabidopsis*. In tobacco, barrel medic and tomato, flower development has been characterized by comparing morphological events to the length of

the flower bud, rather than time. The remaining species (rape, snapdragon, petunia) were only defined in terms of morphological events.

By using the published morphological landmarks of Goldberg, (1988) and Brukhin et al., (2003), we were able to compare *T. cacao* with tobacco and tomato by relating the timing of events to percentage of total length before flower opening. It is important to note that in tobacco and tomato the corolla emerges from the calyx, whereas in *Arabidopsis* and *T. cacao* it does not. This indicates that there will be differences when comparing relative total lengths as measurement will be of the petals rather than the calyx in tomato and tobacco.

From Figure 2-8, several interesting comparisons can be made across the three species in terms of percentage lengths (length of flower bud at morphological event / final length of flower bud immediately before flower opening). The first is that in all three species, the formation of sepal, petal, anther, and carpel primordia are completed within the first 10% of growth. During this initial period, petal primordia form at relatively the same percentage lengths (4% *T. cacao*, 2% tobacco, 3.5% tomato), as does the closure of the calyx in *T. cacao* and tobacco (6.3% *T. cacao*, 3.2% tobacco).

The relative timing of style formation is relatively the same amongst the three species, when converted to percent total mature size with *T. cacao* style primordia forming at 17% of total bud length, and tobacco and tomato at 13% and 20% respectively. *T. cacao* and tobacco petals also reach the top of the anthers at similar percentage lengths (12% *T. cacao* at the formation of the constriction the petal length equals the top of the anthers, 15% tomato). Finally, it appears that in all three species the focus for growth has moved beyond formation and into differentiation and elongation when 30% of final length is reached.

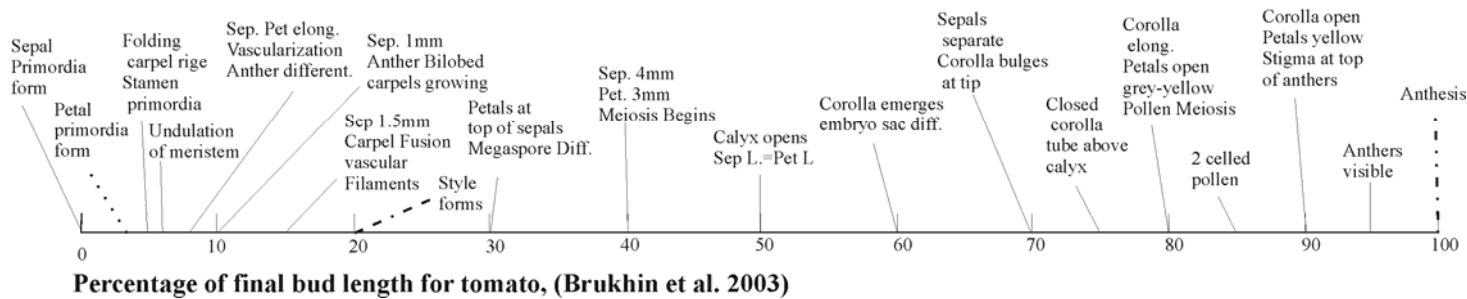
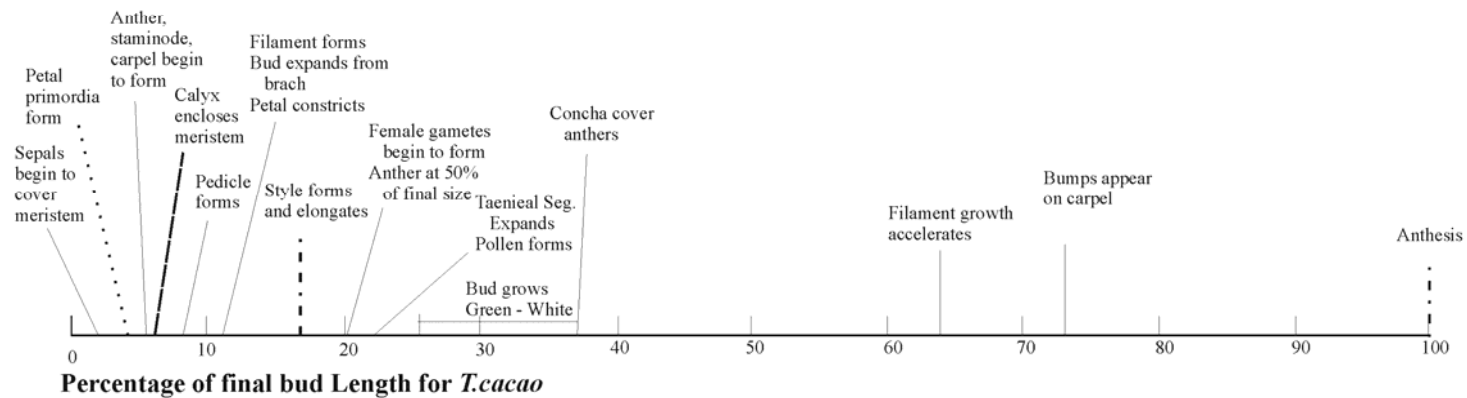
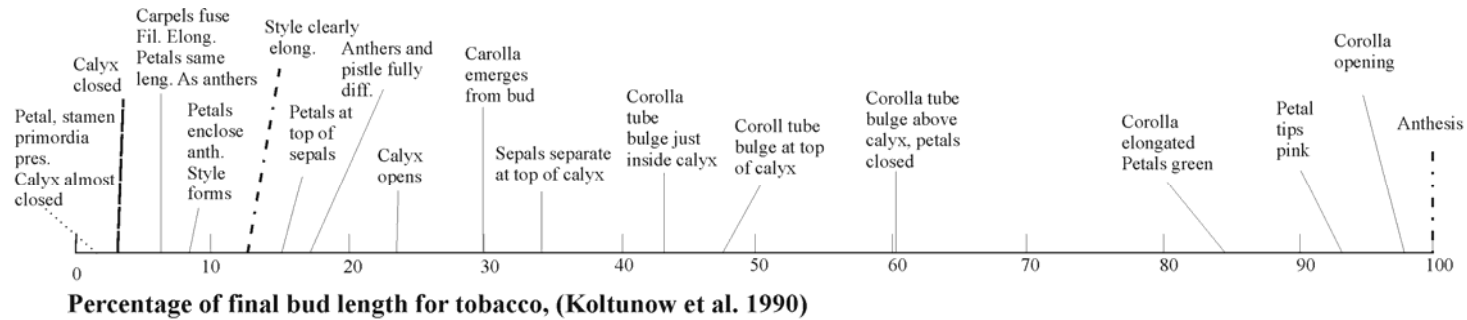


Figure 2-8: Comparison of stages of flower growth of *T. cacao*, tobacco, and tomato. Percentage of total flower bud length is used as a common axis. Similar label lines represent common stages among species.

One difference that can be observed among all of the well studied species mentioned above is in the relative patterning of the primordia. Barrel medic is unidirectional and initiates sepals first followed by “common primordia” and carpels concurrently. Once the common primordia are formed, they differentiate to form petal and sepal primordia. In rape, the order of initiation does not follow the same sequence. After the stamen primordia form, they are followed by the petal primordia along with the carpel. In *Arabidopsis*, the order is reversed with the initiation of floral organs occurring in the sequence: sepals, petals, stamens, and carpel. There is some argument in the literature as to whether or not the stamens and petals initiate together in *T. cacao*. Zamora et al. (1960) were unable to define whether the stamen and carpel primordia initiated together or not. Bayer and Hoppe (1990), consistent with our studies, suggested that each organ is initiated sequentially, although very close together. In snapdragon each whorl is initiated sequentially. In tomato, development is sequential, however the petal and stamen are initiated together. In *S. latifolia* whorls are initiated separately and sequentially. (The data on tobacco were not sufficient to make these comparisons.) Based on these observations alone, these comparisons indicate that there seems to be no correlation between evolutionary relationship and sequence of floral primordia initiation. This could in turn imply that the genes involved in initiation are controlled in different manners across the species.

As we have stated above and as stated by Buzgo et al., (2004), there is a great diversity in floral structure and relative order of floral developmental events across species. This has been illustrated by authors who have refined developmental models of floral ontogeny to fit the specifics of a given species. This results in additions and expansions to current models, for example: the addition of “D” and E’ class genes in the ABC model by Thiessin et al., (2000), and the “shifting borders” hypothesis by Kramer et al., (2004). Moreover it has become apparent that molecular data alone are not sufficient to define homology (Remane, 1952), and furthermore it is suggested that only through the combination of molecular data, detailed flower developmental information and

phylogenetic information that we will be able to accurately compare among species (Soltis et al., 1999; Soltis and Soltis, 2000; Soltis et al., 2002).

As the study of flower development advances, especially with the abundance of molecular data becoming available through genomics projects, it becomes increasingly important to have a good model of normal or typical flower development (Brukhin et al., 2003). Unfortunately, there is sparse literature providing complete information on the normal development of flowers in enough species to be able to draw generalizations. It is obvious from the literature that there is not consistency among the different morphological landmarks used by researchers among different species, nor on the basis of comparison. Furthermore, there are not enough data from comparable studies on multiple plant species to assess how generic are the developmental features of flowers in the model species such as *Arabidopsis*. With the application of genomics to the study of evolution and development, it will be important that the development and morphology of the plants under study be standardized, i.e. described in similar terms so that the molecular data can be applied and compared across species in a consistent way.

In this thesis, we have described the flower developmental program in *T. cacao* and have compared it to other species in the literature in terms of relative developmental time and size. While both of these perspectives show that flowering is indeed complex, we were able to define developmental benchmarks that are in agreement with Buzgo et al., (2004) for comparisons among species. These benchmarks will be useful in studies on conservation of flower development among all flowering plant species.

2.5 References

- Bartolome R** (1951) Cacao. The Philippine Journal of Agriculture **16**: 1-53
- Bayer C, Hoppe JR** (1990) Die Blütenentwicklung von *Theobroma cacao* L. (Sterculiaceae). Beitr. Biol. Pflanzen **65**: 301-312
- Boss PK, Bastow RM, Mylne JS, Dean C** (2004) Multiple pathways in the decision to flower: Enabling, promoting, and resetting. Plant Cell **16**: S18-S31
- Brukhin V, Hernould M, Gonzalez N, Chevalier C, Mouras A** (2003) Flower development schedule in tomato *Lycopersicon esculentum* cv. sweet cherry. Sexual Plant Reproduction **15**: 311-320
- Buzgo M, Soltis DE, Soltis PS, Ma H** (2004) Towards a comprehensive integration of morphological and genetic studies of floral development. Trends in Plant Science **9**: 164-173
- Cheeseman EE** (1932) The economic botany of cacao. Tropical Agriculture Supplement
- Goldberg RB** (1988) Plants: novel developmental processes. Science **240**: 1460-1467
- Jack T** (2001) Relearning our ABCs: new twists on an old model. Trends in Plant Science **6**: 310-316
- Jack T** (2004) Molecular and genetic mechanisms of floral control. Plant Cell **16**: S1-S17
- Kanno A, Saeki H, Kameya T, Saedler H, Theissen G** (2003) Heterotopic expression of class B floral homeotic genes supports a modified ABC model for tulip (*Tulipa gesneriana*). Plant Molecular Biology Reporter **52**: 831-841
- Kramer EM, Jaramillo MA, Di Stilio VS** (2004) Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. Genetics **166**: 1011-1023
- Meinke DW, Cherry JM, Dean C, Rounsley SD, Koornneef M** (1998) Arabidopsis thaliana: a model plant for genome analysis. Science **282**: 662, 679-682
- Meyerowitz EM, Running MP, Sakai H, Williams RW** (1998) Multiple modes of cell division control in *Arabidopsis* flower development. Symp Soc Exp Biol **51**: 19-26
- Rajamony L, Mohanakumarn N** (1991) A note on the floral biology of cacao (*Theobroma cacao* L.). South Indian Horticulture **39**: 168
- Remane A** (1952) Die Grundlagen des Natürlichen Systems der Vergleichenden Anatomie und der Phylogenetik. Geest and Portig, Leipzig, Germany
- Simpson GG, Dean C** (2002) Arabidopsis, the Rosetta stone of flowering time? Science **296**: 285-289
- Smith NJH** (1992) Tropical forests and their crops. Comstock Pub. Associates, Ithaca, N.Y.
- Soltis DE, Soltis PS, Albert VA, Oppenheimer DG, dePamphilis CW, Ma H, Frohlich MW, Theissen G** (2002) Missing links: the genetic architecture of flower and floral diversification. Trends in Plant Science **7**: 22-31
- Soltis ED, Soltis PS** (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Molecular Biology **42**: 45-75

- Soltis PS, Soltis DE, Chase MW** (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402-404
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H** (2000) A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149
- Walker-Larsen J, Harder LD** (2000) The evolution of staminodes in angiosperms: Patterns of stamen reduction, loss, and functional re-invention. *American Journal Of Botany* **87**: 1367-1384
- Wood GAR, Lass RA** (1985) *Cocoa*, Ed 4. Longman Scientific and Technical, New York
- Zamora PM, Orlando NM, Capinpin JM** (1960) Ontogenic and embryological studies in *Theobroma cacao* Linn. *Philippine Journal of Agriculture* **43**: 613-636

Chapter 3

Design and Testing of Generalized PCR Primers for *Theobroma cacao* and other Rosidae Plant Species



Air

3.1 Introduction

Gene discovery is a central component of all genomics research. The advent of the Polymerase Chain Reaction (PCR), and refinement of methods for cDNA and genomic library construction and screening has been greatly simplified and applied in non-model species with little or no initial genomic information. Currently, the most common methods include the use of DNA libraries, the sequencing of entire genomes, and the amplification of DNA fragments using degenerate primers.

The degenerate primer method uses DNA sequence information from other species to create degenerate oligonucleotide primers for PCR to amplify portions of specific genes of interest. Degenerate primers are pools of short (17-30 bp) oligonucleotides, which vary in base composition at specific positions. There are several benefits gained from using a PCR-based approach that are enhanced when using degenerate primers. The first being if the primers designed amplify portions of specific genes of interest in the target species then they most likely can be used to amplify other closely related gene segments from the species of interest. The resulting DNA fragments can be analyzed for Single Nucleotide Polymorphisms (SNPs) or allelic variations and for their potential use as molecular markers. The second major benefit is that the primers can be tested for use in evolutionarily closely related species. Having this resource at hand can aid other researchers in the study of their own species. Finally, the orthologous gene sequences among related species can be delineated and genetic distances can be calculated and tested.

The approach used in these experiments took advantage of publicly available *Arabidopsis* and *Gossypium hirsutum* (cotton) DNA sequences to create degenerate PCR primers that were in-turn used to amplify cDNA fragments from genes in *Theobroma cacao* mRNA. This experimental approach has been used extensively before (Kim et al., 2002; Prosser et al., 2002). It is intended that the target genes

identified here will act as a set of control genes for hybridization experiments such as RT-PCR, northern blots, and *in-situ* hybridization.

This method allowed the amplification of specific cDNAs of interest from *T. cacao* quickly and effectively, rather than having to use a “blind” EST type approach. *Arabidopsis* and cotton were used as the template species to design degenerate primers suitable to amplify cDNAs from *T. cacao*. Both of these species were chosen because of the rather large respective bodies of DNA sequence data available for each. Cotton has a publicly available fiber database with around 9,000 unigene sequences (www.genome.clemson.edu). *Arabidopsis* has its entire genome sequence available (Mayer et al., 1999; Salanoubat et al., 2000; Tabata et al., 2000).

A second reason *Arabidopsis* and cotton were chosen as template species for primer construction was because *T. cacao* and cotton belong to the plant family Malvaceae (Whitlock and Baum, 1999; Whitlock et al., 2001). Additionally, *Arabidopsis*, cotton, and *T. cacao* all belong to eurosids II (APGII, 2003). For phylogenetic comparisons, we also used information gained from the *Populus* genome that was recently released (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>). *Populus* is in the eurosids I, and was treated as an out-group for phylogenetic analyses.

The objectives of the research presented here were to: 1) isolate and characterize a set of fifty specific *T. cacao* cDNAs, 2) identify the utility of these primers as molecular markers, 3) to gauge their use in other closely related species (eurosids II), 4) test their applicability to amplify other gene family members, and 5) to determine the nuclear DNA sequence similarity between a subset of genes common to *Arabidopsis*, cotton, and *T. cacao*.

3.2 Methods

3.2.1 Preliminary data

To verify the expected phylogenetic relationships of our study we performed a preliminary analysis with two different genes. The DNA sequences of the chloroplast genes *atpB*, and *nadH* were obtained for *T. cacao*, cotton, *Arabidopsis*, and *Populus* from NCBI, the National Center for Biotechnology Information (AP000423, AJ233090, AF031446, AF209658, AJ235465, AF238063, AF287916, U55340, AF315314, AJ429105). The sequences were aligned using MUSCLE v.3.552 (Edgar, 2004), and a phylogenetic tree was constructed using the analysis program MEGA 2 (Neighbor Joining, Kumar 2 Parameter, complete deletion of swapped sites, 100 bootstrap replicates were performed) (Kumar et al., 2001).

3.2.2 RNA extraction and cDNA construction from *T. cacao*

RNA was extracted from leaf tissue following Antunez de Mayolo, (2003). The protocol is as follows: Total RNA was extracted from PSU-SCA6 plants. RNA extractions initially followed Chang et al. (1993). Frozen leaf samples were ground with a mortar and pestle in liquid nitrogen to a fine powder. Approximately 1 gm tissue was re-suspended in 15 mL extraction buffer containing 2% CTAB (hexadecyltrimethylammonium bromide), 2% PVP (polyvinylpyrrolidone), 100 mM Tris base, 25 mM diNaEDTA, 2 M NaCl, 0.5 gmL⁻¹ spermidine, and 2% β-mercaptoethanol, then incubated at 65 °C for 10-30 min. 15 mL chloroform was added to the sample, which was then mixed and centrifuged for 20 min at 8000 x g. The top aqueous layer was transferred to tubes containing 5 mL 8 M LiCl (DEPC treated), and stored overnight at 4 °C, followed by centrifugation for 30 min at 8000 x g. Following LiCl RNA precipitation, RNA was recovered following the Qiagen RNeasy based method (Qiagen, Valencia, CA) as follows: pellets were re-suspended in 500 μL RLC Qiagen buffer, samples in RLC buffer

were added 0.5 volumes ethanol (100%) was added and the mix was transferred to a RNeasy mini column, and centrifuged for 15 sec at 8000 x g. Samples were washed with 700 μ L RW1 buffer, then again using RPE buffer; after each step samples were centrifuged (15 sec at 8000 x g). RNA samples were eluted with two 70 μ L RNase-free water aliquots.

cDNA was made from total RNA using the RETROscript cDNA kit (Ambion, Austin, TX) using the manufacturer's instructions and reagents. Briefly, 1-2 μ g of RNA was mixed with 2 μ L of random decamer solution, this mixture was heated to 70 °C, and then placed on ice. 10 x reverse transcription buffer was then added to 1X dilution, as well as 4 μ L dNTP mix, 1 μ L RNase inhibitor and 1 μ L of reverse transcriptase. This mixture was incubated at 42 °C for one hour then heated to 95 °C for 10 min to stop the reaction.

3.2.3 Selection of genes, primer design, cloning, and sequencing of *T. cacao* cDNAs

Fifty genes were chosen for amplification from *T. cacao* based on strong sequence similarity ($>1e-20$) between *Arabidopsis* and ESTs in the Clemson cotton database (www.genome.clemson.edu) (Appendix Ai). Genes were selected based on two factors: 1) the sequence similarity between *Arabidopsis* and cotton of which, the genes were ranked according to e-value, and 2) the proportions of genes in each NCBI Clusters of Orthologous Groups (COGS) category that have been annotated from *Arabidopsis* genome sequence and reported in the COGS database (Tatusov et al., 1997). At least one and no more than five genes in the 'set of 50' represented each COGS category. Seven genes were selected that were based on controls used in commercial microarray chips (TAKRA-Bio Inc., Otsu, Japan), these included genes that were expected to be constantly expressed to some degree in all tissues across all species. Two of these genes would be expected to be present in all kingdoms of life (β -tubulin, and actin), three were found only in eukaryotes (eukaryotic initiation factor, GADPH, and ubiquitin), and the

remaining two were only found in plant cells (cellulose synthase, and rubisco). We also attempted one flower specific gene (*LEAFY*) and one hormone response gene. In total, primers were created for 59 genes.

Primers were designed by first aligning the orthologous genes using the Vector NTI (ver. 7.0) AlignX software (Invitrogen, Carlsbad, CA). Using this software a consensus sequence was produced, and any nucleotides that did not agree across the orthologous genes were replaced by N's. The consensus sequence was processed using Primer 3 software (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi), under the constraints of an annealing temperature of 60°C, and as few as possible N's in the resulting primers (less than 3). The resulting primers were synthesized commercially (Integrated DNA Technologies, Coralville, IA), as listed in Appendix Aii.

T. cacao cDNA was used as template to amplify PCR fragments using the conserved sequence primers. In short, amplification was carried out in a reaction mix containing 2 µL of cDNA template, 2.5 µL of 10x Buffer (PGC Scientific, Frederick, MD), 2.5 µL of 2 mM dNTPs, 1 µL of each forward and reverse 10 mM primer, 1 µL of *Taq* DNA Polymerase (5u/µL)(PGC Scientific, Frederick, MD). Distilled water was used to bring the reaction mix to a total volume of 25 µL. The PCR mix was cycled in a Perkin-Elmer 9600 thermocycler (Boston, MA) for 4 min at 95 °C, then thirty cycles of 94 °C for 30 sec, 50 °C for 30 sec, and 72 °C for 40 sec. Finally, the reaction was held at 72 °C for 7 min and then held at 4 °C. PCR products were separated on 2% agarose gels and fragments that matched the expected fragment sizes predicted by the Primer 3 software were excised from the gel and purified using the QiaQuick Gel Extraction Kit (QIAGEN Inc. Valencia, CA) following the manufacturer's instructions. An example of a typical amplification may be seen in Figure 3-1.

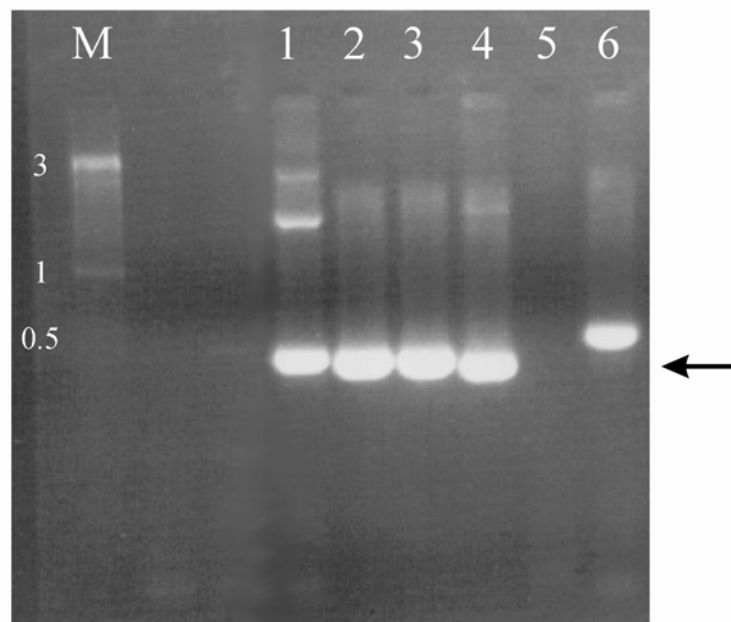


Figure 3-1: Example of a typical PCR amplification using heterologous primers. “M” lane represents a DNA Ladder II (PGC Scientific, Frederick, MD), sizes in kb are indicated next to the marker. Lanes “1”-“4” represent *CELLULOSE SYNTHASE* cDNA amplification products. Lane “5” represents a no template negative control. Lane “6” represents a 600 bp positive control consisting of a *T. cacao nadH* amplification product. Arrow indicates both the size range that was expected, and the bands that were excised.

The DNA fragments were cloned into the TOPO[®] vector using the Invitrogen TOPO[®] cloning kit (Invitrogen, Carlsbad, CA) as per manufacturer instructions. The resulting vectors were transformed into competent *E. coli* cells using heat shock at 42 °C for 30 seconds and then grown in SOC medium at 37°C for one hour. The cells were spread on LB+AMP plates and were grown overnight. Eight of the resulting colonies were selected and the plasmids were purified using the QIAprep Spin Miniprep Kit (QIAGEN Inc. Valencia, CA). These were sequenced using a PCR based method with the standard M13 forward primer and visualized using a Perkin-Elmer 3700 DNA sequencer. The above protocol was repeated for all primers. Representative sequence for each cDNA is shown in Appendix B.

3.2.4 Data analysis

All sequences were compared against the database at NCBI using BLAST and sequences that did not match the original gene were discarded. Of those sequences that did match the original gene, vector sequence was trimmed, imported into Vector NTI™ (Invitrogen, Carlsbad, CA) and the percentage nucleotide difference was calculated relative to the original cotton and *Arabidopsis* sequences using Vector NTI™ (Invitrogen, Carlsbad, CA). The files were imported to MEGA 2.1 (Kumar et al., 2001) and p distances were calculated. P-distance is the proportion of nucleotide sites at which two sequences being compared are different. It is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared with no correction for multiple changes made at the same site (Kumar et al., 2001).

As a further test, subsets of 29 *Populus* orthologous genes were selected from the *Populus* genome sequence (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>). These were aligned with the respective cotton, *Arabidopsis*, cotton and *T. cacao* sequences using ClustalW (Chenna et al., 2003) (Alignments may be found at <http://guiltinanlab.cas.psu.edu/Research/Cocoa/flowers.htm>). These data were exported into FASTA format and imported into MEGA 2.1 (Kumar et al., 2001) for the construction of phylogenetic trees using the following settings: Neighbor Joining, p-distance-parameter, deletion of swapped sites, bootstrap of 1000 replications. The resulting trees are presented in Appendix Ci.

Additional comparisons were carried out on these sequences by selecting the 19 cDNAs, which had separate *T. cacao* sequences for the same gene. These sequences were compared to the PlantTribes database (<http://fgp.bio.psu.edu/cgi-bin/tribedb/tribe.cgi?action=home>, stringency level 3) and potential gene family members were identified from both *Arabidopsis* and *Oryza sativa v. japonica*. These sequences were saved in FASTA format, combined with the *T. cacao*, cotton, *Arabidopsis* and *Populus* data described above, and the sequences were aligned using MUSCLE v.3.52 (Edgar,

2004) (Alignments may be found at <http://guiltinanlab.cas.psu.edu/Research/Cocoa/flowers.htm>) imported into MEGA 2.1 (Kumar et al., 2001). Phylogenetic trees were constructed using the following settings: Neighbor Joining, Kumar 2 Parameter, deletion of swapped sites, bootstrap of 1000 replications with replacement. The Kumar 2 Parameter was selected as it measures genetic distance but is able to correct for multiple substitutions (Kumar, 1980).

Trees where the *T. cacao* sequences that were separated phylogenetically by *Arabidopsis* genes were selected and the *O. sativa* sequences were removed from the analysis (the *O. sativa* gene family members had diverged too far to get a good alignment with the respective *T. cacao*, *Arabidopsis*, cotton and *Populus* sequences (C. dePamphilis, pers. comm.). These sequences were re-aligned using MUSCLE v.3.52 (Edgar, 2004) (Alignments may be found at <http://guiltinanlab.cas.psu.edu/Research/Cocoa/flowers.htm>) then imported into MEGA 2.1 (Kumar et al., 2001). Phylogenetic trees were constructed using the following settings: Neighbor Joining (Kumar 2 Parameter), bootstrap of 100 replications with replacement.

To assess possible SNP sites, *T. cacao* sequences of the same gene were compared whenever at least four clones had been sequenced. A putative SNP was identified if a base substitution was observed in an identical position in sequences that had at least four representative samples of that gene, and two of those samples showed the identical base substitution.

3.2.5 Use of degenerate primers in other species

To assess the applicability of these primers in other species of eurosids I and II, PCR was carried out on DNA obtained from the dePamphilis laboratory (Department of Biological Sciences, The Pennsylvania State University). The species selected were: *Carica papaya* (eurosids II), *Aesculus glabra* (Eurosids II), *Citrus limon* (Eurosids II), and *Cucumis sativus* (eurosids I).

PCR was carried out using the REDTaq™ PCR kit (Sigma-Aldrich, ST. Louis, MO). The reaction mix was as follows: 5 µL of REDTaq™, 1 µL of each forward and reverse primer (10 mM), 1 µL of DNA (5 ng/µL). After an initial denaturation step, the reaction was cycled in a Perkin-Elmer 9600 thermocycler (Boston, MA) for thirty cycles at 94 °C for 30 sec, 50 °C for 30 sec, and 72 °C for 40 sec. Finally, the reaction was held at 72 °C for 7 min and then held at 4 °C. PCR products were separated on 2% agarose gels. The gels were scored by recording successful reactions (band(s) produced) as a “1”, and unsuccessful reactions as a “0”. Minitab® (ver. 14) (Minitab, State College, PA) was used to perform cluster analyses (single linkage method) to identify relationships among the species based on the PCR results obtained. Our assumption was that as genetic distance increased, proportionally larger numbers of genes would fail to amplify.

3.3 Results

3.3.1 Preliminary studies

In preliminary studies we compared the chloroplast genes *ndhF* and *atpB* (sequences derived from Whitlock, (2002) and Bayer et al., (Unpublished)) in cotton, *Arabidopsis* and *T. cacao*. We found that the *Arabidopsis* and *T. cacao* genes had an average of 91% similarity, and the cotton and *T. cacao* genes had a 98% similarity. This was in agreement with taxonomic studies that put *Arabidopsis*, cotton and *T. cacao* in the eurosids II, and *T. cacao* and cotton together in the family Malvaceae (Whitlock, 2002). We also created a dendrogram with the above three species and used sequences from *Populus* (member of eurosids I) to root the tree (Figure 3-2).

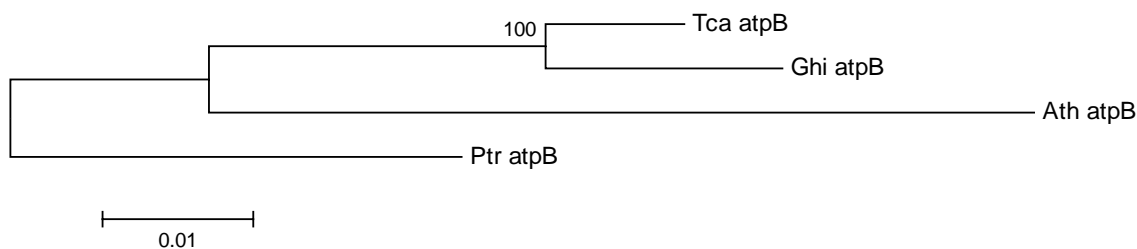


Figure 3-2: Phylogenetic tree of *T. cacao* (*Tca*), cotton (*Ghi*) *Arabidopsis* (*Ath*), and *Populus* (*Ptr*) based on chloroplast *atpB* sequence. Bootstrap support of 100 replications is shown.

3.3.2 *T. cacao* cDNA isolation

By creating degenerate primers based on cotton and *Arabidopsis*, we were able to amplify, clone, and sequence portions of forty-two *T. cacao* cDNAs. These genes were selected based both on the sequence similarity between *Arabidopsis* and cotton and on the number of genes represented in each functional category of the NCBI COG (Clusters of Orthologous Groups) database.

Out of the fifty-nine genes attempted, forty-two putative *T. cacao* cDNAs were correctly amplified (as assessed by comparison to their respective genes at the NCBI database) using this method. Out of the fifteen that did not work; five did not have a good enough conserved regions between *Arabidopsis* and cotton to produce primers. Four did not amplify from *T. cacao* with the primers designed, or did not sequence correctly. Six did not match correctly to their expected genes using the BLAST program (i.e. we amplified an alternative gene with a high BLAST value), and three matched to the correct gene but did not have a high enough BLAST alignment score to allow confidence in the result. These results are summarized in Table 3-1, and representative sequences are reported in Appendix B.

3.3.4 Data analysis

We compared the *T. cacao* sequences obtained to the original cotton and *Arabidopsis* gene sequences from which the primers were derived. Overall averages of DNA sequence identity between optimally aligned pairs of orthologous sequences from 285 different cDNAs were calculated (193 *T. cacao*, 42 cotton, and 50 *Arabidopsis* cDNA sequences, respectively). When the same *T. cacao* cDNAs were compared to each other we observed an average of 95.9% (s.d.= 6.6%) identity. When we compared *T. cacao* sequences with cotton cDNAs, we observed an average of 88.1% (s.d.= 9.3%) identity. When *T. cacao* and *Arabidopsis* were compared, we observed an average of 78.7% (s.d.= 8.9%) identity, and for cotton and *Arabidopsis* we saw an average of 81.4% (s.d.=4.3%) identity. These data were expressed in terms of p-distances in MEGA 2.1 (Kumar et al., 2001) and were illustrated in Figure 3-3. These distances showed that *Arabidopsis* and *T. cacao*, and *Arabidopsis* and cotton had similar distances from each other (p-dist.=0.189, s.d=0.033; p-dist.=0.185, s.d.=0.034, respectively), and *T. cacao* and cotton were closer in p-distance to each other (p-dist.=0.095, s.d.=0.04) than to *Arabidopsis*. These results are in agreement with the results obtained in the preliminary comparison of the chloroplast genes *ndhF* and *atpB* (Whitlock, 2002).

3.3.4.1 SNPs

Putative SNPs were identified in *T. cacao* genes where the same nucleotide was substituted in the identical position for at least two instances in different *T. cacao* sequences for a particular gene. Of the forty two cDNAs, 17 had putative SNP substitutions according to the criteria defined in section 3.2.4. The cDNAs in which these substitutions were observed are recorded in Table 3-1.

Table 3-1: Genes extracted from *T. cacao* using degenerate primers. Columns 1 and 2 list the putative functional gene category and assignments. Column 3 lists the fragment sizes from the prediction based on the primer design phase of the method, the observed fragment sizes are calculated from the sequenced clones. Columns 4 and 5 list the BLAST scores and the accession number of the top match. Column six lists the sequences with putative SNP sites based on the comparison of multiple clones (N/A=did not meet criteria for SNP selection).

Gene Category	Functional Assignment	Fragment Size (bp) Pred/Obs	BLAST Score	Top BLAST Hit (Accn No.)	Possible RFLP/ SNP sites
Control	β -tubulin				
Control	Actin				
Control	Eukaryotic Initiation factor				
Control	GADPH				
Control	Ubiquitin				
Control	Cellulose Synthase				
Control	Rubisco (small Subunit)				
Amino acid metab. & trans.	Glutamate Decarboxylase	289/288	4e-65	AF353615	N
Amino acid metab. & trans.	Glutamate Dehydrogenase	241/243	1e-34	U48695	N
Amino acid metab. & trans.	Ketol-Acid Reductoisomerase Precursor	255/273	9e-26	Y17796	Y
Amino acid metab. & trans.	Putative 3-Isopropylmalate Dehydratase Large Subunit	361/361	3e-63	NM_117417	N/A
Carbohydrate Metab & Trans.	Adenosine Kinase	353/353	3e-54	NM_111817	Y
Cell Div & Xome Part.	Actin 12	250/251	1e-120	AF124139	Y
Cell Div & Xome Part.	Actin 2 7	280/280	1e-123	AY096397	N
Cell Div & Xome Part.	Actin 58	327/327	2e-60	U60500	Y
Coenzyme Metabolism	Ubiquitin-Activating Enzyme	307/307	2e-33	AC004165	Y
Energy Prod. & Conversion	(S)-2-Hydroxy-Acid Oxidase, Peroxisomal	299/308	4e-67	AF082874	N/A
Energy Prod. & Conversion	Betaine-Aldehyde Dehydrogenase Precursor	326/331	7e-.9	NM_069653	N/A
Energy Prod. & Conversion	Isocitrate Dehydrogenase	302/301	3e-72	AF367443	Y
Energy Prod. & Conversion	Pyruvate Dehydrogenase E1 Component Alpha Subunit	226/226	2e-42	AF360306	N
Energy Prod. & Conversion	Turgor-Responsive Protein 26g	248/248	8e-17	AY102145	Y
Function Unknown	Putative Udp-Glucose	241/231	7e-14	AY096627	N
General Function Pred. only	Esterase D	276/276	2e-48	NM_129716	Y
Inorg. Ion Trans. & Metab	Nadph-Cytochrome P450 Reductase	272/273	2e-73	AF302498	N/A
Inorg. Ion Trans. & Metab	Putative Ascorbate Peroxidase	332/333	7e-70	AF159360	Y
Lipid Metabolism	Chalcone Synthase 1	242/202	2e-54	AY087778	Y
Lipid Metabolism	Putative Beta-Ketoacyl-Coa Synthase	376/376	9e-34	NM_128221	Y
Molecular Chaperones	EndI3 Protein	245/245	9e-54	AY087749	Y
Molecular Chaperones	Dtdp-Glucose 4,6-Dehydratase	234/233	4e-37	NM_130333	N/A
Molecular Chaperones	T-Complex Protein 1, Epsilon Subunit	311/310	2e-48	AY075611	N/A
Nucleotide Metab. & Trans.	Amidophosphoribosyltransferase Precursor	277/274	1e-65	AC007195	N/A
Nucleotide Metab. & Trans.	Ethylene-Inducible Protein Hever	289/287	9e-51	M88254	Y
Outer Memb, Cell Wall Biogen.	Cellulose Synthase Catalytic Subunit	439/439	1e-20	NM_100153	N/A
Outer Memb, Cell Wall Biogen.	Udp-Glactose 4-Epimerase	316/450	9e-52	AJ005081	Y
Replicat, Repr, Recomb	Alcohol Dehydrogenase Class III	308/308	8e-36	U77637	Y
Secretion And Motility	Preprotein Translocase Secy Subunit	385/386	3e-39	NM_127427	N
Transl, Ribosomal Struct.	40s Ribosomal Protein S3	257/257	3e-55	AY46041	Y
Transl, Ribosomal Struct.	CAB11	291/286	6e-52	X57706	N/A
Transl, Ribosomal Struct.	Elongation Factor 2	256/233	3e-73	AY035011	Y
Transl, Ribosomal Struct.	Hypothetical 47.9 Kda Protein (Elongation Factor, by BLAST)	305/256	8e-82	NM_104485	N/A
Hormone Response	Ethylene	304/180	5e-36	AJ309394	N/A
Flower Specific	Leafy	403/403	9e-63	AF450278	N

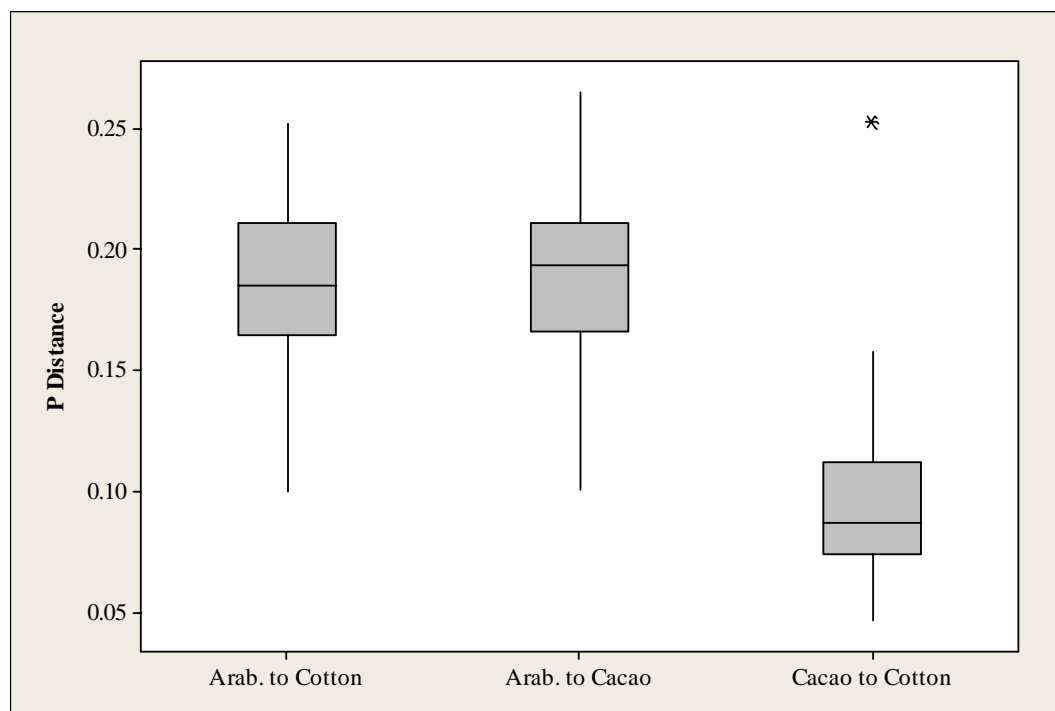


Figure 3-3: Plots depicting p-distance among *T. cacao*, cotton and *Arabidopsis* based on the combined data from all forty-three extracted cDNAs (n=193 *T. cacao*, n=42 cotton, n=50 *Arabidopsis*). Each box and whisker plot show the means, upper and lower quartile among the *Arabidopsis*/cotton, *T. cacao*/cotton, and *T. cacao*/*Arabidopsis* comparisons. It can be easily seen that the *Arabidopsis*/cotton and *Arabidopsis*/*T. cacao* have similar distances, while *T. cacao*/cotton have a much lower p-distance reflecting the fact that *T. cacao* and cotton are more closely related than either is to *Arabidopsis*. * indicates an outlier

3.3.4.1 Phylogenetic analysis

Twenty nine genes were selected for further analysis and by comparing dendrograms for each species it was observed that 24 of the 29 phylogenies were in accordance with what is expected based on our current understanding of these species phylogenetic relationships (Whitlock and Baum, 1999; Whitlock, 2002). A general example of the expected phylogeny is shown in Figure 3-4. This figure shows that *T.*

cacao and cotton are the most closely related of the four species, which could be expected because they are both classified in the same family (Malvaceae). *Arabidopsis* is expected to be more closely related to *T. cacao* and cotton since it is in the same major rosid group (eurosids II), than to *Populus* which is included as an out-group as it belongs in a different rosid group (eurosids I) (APGII, 2003).

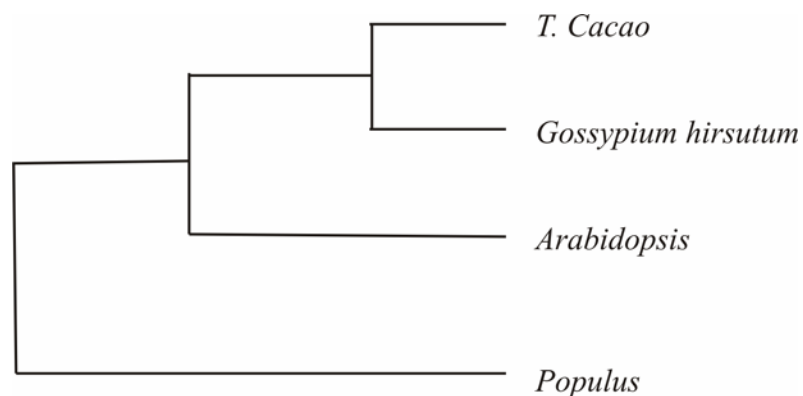


Figure 3-4: Generalized dendrogram of expected phylogenetic relationships among *T. cacao*, cotton, *Arabidopsis* and *Populus* (APGII, 2003).

The dendrograms presented in Appendix Ci showed differences between possible alleles, and simple sequence differences. Of the 29 cDNAs observed 10 were represented by single *T. cacao* sequences and were not able to be used for this analysis; 19 had two or more different *T. cacao* sequences of the same gene. These could represent different gene families, alleles, PCR error, or sequence read errors. The 19 cDNAs were compared to potential gene family members from *Arabidopsis* and *O. sativa* (Appendix Cii). It was found that five *T. cacao* genes had additional *Arabidopsis* gene family members clustering with the different *T. cacao* sequences for the respective gene. These five genes were: glutamate decarboxylase, ketol-acid reductoisomerase precursor, actin 12, actin58, and cellulose synthase and their respective phylogenetic trees are presented in Appendix Ciii.

3.3.5 Fragment amplification in other flowering plant species

PCR amplification, using a subset of 34 primers, was carried out on DNA isolated from *T. cacao*, *Carica papaya*, *Aesculus glabra*, *Citrus limon*, and *Cucumis sativus*. Successful amplification was recorded if products of appropriate size occurred in each species respective to each primer (Table 3-2). In one case (amidophosphoribosyltransferase precursor), there was no amplification product observed in *T. cacao* and so this gene was disregarded from further analysis. From the results in Table 3-2 it can be seen that *Citrus limon*, *Carica papaya*, and *Aesculus glabra* produced at least one amplification product (75%, 78%, and 72%, respectively) for the primers tested. Using *C. sativus* DNA template, 51% of the primer pairs yielded amplification product. Based on our current understanding of the phylogenetic relationships of these species, these results are consistent with *Citrus limon*, *Carica papaya*, and *Aesculus glabra* (eurosids II) being more closely related to *T. cacao* than *Cucumis sativus* (eurosids I) (APGII, 2003).

To further check the validity of these results a cluster analysis was performed using Minitab® V.14.1 software (Minitab, State College, PA) with the above data (Table 3-2, Figure 3-5). Cluster analysis showed the expected relationships according to the respective phylogenies of these species reported by APG II (APGII, 2003).

Table 3-2: PCR results using *T. cacao* degenerate primers in reactions containing DNA template isolated from *Citrus limon*, *Carica papaya*, *Cucumis sativus*, *T. cacao*, and *Aesculus glabra* (1, Positive; 0, Negative).

	<i>T. cacao</i>	<i>C. sativus</i>	<i>C. papaya</i>	<i>A. glabra</i>	<i>C. limon</i>
Glutamate Decarboxylase	1	1	1	1	1
Glutamate Dehydrogenase	1	1	1	1	1
Ketol-Acid Reductoisomerase Precursor	1	0	0	0	0
Putative 3-Isopropylmalate Dehydratase Large Subunit	1	1	0	1	1
Glutamate Decarboxylase	1	1	1	1	1
Adenosine Kinase	1	1	1	1	1
Actin 12	1	0	1	1	1
Actin 27	1	1	1	1	1
Actin 58	1	1	1	1	1
Ubiquitin-Activating Enzyme (S)-2-Hydroxy-Acid Oxidase, Peroxisomal	1	0	1	1	1
Betaine-Aldehyde Dehydrogenase Precursor	1	0	1	0	0
Isocitrate Dehydrogenase	1	1	1	1	1
Pyruvate Dehydrogenase E1 Component Alpha Subunit	1	1	1	1	1
Turgor-Responsive Protein 26g	1	0	0	0	0
Putative Udp-Glucose	1	0	1	1	1
Chalcone Synthase 1	1	0	1	1	1
Putative Beta-Ketoacyl-Coa Synthase	1	1	1	0	0
Dtdp-Glucose 4,6-Dehydratase	1	0	0	0	0
End13 Protein	1	0	0	1	1
T-Complex Protein 1, Epsilon Subunit	1	1	1	0	0
Amidophosphoribosyltransferase Precursor	0	0	0	0	0
Ethylene-Inducible Protein Hever	1	0	0	0	0
Esterase D	1	0	0	0	0
Cellulose Synthase Catalytic Subunit	1	1	1	1	1
Udp-Glactose 4-Epimerase	1	0	1	1	1
NMTR1	1	0	1	1	1
TRANSCR1	1	1	1	1	1
Alcohol Dehydrogenase Class III	1	1	1	1	1
Preprotein Translocase Secy Subunit	1	0	1	0	1
β -tubulin	1	1	1	1	1
Actin	1	1	1	1	1
Eukaryotic Initiation Factor	1	0	1	1	1
Leafy	1	0	1	1	1
Total	33	17	26	24	25

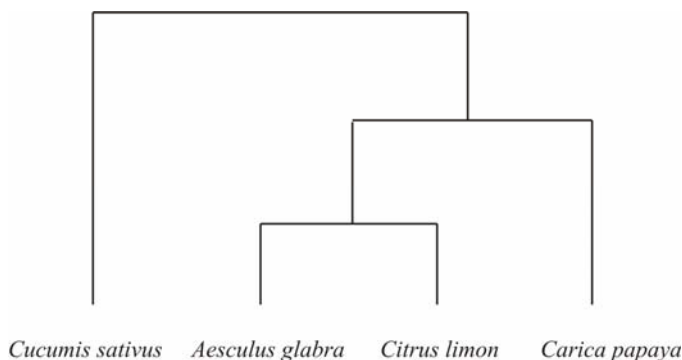


Figure 3-5: Dendrogram showing cluster analysis of PCR using *T. cacao* degenerate primers in *Citrus limon*, *Carica papaya*, *Cucumis sativus*, *T. cacao*, and *Aesculus glabra*

3.4 Discussion

3.4.1 Relationships of *T. cacao*, cotton and *Arabidopsis* and gene families

Based on a comparison of the sequence of 42 cDNA fragments, it was found that *T. cacao* differed in coding sequence from *Arabidopsis* an average of 22% (p-distance of 0.189, s.d.=0.033), and cotton 12% (p-distance of 0.095, s.d.=0.04).

Furthermore, when phylogenetic trees were constructed with our data (24 out of 29 genes tested), it was seen that the *Arabidopsis*, *T. cacao*, cotton, and *Populus* phylogenies were in agreement with our current understanding of the phylogenetic relationships of these respective species. Typical bootstrap support (based on 1000 replications with replacement) ranged from 28 – 100 with an average of ~70. Additional evidence to support these observations comes from data presented from chloroplast gene sequences which show similar phylogenies (Whitlock and Baum, 1999; Whitlock et al., 2001).

It was also noted that the p-distance between *T. cacao* and cotton was lower indicating that they were more closely related to each other than either *T. cacao* or

cotton were to *Arabidopsis*. A further line of research based off this result would be to test if synonymous distances are consistent between *Arabidopsis* and cotton, and *Arabidopsis* and *T. cacao* for all genes individually. This would provide a test of the consistency of the molecular clock across varied nuclear genes and would indicate whether it moves at the same rate between the family members cotton and *T. cacao* with respect to *Arabidopsis*.

In cases where different sequences were observed from a single cDNA we investigated the potential of identifying multi-gene families. Gene family members for *T. cacao* were identified by computing phylogenetic trees that had potential gene family members for *Arabidopsis* (identified by the PlantTribes database (<http://fgp.bio.psu.edu/cgi-bin/tribedb/tribe.cgi>)). The five *T. cacao* genes that had suspected gene family members were: glutamate decarboxylase, ketol-acid reductoisomerase precursor, actin 12, actin58, and cellulose synthase (Appendix Ciii).

The *Arabidopsis* glutamate decarboxylase gene family consists of seven potential gene family members (Appendix Ciii). When clustered with *T. cacao*, cotton and *Populus*, it is found that of the three *T. cacao* sequences, two (TcaAA2c and TcaAAA2f) are the same gene and differences between them may be accounted for by PCR error, sequencing error, or allelic differences. They cluster with the cotton sequence and with the *Arabidopsis* sequence that was used to create the respective primer. The third *T. cacao* sequence (TcaAAA2e) is found to be more similar to alternate *Arabidopsis* sequences. These data provide some evidence that *T. cacao* has at least two gene family members of glutamate decarboxylase and that possibly a single gene in the common ancestor of *Arabidopsis*, cotton and *T. cacao* had undergone duplication in the lineage leading to *T. cacao*. However, this hypothesis will require further investigation to provide evidence for *T. cacao* gene family members.

In the case of ketol-acid reductoisomerase precursor, there are six potential gene family members in *Arabidopsis* (Appendix Ciii). We observed that the *T.*

cacao sequences isolate into two separate clades. The first clade contains a single *T. cacao* sequence (TcaAAA4b) that clusters with the respective cotton and *Arabidopsis* sequences that were used to make primers, as well as *Populus*, and an additional *Arabidopsis* gene. The second clade contains four *T. cacao* sequences (TcaAAA4f, TcaAAA4a, TcaAAA4e, and TcaAAA4g) that cluster with an *Arabidopsis* gene (At1g31480). The three *T. cacao* sequences (TcaAAA4f, TcaAAA4a, and TcaAAA4e) have slight differences in sequence but are possibly the same gene and may be accounted for by PCR error, sequencing error or allelic differences. The fourth sequence (TcaAAA4g) is clearly different, as it does not cluster with the other *T. cacao* genes. These results indicate that there are at least two (possibly three) different gene family members of ketol-acid reductoisomerase and that the common ancestor for *T. cacao* may have been a single gene before the separation of eurosids I and eurosids II.

The actin 12 and actin 58 phylogenetic trees show low bootstrap values (of which are >50%) (Appendix Ciii), lowering confidence in the actual branching topology. Some differential clustering of *T. cacao* sequence can be observed among the 18 and 19 respective *Arabidopsis* gene family members of actin 12 and 58, but these data will require further evidence to be able to draw robust conclusions.

Cellulose synthase is observed to have 30 *Arabidopsis* gene family members (Appendix Ciii). Among these, the four *T. cacao* sequences are observed to cluster with three different *Arabidopsis* sequences. The first cluster contains two *T. cacao* sequences (TcaOMWAL1f, and OMWAL1a), the cotton sequence, the *Populus* sequence and the *Arabidopsis* sequence Atg02730. The two *T. cacao* sequences are possibly the same gene and the sequence difference could be due to PCR error, sequencing error, or that they are different alleles. The second group containing *T. cacao* sequence (TcaOMWAL1) contains three *Arabidopsis* sequences including that used to create the primer. The third group contains the *T. cacao* sequence TcaOMWAL1c and two other *Arabidopsis* sequences. These data indicate that there is the possibility of three

potential gene family members and that the common ancestor cellulose synthase gene may have been a single gene before the eurosids I, eurosids II speciation event.

3.4.2 Use of amplicons as DNA markers in *T. cacao* and in other species

This method highlighted another important benefit of creating degenerate primers. Through the sequencing of multiple clones, polymorphisms were identified and through phylogenetic analysis, it was seen that some of the sequences extracted could belong to multiple members of genes family, or alleles of the same gene. Furthermore, some of these differences fell over restriction sites which allow their potential use as PCR-RFLP DNA markers (Deng, 1988). The 17 potential SNP sites identified in this experiment require verification as to their use as DNA markers, however the potential is obvious. The advantage of using coding sequence rather than non-coding sequence (such as simple sequence repeat markers (SSRs)) is that the polymorphisms are located directly on the gene of interest, thus producing very reliable markers.

These results indicate that on average 75% of the primers created were able to amplify DNA from four diverse plant species. This was further corroborated by a dendrogram produced with this data that showed similar relationships to the currently accepted phylogenetic relationships (APGII, 2003). This preservation of phylogenetic relationships suggests that the PCR products amplified were due to the sequence similarity of the species and in turn reflected the evolutionary relationships among these species and their respective sequence divergence from *T. cacao*. By combining these data the following primers would be most applicable for further phylogenetic studies, as they simultaneously show orthology among *T. cacao*, cotton, *Arabidopsis*, and *Populus*, and are able to produce PCR product across a wide range of rosid species. These primers amplify cDNAs from the genes Glutamate Decarboxylase, Glutamate Dehydrogenase, Adenosine Kinase, Actin 27, Actin 58, (S)-2-Hydroxy-Acid Oxidase, Peroxisomal,

Isocitrate Dehydrogenase, Pyruvate Dehydrogenase E1 Component Alpha Subunit,
Cellulose Synthase Catalytic Subunit, Alcohol Dehydrogenase Class III, and β -tubulin.

3.5 References

- Antunez de Mayolo G** (2003) Genetic engineering of *Theobroma cacao* and molecular studies on cacao defense responses. The Pennsylvania State University, University Park (Ph.D. Thesis)
- APGII** (2003) An update of the angiosperm phylogeny group classification for orders and families of flowering plants: APG II. *Botanical Journal Of The Linnean Society* **141**: 399-436
- Bayer C, Fay MF, de Bruin AY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, Chase MW** (Unpublished) Support for an expanded family concept of Malvaceae with a recircumscribed order Malvales: a combined analysis of plastid atpB and rbcL DNA sequences. *In*,
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD** (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**: 3497-3500
- Deng GR** (1988) A sensitive non-radioactive PCR-RFLP analysis for detecting point mutations at 12th codon of oncogene c-Ha-ras in DNAs of gastric cancer. *Nucleic Acids Research* **16**: 6231
- Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797
- Kim HU, Cotter R, Johnson S, Senda M, Dodds P, Kulikauska R, Tang W, Ezcara I, Herzmark P, McCormick S** (2002) New pollen-specific receptor kinases identified in tomato, maize and *Arabidopsis*: the tomato kinases show overlapping but distinct localization patterns on pollen tubes. *Plant Mol Biol* **50**: 1-16
- Kumar M** (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111-120
- Kumar S, Tamura K, Jakobsen IB, Nei M** (2001) *MEGA2: Molecular Evolutionary Genetics Analysis* software. *Bioinformatics* **17**: 1244-1245
- Mayer K, Schuller C, Wambutt R, Murphy G, Volckaert G, Pohl T, Dusterhoft A, Stiekema W, Entian KD, Terry N, Harris B, et al.** (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**: 769-+
- Prosser I, Phillips AL, Gittings S, Lewis MJ, Hooper AM, Pickett JA, Beale MH** (2002) (+)-(10R)-Germacrene A synthase from goldenrod, *Solidago canadensis*; cDNA isolation, bacterial expression and functional analysis. *Phytochemistry* **60**: 691-702
- Salanoubat M, Lemcke K, Rieger M, Ansoerge W, Unseld M, Fartmann B, Valle G, Blocker H, Perez-Alonso M, Obermaier B, Delseny M, et al.** (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* **408**: 820-822

- Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, Miyajima N, Sasamoto S, Kimura T, Hosouchi T, Kawashima K, et al.** (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* **408**: 823-826
- Tatusov RL, Koonin EV, Lipman DJ** (1997) A genomic perspective on gene families. *Science* **278**: 631-637
- Whitlock BA** (2002) Phylogenetic relationships and floral evolution of Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) based on sequences of the chloroplast gene, *ndhF*. *Systematic Botany* **26**: 420-437
- Whitlock BA, Baum DA** (1999) Phylogenetic relationships of *Theobroma* and *Herrania* (Sterculiaceae) based on sequences of the nuclear gene *Vicilin*. *Systematic Botany* **24**: 128-138
- Whitlock BA, Bayer C, Baum DA** (2001) Phylogenetic relationships and floral evolution of the Byttnerioideae ("Sterculiaceae" or Malvaceae s.l.) based on sequences of the chloroplast gene, *ndhF*. *Systematic Botany* **26**: 420-437

Chapter 4

Spatial and Temporal Expression Patterns of Several Key Floral Regulatory Factors during Flower Development in *T. cacao*



Fire

4.1 Introduction

The comparative analysis presented in Chapter 2 allowed a conclusion to be made that there is a high degree of conservation between the developmental programs of *T. cacao* and *Arabidopsis*. However, it is unknown if the conservation of the floral developmental program in *T. cacao* extends to the level of gene expression. It has been shown in other species, (such as tomato, *N. tabacum*, and *Vitis vinifera*) that the respective orthologs of the *Arabidopsis* *LEAFY* (*LFY*) and ABC genes have similar functions (Jack, 2004). It can therefore be expected that this will most likely also be true for *T. cacao* and *Arabidopsis*. The goal of this chapter is therefore to test the hypothesis that selected genes involved in the developmental program of *T. cacao* are expressed in similar tissues and at similar times as their orthologous counterparts in *Arabidopsis*.

4.1.1 *LEAFY* and the ABC model

As stated in Chapter 1, there are two major unifying principles in flower development (Jack, 2004). The first is the ABC model. The ABC model was developed (through analysis and subsequent gene characterization of flowering mutants) to explain how floral organs are able to develop in the correct place, at the correct time, and was expanded through analysis and subsequent gene characterization of flowering mutants. It was subsequently shown to be applicable to a wide variety of both monocot and eudicot species (Bowman et al., 1991; Coen and Meyerowitz, 1991; Meyerowitz et al., 1998; Ambrose et al., 2000; Theissen et al., 2000; Fornara et al., 2003).

The ABC model refers to the idea that combinations of three classes of gene products interact to determine the identity of the cells forming the four whorls of the

flower. Simplified, the A class of genes are responsible for the formation of the sepals, the combination of the A and B genes are responsible for the identity of petals, the combination of B and C genes are responsible for stamen identity, and the C genes alone are responsible for the identity of carpels. Finally, A and C genes are mutually repressive. The ABC model has been extended to include D and E class genes (Angenent and Colombo, 1996; Theissen et al., 2000); however, for this discussion these additional classes are not relevant.

The primary A class genes that have been identified in *Arabidopsis* are *APETALA1* (*API*) and *APETALA2* (*AP2*); B class are *APETALA3* (*AP3*) and *PISTILLATA* (*PI*); C class are *AGAMOUS* (*AG*). All of these genes, with the exception of *AP2*, are termed MADS-box genes due to their characteristic structure that contains (from 5' to 3'): an N-terminus, the MADS DNA binding/nuclear localization/dimerization domain, an intervening region, a K domain for mediating protein-protein interactions and dimerization functions, and the C domain for transcriptional activation and the formation of higher order complexes (McGonigle et al., 1996; Fan et al., 1997; Riechmann and Meyerowitz, 1997; Egea-Cortines et al., 1999; Moon et al., 1999; Honma and Goto, 2001; Immink et al., 2002; Yang et al., 2003).

The expression of these genes has been examined in several species. In *Arabidopsis*, *API* is initially expressed throughout the floral meristematic tissues and is eventually reduced to just the outer two whorls as *AG* inhibits *API* expression (Mandel et al., 1992; Gustafson-Brown et al., 1994). *AP2* is found to be expressed throughout the meristematic tissues during floral organ identity (Jofuku et al., 1994). However, it is only active in whorls 1 and 2, due to inhibition via siRNA mechanisms in whorls 3 and 4 (Chen, 2004). *AP3* and *PI* genes are expressed exclusively in whorls 2 and 3 during organ identity while *AG* is expressed in the central two whorls of the developing floral meristem (Yanofsky et al., 1990; Jack et al., 1992; Goto and Meyerowitz, 1994). Of the genes discussed, homologs of *AG* seem to have some plasticity in their expression. For example in *Petunia spp.*, *AG* is expressed across the entire meristem during early floral

stages and its expression is reduced to the central two whorls later in organogenesis (Zahn pers. comm.).

The other unifying principal of flower development is the role of the *LFY* gene that acts as both an integrator of the floral pathways (Blazquez et al., 1997; Nilsson et al., 1998) and an activator of the floral organ identity genes (ABC genes) (Weigel and Meyerowitz, 1993; Parcy et al., 1998; Lenhard et al., 2001; Lohmann et al., 2001). Like the ABC genes, *LFY* orthologs have been identified in other species (Gocal et al., 2001) and are generally expressed across the entire meristem during the phase change from vegetative to inflorescence development and throughout the identity phase of organogenesis. *LFY* generally reduces in expression once the organs are specified (Weigel et al., 1992).

4.1.2 Methods of analyzing gene expression

There are many methods available to analyze gene expression. These among others, include northern blots, Real Time PCR, and microarray analysis. Northern blots require relatively large quantities of RNA (10-100 µg) and can be difficult when there only small amounts of tissue such as developing flower buds (Jackson, 1992). While Real Time-PCR requires less RNA (1-100 ng) than northern blots, a separate fluorescent primer is required for each gene studied (Freeman et al., 1999). Microarray techniques require large amounts of highly purified RNA which may be difficult to obtain if dealing with a specific part of the tissue (Jackson, 1992). In an effort to use small amounts of tissue for micorarray analysis, RNA amplification has been attempted; however, it is unknown if the amplification preserves the original proportions of respective genes in the sample (Iscove et al., 2002). Microarrays have been extensively used in model species such as *Arabidopsis* (Girke et al., 2000; Ramonell et al., 2002), and have led to large collections of data at web-based sites (Zimmermann et al., 2004). These community resource websites allow global searches of genes of interest so that

researchers who do not have the ability to run their own “wet” laboratory microarray experiments can see what others have done pertaining to their respective research.

However, all of the methods described above share the limitation of not being able to provide precise spatial information regarding the expression of the gene of interest. *In situ* hybridization allows the visualization of gene expression using high magnification microscopic resolution. *In situ* hybridization is a technique where tissue is fixed, embedded in paraffin, sectioned, and adhered to a glass slide. Labeled DNA or RNA fragments of specific genes of interest are used as probes and are hybridized to the tissue sections. Hybridization of the probe reveals the localization of RNA in the tissue section. The spatial resolution of this method far exceeds that of northern blotting or RT-PCR methods by allowing identification of expression down to a single cell layer, and in some cases has been used to localize mRNA at a sub-cellular level (Jackson, 1992; Scutt, 2001). This allows precise determination of tissue and time specific expression patterns common in regulatory genes.

4.1.3 *In situ* hybridization and the ABCs of *T. cacao*

In order to assess if the conservation of floral developmental sequences identified between *Arabidopsis* and *T. cacao* in Chapter 2 are extended to the expression level we have used *in situ* hybridization to investigate expression of selected ABC genes (*AP1*, *AP3*, *PI* and *AG*) and *LFY* in *T. cacao* and to compare that with expression of these genes in *Arabidopsis*. These results were compared to *Arabidopsis* data that are available in the literature and on the world-wide-web. Some comparisons were also made to other species.

4.2 Methods

In situ analysis was carried out by combining the radioactive protocol currently in use in the Ma lab (Zahn pers. comm.) with the protocol published by the Barton lab (Lynn et al., 1999). The Barton lab protocols are modifications of the protocols developed in the labs of Hake, Irish and Meyerowitz. Some references for these protocols include Meyerowitz, (1987) and Jackson, (1992).

4.2.1 Gene selection and isolation

In total, six cDNAs were analyzed. Fragments of DNA encoding portions of the *LFY* and *CELLULOSE SYNTHASE* were isolated from *T. cacao* using the previously described degenerate primer method (see Chapter 3.2). *Arabidopsis* cDNAs encoding *APETALLA1* (*API*), *APETALLA3* (*AP3*), *PISTALATA* (*PI*), and *AGAMOUS* (*AG*) were obtained in the following plasmids (pAM128, pD793, pcPINX, and pCIT565, respectively) from the Ma laboratory (Dept. of Biology, The Pennsylvania State University) and had been used previously for *in situ* hybridization (Yanofsky et al., 1990; Jack et al., 1992; Mandel et al., 1992; Goto and Meyerowitz, 1994; Gustafson-Brown et al., 1994). All inserts were sequenced and compared to the NCBI database to confirm their identity prior to use. The orientations of each probe were also elucidated, and the restriction sites surrounding the inserts were verified.

4.2.2 Southern analysis of the *T. cacao* *LEAFY* gene

Southern analysis of the cacao *LFY* gene was carried out by digesting three samples of CsCl purified genomic DNA from the *T. cacao* genotype PSU-SCA6 with EcoRI, HindIII, and BamHI (Promega, Madison, WI). These were digested in 5 μ L of 5x Buffer, 1 mg of genomic DNA, 1 μ L of enzyme (5 u/ μ L), made up to 50 μ L with

water and incubated overnight at 37 °C. DNA was separated on a 0.8% agarose gel, denatured for 15 min in 0.5 M NaOH, 1.5 M NaCl, neutralized for 15 min in 0.5 M Tris-HCl (pH 7.4), 1.5 M NaCl. The gel was blotted overnight to Hybond™ N+ positively charged membrane (Amersham, Upsala, Sweden) using 10X SSC. The DNA was UV cross-linked to the membrane using a UV Stratalinker 1800 (Stratagene) at 1200 μJoules X100 to bind the DNA to the membrane.

Probes were PCR amplified from the pCR-Topo® cloning vector (Invitrogen, Carlsbad, CA) containing the *LFY* gene. Amplification was carried out in a reaction mix containing 1 μL of plasmid DNA template (10 ng/μL), 2.5 μL of 10x Buffer (PGC Scientifics, Frederick, MD), 2.5 μL of 2 mM dNTPs, 1 μL each of forward and reverse 10 mM primer, 1 μL of Taq DNA Polymerase (5u/μL) (PGC Scientific, Frederick, MD). Deionized water was used to bring the reaction mix to a total volume of 25 μL. The PCR mix was cycled in a Perkin-Elmer 9600 thermocycler (Boston, MA) for 4 min at 95 °C, then thirty cycles of 94 °C for 30 sec, 50 °C for 30 sec, and 72 °C for 40 sec. Finally, the reaction was held at 72 °C for 7 min and then held at 4 °C.

Labeling was carried out using the Megaprime DNA labeling kit (Amersham Biosciences, Buckinghamshire, UK) using the manufacturer's instructions. The entire PCR product (and 0.5 ng of 1 kb ladder (PGC Scientifics, Frederick, MD)) was mixed with 5 μL of the included primer mix and boiled for 5 min then quenched on ice. To this mix was added: 5 μL of reaction buffer, 12 μL of the included dATP, dGTP, and dTTP mix, 5 μL of dCTP-P³², and 2 μL of Klenow enzyme (provided in the Megaprime kit). This was incubated at 37 °C for one hour. The probe was cleaned using a DTR Gelfilter cartridge (Edge Biosystems, Gaithersburg, MD). Then boiled for 5 min and quenched on ice just before use.

The membrane was pre-hybridized for 2 hours with the ExpressHyb Hybridization solution (BD Biosciences, Palo Alto, CA) at 65 °C. The pre-hybridization solution was replaced with hybridization solution and the probe was added.

Hybridization took place in a rotating incubator overnight at 65 °C. The following morning the membrane was washed three times for 30 min at 65 °C in 1 x SSC, 0.1% SDS. After washing, the membrane was visualized using storage phosphor imaging (Molecular Dynamics, www.mymdn.com).

4.2.3 Dot blot hybridization of heterologous probes

Dot blot hybridization of the genes used for *in situ* hybridization was carried out by spotting 1 µL of colonies grown overnight onto Hybond™ N+ positively charged nylon membranes (Amersham, Uppsala, Sweden) in duplicate. The following colonies containing plasmids with each of the respective gene sequences were as follows: *CELLULOSE SYNTHASE* (from *T. cacao*), *LFY* (from *T. cacao*), *AP1*, *AP3*, *PI*, and *AG* (from *Arabidopsis*) (see above for plasmid names). The membrane was placed on media containing LB + ampicilin and grown overnight at 37 °C. The following morning two pieces of Whatman® filter paper were cut and one was saturated in denaturing solution (0.5 M NaOH, 1.5 M NaCl made fresh), while the other was saturated in neutralization solution (0.5 M Tris-HCl (pH 7.4), 1.5 M NaCl). The membrane was placed colony side up on the denaturing filter paper for four minutes, and then transferred to the neutralizing membrane for four minutes, after which the membrane was dried for 30 min. The DNA was UV cross-linked to the membrane using a UV Stratalinker 1800 (Stratagene) at 1200 µJoules X100 to bind the DNA to the membrane.

Labeling was carried out using the Megaprime DNA labeling kit (Amersham Biosciences, Buckinghamshire, England) using the manufacturer's instructions. The hybridization was carried out as described above in section 4.2.2.

4.2.4 *In situ* hybridization probe preparation

After assessment of the orientation of the insert in the vector, two probes for each cDNA fragment were produced: Firstly a negative control that is orientated in the sense direction and would be expected not to hybridize to the mRNA transcript of interest, and secondly, a probe that would be orientated in the anti-sense direction and will hybridize to the mRNA of interest.

These plasmid templates were linearized by digesting 5 µg of each plasmid with the appropriate restriction enzymes (Table 4-1). These reactions were phenol/chloroform extracted once and the supernatant transferred to RNase-free eppendorf tubes. These were precipitated with 1/10 volumes of RNase-free 3 M NaOAc and 2.5 volumes of ethanol, and washed in RNase-free 70% Ethanol. The linearized plasmids were re-suspended in RNase-free TE buffer at 1 mg/ml.

Table 4-1: List of restriction enzymes and RNA Polymerases used for *in situ* probe preparation

Gene	Plasmid name	Insert size (bp)	Restriction Enzyme	RNA Pol. used	Probe +ve/-ve control
<i>CELLULOSE SYNTHASE</i>	TcCsTOPO	439	NotI	T7	+ve
			NcoI	SP6	-ve
<i>LEAFY</i>	TcLFYTOPO	403	NcoI	SP6	+ve
			NotI	T7	-ve
<i>API</i>	pAM128	550	NcoI	SP6	-ve
			NotI	T7	+ve
<i>AP3</i>	pD793	536	PstI	T7	+ve
			XbaI	SP6	-ve
<i>PI</i>	pcPINX	507	XbaI	SP6	-ve
			MluI	T7	+ve
<i>AG</i>	pCIT565	541	BamHI	T7	+ve
			XhoI	SP6	-ve

The probes were labeled with digoxigenin using transcription reactions. They were set up using the Promega Transcription Kit (with SP6 and T7 RNA Polymerase) (Promega, Madison, WI) as follows: 1.5 μ L of 2.5 mM UTP, 1.5 μ L of 2.5 mM digoxigenin-UTP (Roche, Indianapolis, IN), 10 μ L of 5x Transcription buffer, 5 μ L of 100 mM DTT, 2 μ L of RNasin, 3 μ L each of 2.5 mM ATP, GTP, and CTP nucleotide, 1 μ L of linearized plasmid, 19 μ L of RNase free water, and 1 μ L of either SP6 or T7 RNA Polymerase. The transcription reaction components were mixed then incubated at 40 °C for 1.5 hours.

After incubation, 1 μ L of DNase was added and the mixtures were incubated at 37 °C for 15 min. Once complete, the mixtures were made up to total volumes of 200 μ L with TE + 10 mM DTT. 2 μ L of 10 mg/mL tRNA (Invitrogen, Carlsbad, CA) was added to each tube and these mixtures were extracted with RNase free phenol, then RNase free chloroform/isoamyl alcohol. The resulting supernatants were precipitated with 1/10 vol NaOAc and 2.5 vol of 100% ethanol. The precipitates were washed with 70% ethanol, dried, and re-suspended in 50 μ L of RNase free TE buffer.

The labeled probes in 50 μ L TE were hydrolyzed with 30 μ L and 20 μ L of freshly made 200 mM Na₂CO₃, and 200 mM NaHCO₃, respectively. This mixture was incubated at 60 °C for a period of time calculated by the following formula: $t=(L_0-L_f)/(K)(L_0)(L_f)$, where L_0 =starting length (in kb), L_f =final length (kb) = 0.1-0.075 kb, $K=0.11$, t = incubation time in min. These times were ($L_f=0.1$ kb): *CELLULOSE SYNTHASE*, 70 min; *LFY*, 68 min; *API*, 74 min; *AP3*, 73 min; *PI*, 72 min; *AG*, 74 min.

The reactions were stopped by incubation on ice and adding 3 μ L of 3 M sodium acetate (pH 6.0) and 5 μ L of 10% glacial acetic acid. The probes were then precipitated by adding 1 μ L 1 M DTT, 1 μ L 10 mg/mL tRNA carrier, 8 μ L 3 M sodium

acetate, and 250 μ L ethanol. These were stored in the -20°C freezer until the probes were prepared for hybridization.

4.2.5 Tissue selection and fixation

Floral meristems were selected from cushions that were 14 days or younger. In many cases, a more mature flower was also present on the pedicel that contained the young bud. In these cases, both flower buds were collected. Collection was done by cutting the pedicel as close to the flower cushion as possible with a sterile scalpel and dropping the tissue into vials containing FAA (50% ethanol, 5% acetic acid, 3.7% formaldehyde, and 41.3% water). This mixture was placed under vacuum overnight.

The following morning the fixative was replaced by the following ethanol series 50%, 50%, 60%, 70%, 85%, 95% for 30 min each step. The mixture was left overnight in 95% ethanol. The 95% ethanol was replaced with 100% ethanol and the mixture was incubated for one hour. This step was repeated.

The tissue was cleared in the following solutions made from fresh ethanol and histosolve[®] (Thermo, Pittsburgh, PA): 25% histosolve[®]:75% ethanol, 50% histosolve:50% ethanol, 75% histosolve[®]: 25% ethanol. Each of the incubations was performed at room temperature for 30 min. The final solution was replaced with 100% histosolve[®] for one hour, and this wash was repeated two times.

The tissue was infiltrated by adding a layer of paraplast[®] chips (Structure Probe Inc. West Chester, PA) (about 20) to the 100% histosolve[®] solution, and incubating overnight at room temperature. The following morning the solution was placed at 42°C until the paraplast[®] was in solution. Once in solution, another layer of paraplast[®] chips were added. This procedure was repeated five times after which the

molten paraplast[®]/histsolve[®] solution was poured off and replaced with molten paraplast[®] and incubated at 60 °C. The paraplast[®] was replaced each night and morning for a total of five days. The tissues were embedded in paraplast[®] blocks, hardened for 24 hours at room temperature and sectioned to 8 µm using a Shandon Finesse Microtome (Thermo Electron Corp., Pittsburgh, PA). The sections were placed on ProbeOn[®] Plus slides (Fisher Scientific, Pittsburgh, PA) and incubated for one hour on a slide warmer at 42°C and then incubated in an oven at 47°C for twelve hours. Then the slides were able to be stored at room temperature until used (for up to three months).

4.2.6 Slide preparation

All solutions and glassware were RNase-free by DEPC treatment and autoclaving, or soaking in 0.2 M NaOH. Paraplast[®] was removed from the slides by two washes of 10 minutes each in histsolve[®]. The slides were washed briefly (10 sec/wash) in the following ethanol series: 100%, 100%, 95%, 85%, 70%, 50%, 30%, 15%, water, water.

The slides were placed in a proteinase K solution (1 mg/ml) for 30 min at 37°C. After incubation the slides were rinsed in water twice and once in 1x PBS (7 mM Na₂HPO₄, x130 mM NaCl, 3 mM NaH₂PO₄), then incubated at room temperature in 2 mg/mL glycine in 1x PBS to block the protease. The slides were washed twice in 1x PBS for 2 min each.

Any remaining positive charges were removed from the slides to reduce background by washing them for 5 min in a solution of 100 mM triethanolamine (pH ~8.0) containing 1.25 mL acetic anhydride/500mL of triethanolamine solution. The slides were then washed twice in 1x PBS for 5 min. Finally, the slides were dehydrated through the following ethanol series: 30%, 50%, 70%, 85%, 95%, 100%, and dried under vacuum for 3-4 hours until hybridization.

4.2.7 Hybridization and signal development

The 2 µg RNA probes that were generated as described in section 4.2.5 were prepared for hybridization by precipitating them using 1/10 vol NaOAc and 2.5 vol of 100% ethanol. The precipitate was washed with 70% ethanol, dried, and re-suspended in 50 µL of RNase-free TE buffer. The probe was separated into two separate 2.5 µg aliquots, one was stored at -20°C for later use and the other was boiled for 3 min, and then quenched on ice.

One µL of probe was spotted to nylon membranes (Amersham Hybond™ N+) (Amersham, Uppsala, Sweden) to check labeling of the probe (see below), and 24 µL was added to 180 µL of OmniPur Hybridization solution (EM Science, Houston, TX), 50% formamide (prewarmed to 50 °C). Once the probe and hybridization solution were mixed, slides from section 4.2.6 were coupled into pairs of common probes and 200 µL of hybridization solution containing 50 ng of probe was pooled on one slide; the other was sandwiched face down. The slides were stored overnight (<16 hours) in an airtight box layered with wet paper towels (humidity chamber) at 46 °C.

The following wash steps were completed using autoclaved but not RNase free glassware and solutions. The following morning, slides were separated in 0.2 x SSC (0.5 x – 1 x SSC for heterologous probes) and washed in 0.2 x SSC (or 0.5 x – 1 x SSC) for three washes of one hour each at 45°C with gentle agitation. The slides were washed one more time in 1 x PBS for five minutes at room temperature before they were separated and placed face up in a tray containing 1% Boehringer block solution (Roche, Indianapolis, IN) in 100 mM Tris (pH 7.5), 150 mM NaCl for 45 min on a rocking platform at room temperature. The blocking solution was replaced by 1% bovine serum albumin in 100 mM Tris (pH 7.5), 150 mM NaCl, 0.3% triton X-100 for 45 min at room temperature on a rocking platform.

Anti-DIG antibody (Roche, Indianapolis, IN) was diluted 1:1250 in the above defined BSA:Tris:NaCl:Triton solution and slides were sandwiched and dipped in the pool of antibody using capillary action to draw the solution up the slides. The solution was drained on paper towels, re-drawn, and stored for two hours at room temperature in humidity chambers.

The slides were separated, arranged face up, and washed four times for 15 min each time in BSA:Tris:NaCl:Triton solution on a rocking platform. The slides were further washed twice for 10 min in 100 mM Tris (pH 9.5), 100 mM NaCl, 50 mM MgCl₂, and then sandwiched together. The slide pairs were dipped in NBT/NCIP substrate solution (Promega, Madison, WI), drained and re-dipped. The slides were stored in the above described humidity chambers in complete darkness for two days to allow color development. Membranes underwent identical washes and color development as described above for the slides, beginning with the first 45 min BSA:Tris:NaCl:Triton wash.

Once developed, the slides were separated in TE buffer (pH 7.5) then dipped quickly through the following ethanol and histosolve[®] series: 30%, 50%, 70%, 85%, 95%, 100%, 100%, histosolve[®], histosolve[®]. The slides were drained and mounted with cyto seal 60 (VWR, West Chester, PA) and covered with a coverslip.

4.2.8 Slide analysis

Slides were viewed on a Nikon Optisport Microscope (Nikon, Garden City, NY) and sections showing signal were photographed using a Spot RT Slider digital camera (Diagnostic Instruments, Sterling Heights, MI). The images were subjected to further analysis using Photoshop[®] CS (Adobe[®], San Jose, CA) where blue hues associated with the NBT/BCIP precipitate were selected (using the select > color range tool) and replaced (using image > adjustments > replace color tool) with colors that

allowed the probe to be more easily recognized. Since the blue hues could vary from experiment to experiment (especially between homologous and heterologous probes), selection of RGB values was done once per *in situ* hybridization experiment and the resulting Photoshop® color range file was used to apply the filter identically on all images of that experiment.

4.3 Results

4.3.1 Gene selection and analysis of *in situ* hybridization probes

4.3.1.1 *CELLULOSE SYNTHASE* and *LEAFY* gene

A 439 bp cDNA fragment was obtained using the degenerate primer method described in Chapter 3. This sequence was shown to have high levels of sequence identity ($>e^{-20}$) to other *CELLULOSE SYNTHASE* genes in other species. We therefore postulated that this fragment was a putative *CELLULOSE SYNTHASE* for *T. cacao*. *CELLULOSE SYNTHASE* is a membrane bound protein that contributes to synthesis of the cellulose component of the plant cell wall. It is expected to be expressed in growing plant tissues where cells are dividing and expanding as the plant cell expands the cell wall (Amor et al., 1995). Due to the fact that flower meristems are undergoing rapid growth and development *CELLULOSE SYNTHASE* was selected as a potential positive control for *in situ* hybridization.

A putative 402 bp *LFY* cDNA fragment was obtained using the degenerate primer method described in Chapter 3.2. This sequence was shown to have high homology ($>e^{-63}$) to the *LFY* sequences in *Arabidopsis*. Furthermore, this sequence was directly compared to the full-length genomic *LFY Arabidopsis* sequence (Accn. number: 30698605:24860061-24864200) and it was found that the *T. cacao* fragment isolated was

similar to the third exon of the *Arabidopsis* gene (e-value = $9e-63$). It was therefore postulated to be a fragment of the *LFY* ortholog for *T. cacao*. The *LFY* gene is regarded as an integrator and initiator of floral meristem development and its function is described in Chapter 1.2.2 (Jack, 2004).

LFY is also reported to be a single gene in *Arabidopsis*. To test copy number in *T. cacao*, a genomic Southern analysis was performed using the *T. cacao LFY* fragment as a probe. This analysis revealed a single band of strong hybridization with genomic DNA digested using three different restriction enzymes, supporting the hypothesis that the postulated *T. cacao LFY* is also a single copy gene (Figure 4-1). It is important to note, however, that the Southern blot was somewhat under exposed so the possibility of the *LFY* probe hybridizing to other corresponding DNA sequences at reduced intensities cannot be ruled out.

4.3.1.2 *Arabidopsis* genes *APETALA1*, *APETALA3*, *PISTILLATA*, and *AGAMOUS*

Plasmids containing cDNAs from the *Arabidopsis* genes *API*, *AP3*, *PI*, and *AG* were obtained from the Ma laboratory. The inserts were sequenced and were shown to be identical to the expected *Arabidopsis* genes (e-value = 0) (Table 4-1). These genes are involved in floral organ identity and are reviewed in Chapter 1.2.3.

Dot blot hybridization was carried out to determine if the *Arabidopsis* genes would cross hybridize to *T. cacao* orthologs if they exist. Two dot blots were created by spotting the *T. cacao* cDNAs *CELLULOSE SYNTHASE* and *LFY*, and the *Arabidopsis* cDNAs *API*, *AP3*, *PI*, and *AG* onto nylon membranes. The two membranes were hybridized separately with digested, radioactively labeled genomic DNA from *Arabidopsis* (Figure 4-2, panel A) and *T. cacao* (Figure 4-2, panel B). All spots were positive irrespective of DNA origin. It was noticed that *T. cacao* DNA did not hybridize as strongly as the *Arabidopsis* DNA, this phenomenon has also been noticed when attempting to hybridize *T. cacao* DNA to *Arabidopsis* microarrays (data not published).

In addition, the signal was weak when using *T. cacao* DNA and there was no negative control spot used in the experiment. However, these results provide some evidence (albeit not well supported) that the *Arabidopsis* genes will cross-hybridize to *T. cacao*, and likewise the *T. cacao* genes *LFY* and *CELLULOSE SYNTHASE* will hybridize to *Arabidopsis* DNA.

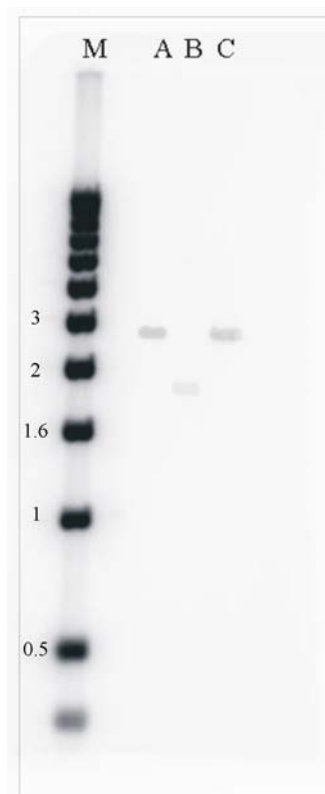


Figure 4-1: Genomic Southern blot of putative *T. cacao* *LEAFY* gene. “M” lane represents a 1kb ladder (PGC Scientific, Frederick, MD), sizes in kb are indicated next to the marker. “A”, “B” and “C” lanes represent genomic *T. cacao* cut with EcoRI, HindIII, and BamHI respectively. Putative *T. cacao* *LFY* was used as a probe.

4.3.2 *In situ* analysis

In situ hybridization was performed for the six genes *CELLULOSE SYNTHASE*, *LFY*, *AP1*, *AP3*, *PI*, and *AG* on tissue sections of varying stages of flower development. The *CELLULOSE SYNTHASE* and *LFY* probes were derived from *T. cacao*, whereas the others were derived from *Arabidopsis*. With the heterologous probes, we observed that the signal was weaker and the slides could not be washed as stringently without removing the probe, resulting in a lower signal to noise ratio for *AP1*, *AP3*, *PI*, and *AG*.

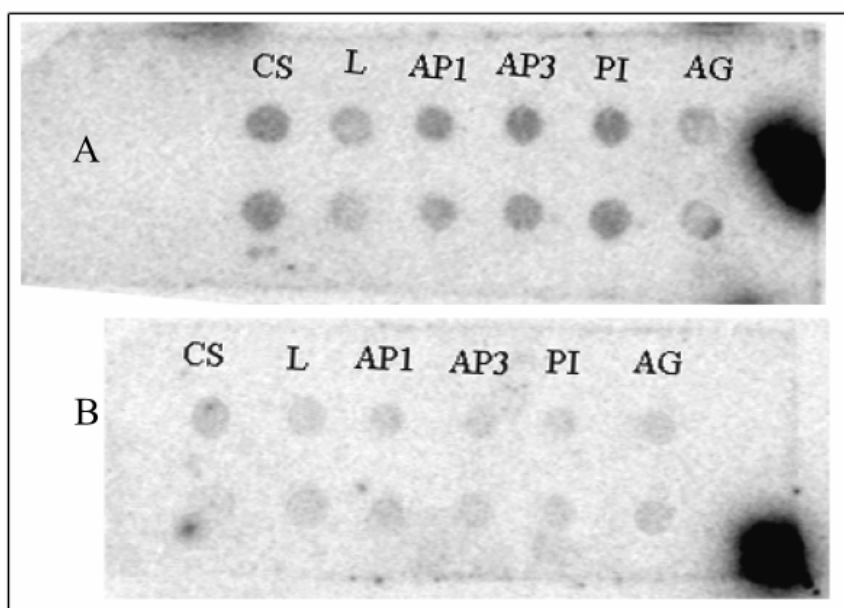


Figure 4-2: Dot Blot showing cross-hybridization of *Arabidopsis* and *T. cacao* DNA to *Arabidopsis* and *T. cacao* cDNA. Hybridization of labeled DNA from **A**, *Arabidopsis*, **B**, *T. cacao* to the following genes: *CELLULOSE SYNTHASE* (CS), *LEAFY* (L), *AP1*, *AP3*, *PI*, and *AG*. CS and L originated from *T. cacao*, all other genes originated from *Arabidopsis*.

4.3.2.1 *CELLULOSE SYNTHASE (CS)*

To evaluate the effectiveness of our *in situ* hybridization protocol, we chose to first hybridize a putative constitutively expressed gene. Since all growing plant cells need to produce new cell walls and/or thicken existing walls, the *CELLULOSE SYNTHASE* gene was expected to be expressed in all but terminally differentiated tissues and all cell types. The localization of the *T. cacao* gene *CELLULOSE SYNTHASE* was observed in stage 3 flowers and late stage 9 flowers. It was seen that the *CELLULOSE SYNTHASE* probe localized throughout all growing tissues suggesting that *CS* was expressed throughout the entire growing flower during the periods from days 3 – 20. A stronger signal was seen within areas undergoing rapid growth, such as the meristem (Figure 4-3, c and d) and the area where the taenial-segment begins to form (Figure 4-3, g and h). By comparison, the negative control (sense probe) did not show any specific strong hybridization

4.3.2.2 *LEAFY (LFY)*

To assess *LFY* expression in the early stages of floral development, both sense (control) and anti-sense *T. cacao* probes were hybridized to sections of stage 3 (data not shown), 5, and 9 (data not shown) buds. In contrast to the sense mRNA control probe, which showed no specific hybridization to the sections (Figure 4-4, a and b), the anti-sense probe hybridized in a clearly defined spatial pattern from the tip of the apical meristem to ~50 μm basal of the tip in stage 5 flowers. Expression of *LFY* in *T. cacao* can be summarized as occurring only in the meristematic cells that are undergoing organogenesis (Figure 4-4, c and d) and was observed in stages 3 – 5 (between days 3 – 7). Once organs have been specified and begin to form, *LFY* expression dissipates from the region where the respective organ has formed. Evidence for this includes no *LFY* expression in the sepal region of stage 5 flower buds (Figure 4-4, c and d), and no strong localized *LFY* expression in a fully identified stage 9 flower (data not shown).

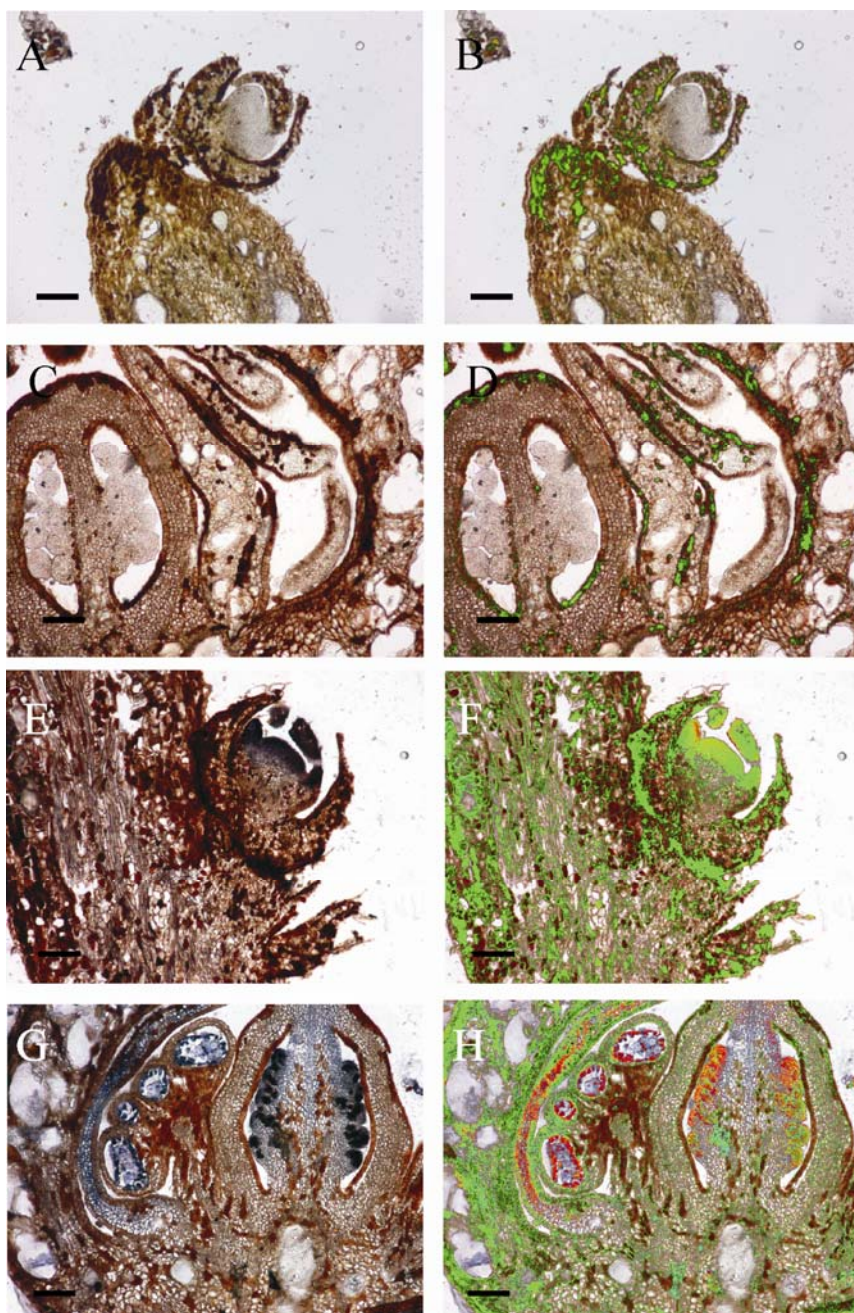


Figure 4-3: *In situ* hybridization of *CELLULOSE SYNTHASE* in PSU-SCA6 plants. **A-D**, Negative control sense *CS* probe hybridized to a stage 3 and late stage 9 flower bud. **E-H**, Hybridization of anti-sense probe to a stage 3 (**E-F**) and a late stage 9 (**G-H**) flower bud. Left hand panels are raw images, right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

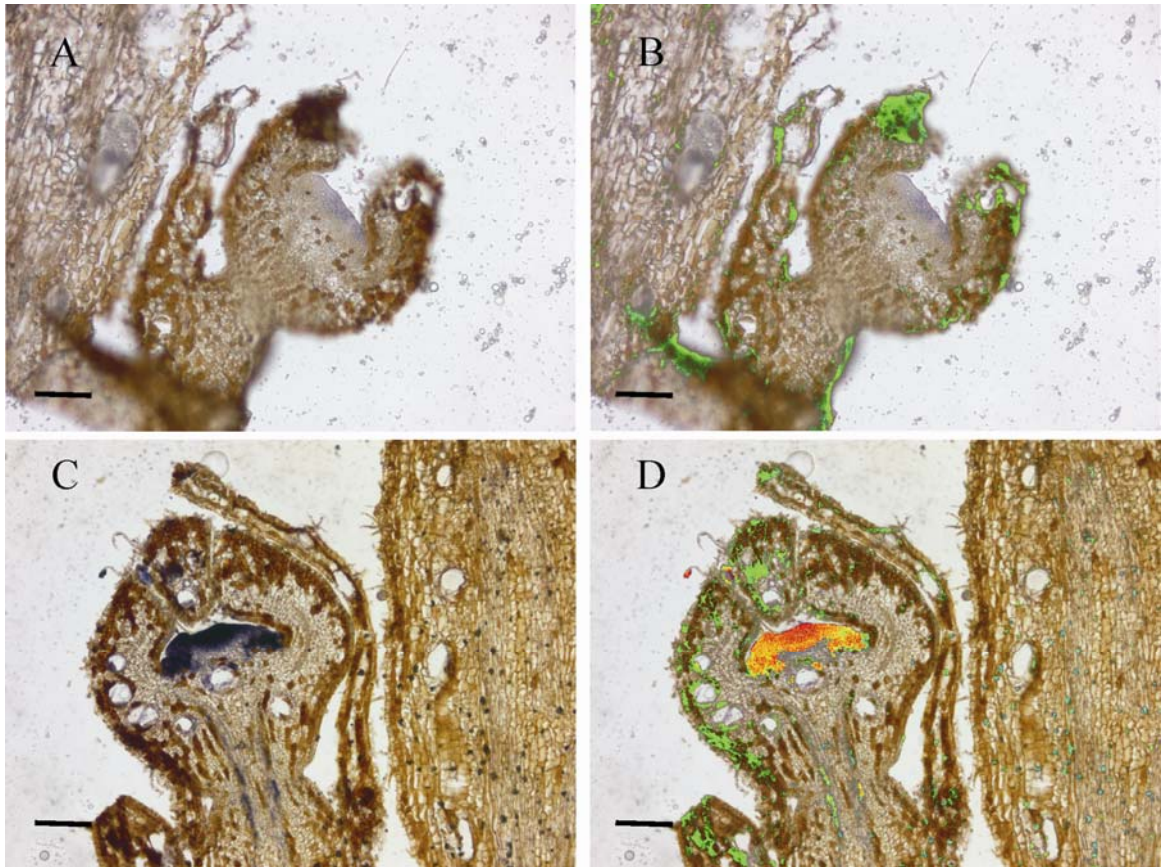


Figure 4-4: *In situ* hybridization of *LEAFY* probe in PSU-SCA6 plants. **A-B**, Negative control sense *LFY* probe hybridized to a stage 4 flower bud. **C-D**, Hybridization of anti-sense probe to a stage 5 flower bud. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

4.3.2.3 *APETALA1* (*API*)

In situ hybridization of the *Arabidopsis*-derived *API* was observed on sections of stages 4, 5 and 6 of *T. cacao* flower buds. Localization was observed across the developing meristematic tissues in stages 4 and 5 (days 4 – 7) (Figure 4-5, c-f). *API* expression extended across all regions of the developing floral meristem to a depth of

~50 μM . By late stage 6 (day 8), localization had diminished to just the petal primordia of the flower (whorl 2). As with *LFY*, *API* gene expression was diminished once the organs were visible (Figure 4-5, e – j). Since we did not observe sections beyond stage 6, it is unknown when *API* expression completely terminates in *T. cacao*. Sense (control) probes on early stage 6 and 7 flowers revealed no specific localized hybridization (Figure 4-5, a and b).

4.3.2.4 *APETALA3* (*AP3*)

Localization of the *Arabidopsis*-derived *AP3* signal was observed in sections of stage 4, stage 6 and stage 7 flower buds. In stage 5 buds, expression was observed exclusively in whorls 2 and 3 (to a ~50 μM depth in stage 4) where the petal and anther primordia are expected to form (Figure 4-6, c and d). This continues through stages 6 and 7 (Figure 4-6, e-h). From this data it is not possible to precisely determine when *AP3* expression is activated or turned off in *T. cacao*; however, we can conclude that it is during stages 4 – 7 (days 6 – 11). An important point is that, in *T. cacao*, *AP3* is not only expressed in the anthers and petals but also in the infertile staminode primordia (Figure 4-6, g and h). This indicates that staminodes, although very different in structure from stamens, also appear to express B class genes, providing some evidence that staminodes may also be subject to the ABC model of flower development. No specific localized expression was observed in the sense control probe (Figure 4-6, a and b).

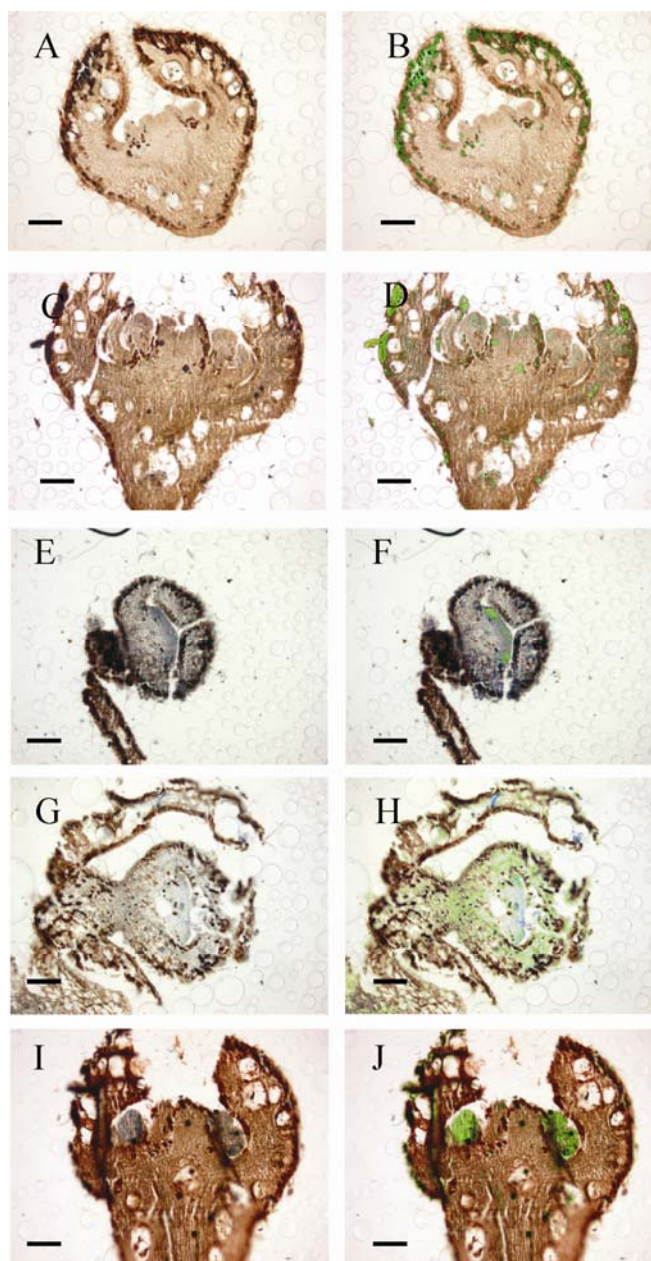


Figure 4-5: *In situ* hybridization of *Arabidopsis API* in PSU-SCA6 plants. **A-D**, Negative control sense *API* probe hybridized to an early stage 6 and stage 7 flower buds. **E-F**, Hybridization of anti-sense probe to a stage 4 flower bud. **G-H**, Hybridization of anti-sense probe to a stage 5 flower bud. **I-J**, Hybridization of anti-sense probe to a late stage 6 flower bud. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

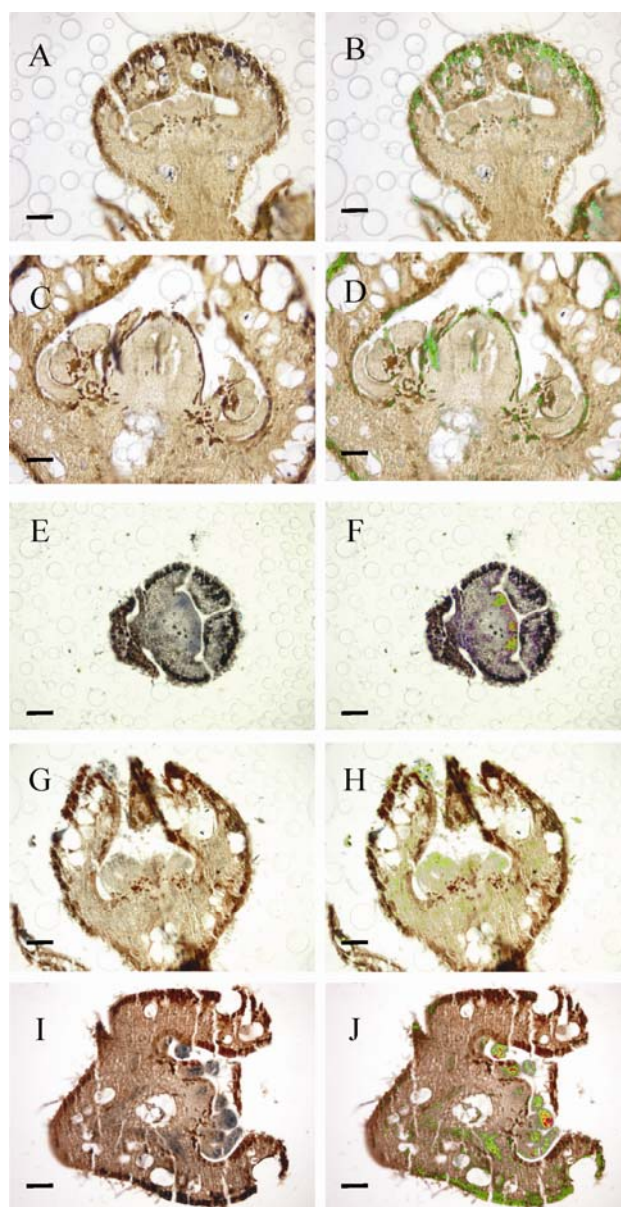


Figure 4-6: *In situ* hybridization of *Arabidopsis AP3* in PSU-SCA6 plants. **A-D**, Negative control sense *AP3* probe hybridized to a late stage 6 and stage 8 flower buds. **E-F**, Hybridization of anti-sense probe to a stage 4 flower bud. **G-H**, Hybridization of anti-sense probe to a stage 6 flower bud. **I-J**, Hybridization of anti-sense probe to a stage 7 flower bud. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

4.3.2.5 *PISTILLATA (PI)*

Localization of the *Arabidopsis*-derived *PI* signal was observed in tissue sections of stages 4, 5, and stage 7 in *T. cacao*. It was observed that localization of the probe in stage 4 (day 5) was exclusively in whorls 2 and 3, regions of the apical meristem where the petal and anther primordia were expected to form (Figure 4-7, c and d). Expression continued in the petal and anther regions throughout stages 5 and 7 (days 6 – 11), however, expression seemed to be reducing in petals by stage 7 (day 11) (Figure 4-7, e-g). Expression was also observed in staminodes in the above stages providing further evidence that staminode identity is provided via the ABC model. No specific localized expression was observed in the sense probe controls.

One other interesting observation made with *PI* was that localization was observed in the vascular bundles of the flower in mid stage six (day 10, the same time that the pedicel begins to form) (Figure 4-8, a-d). Skipper (2002), who suggested that B class genes have a possible role in the development of vascular tissue, has observed this in the B class genes *AP3* and *PI* in *Eranthis hyemalis*.

4.3.2.6 *AGAMOUS (AG)*

The *Arabidopsis* *AG* probe was hybridized to sections of late stage 4, late stage 5, and early stage 6 *T. cacao* flower buds. It was observed that the localization in stages 4 - 6 occurred over the central meristem specifically in the central whorls 3 and 4, just the areas where the anther and carpel were expected to develop (Figure 4-9, b-e). It is unknown how *AG* expression develops beyond stage 6 in *T. cacao*. No localized specific expression was observed in the negative sense controls (Figure 4-9, a).

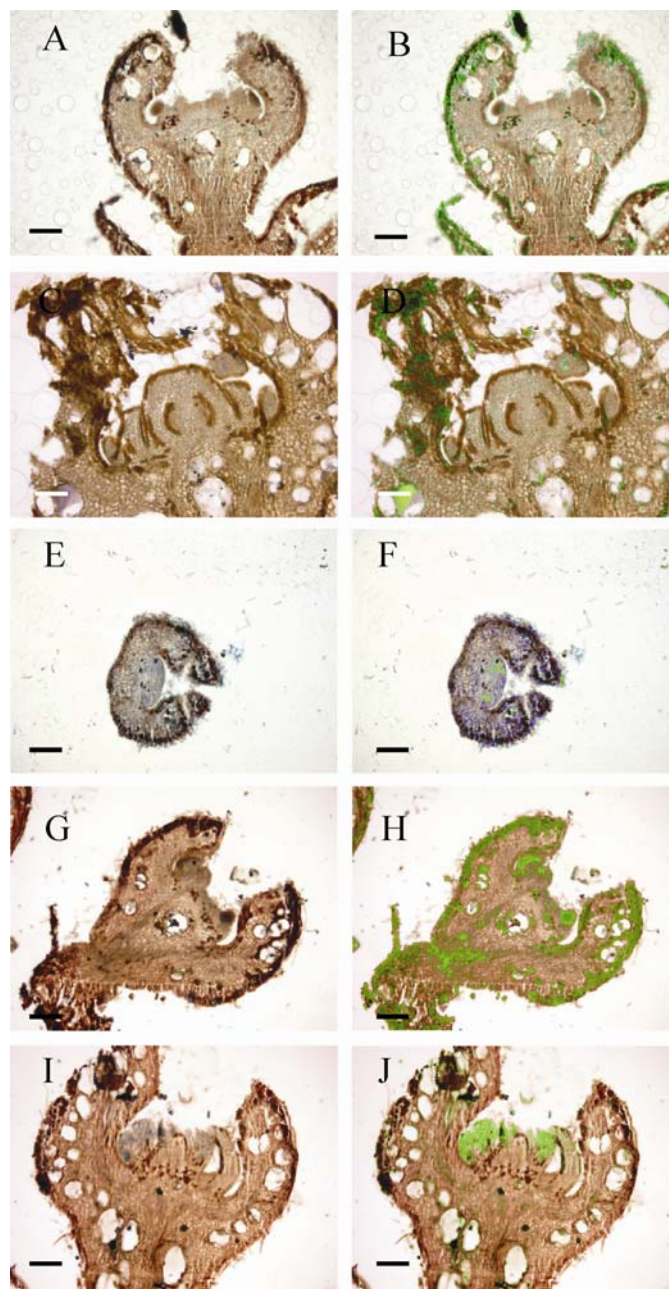


Figure 4-7: *In situ* hybridization of *Arabidopsis PI* in PSU-SCA6 plants. **A-D**, Negative control sense *PI* probe hybridized to late stage 6 and late stage 8 flower buds. **E-F**, Hybridization of anti-sense probe to a stage 4 flower bud. **G-H**, Hybridization of anti-sense probe to a stage 5 flower bud. **I-J**, Hybridization of anti-sense probe to a stage 7 flower bud. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

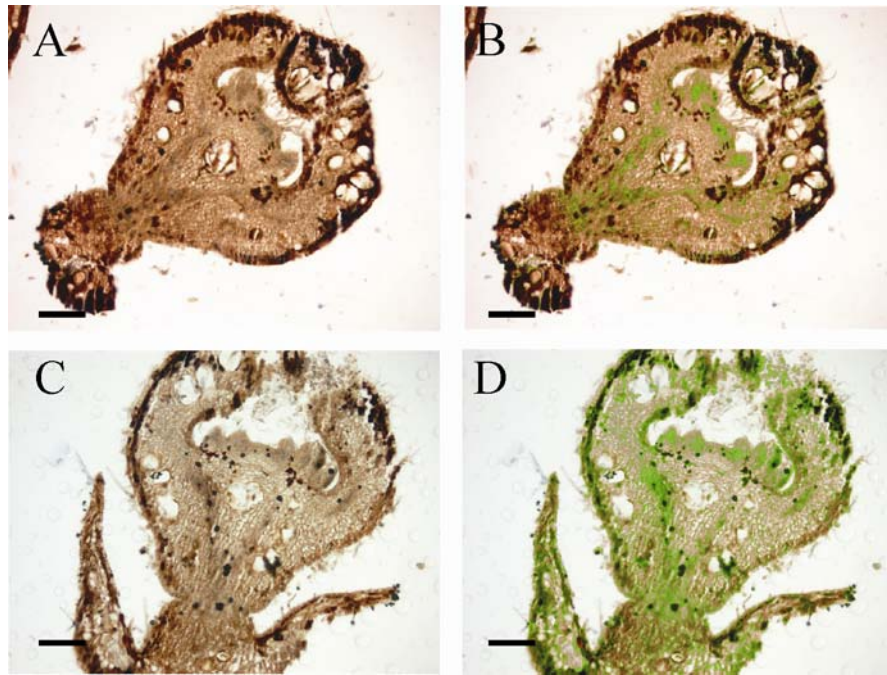


Figure 4-8: *In situ* hybridization of *Arabidopsis PI* in stage 6 PSU-SCA6 flower buds. **A - D**, Hybridization of anti-sense probe to stage 6 flower buds. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

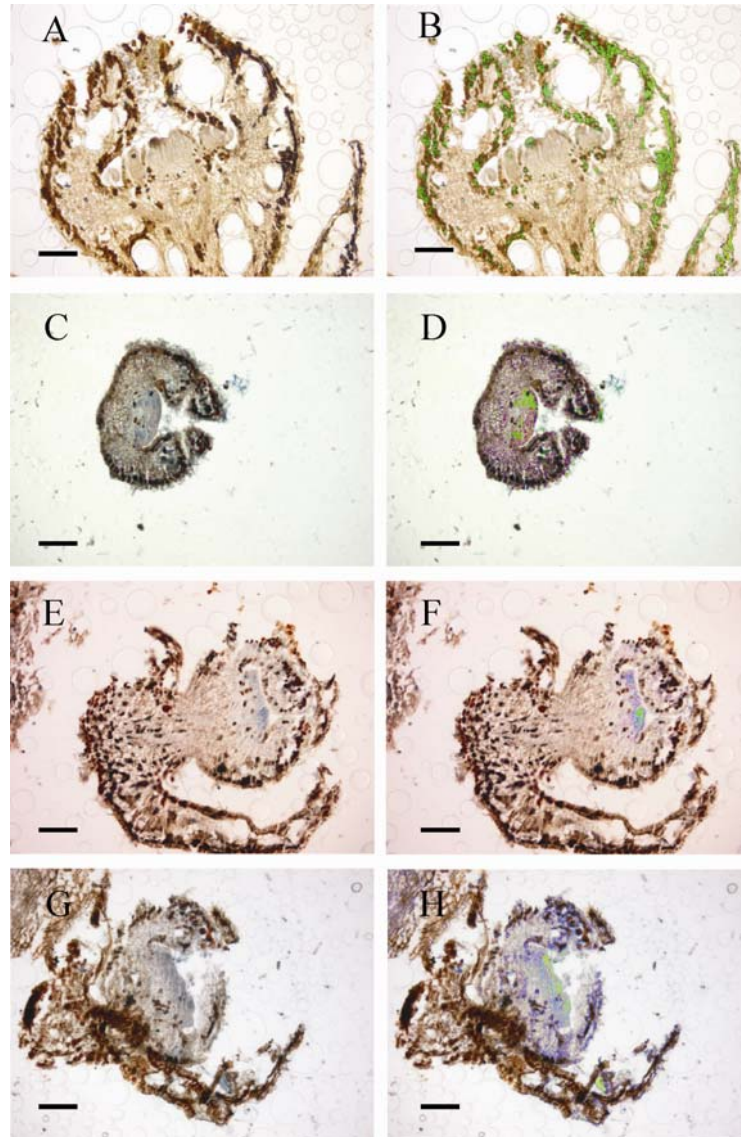


Figure 4-9: *In situ* hybridization of *Arabidopsis* AG in PSU-SCA6 plants. **A-B**, Negative control sense AG probe hybridized to a late stage 6 flower bud. **C-D**, Hybridization of anti-sense probe to a late stage 4 flower bud. **E-F**, Hybridization of anti-sense probe to a late stage 5 flower bud. **G-H**, Hybridization of anti-sense probe to an early stage 6 flower bud. Left hand panels are raw images; right hand panels have had the hues of the NCB/NCIP precipitate selected in Adobe® Photoshop and replaced by green hues (see Methods). (Bar = 100 μ m)

4.4 Discussion

4.4.1 Determination of *in situ* hybridization probes

To determine the copy number of *LFY*-related genes in the *T. cacao* genome, Southern blot hybridization was performed using the *TcLEAFY* genomic PCR fragment as a probe. Although underexposed, this analysis revealed a single hybridizing *LFY* fragment for each of three lanes of genomic *T. cacao* DNA cut with different restriction enzymes, thus supporting the hypothesis that *LFY* is a single gene. These data are supported by the literature where it is recorded that *LFY* is a single copy gene, with the exception of a single case (*Eucalyptus spp.*) where two *LFY* orthologs have been identified (Kelly et al., 1995; Southerton et al., 1998). Unfortunately, due to underexposure of the autoradiogram, the presence of other *LEAFY*-like genomic loci cannot be ruled out.

The ABC genes used in these experiments were obtained from the Ma lab (Department of Biology, Penn State University) and had been used in previous *in situ* hybridization experiments in *Arabidopsis*. The DNA fragments used as probes had previously been checked for specificity and cross hybridization by comparing the DNA sequences to the NCBI database (Zahn pers. comm.). The probes were also designed so that they would not incorporate conserved regions such as the MADS box (Yanofsky et al., 1990; Jack et al., 1992; Mandel et al., 1992; Goto and Meyerowitz, 1994; Gustafson-Brown et al., 1994).

Dot blot hybridization of the *Arabidopsis AP1*, *AP3*, *PI* and *AG*, cDNAs and *T. cacao LFY* and *CELLULOSE SYNTHASE* cDNAs to genomic DNA from *T. cacao* revealed some hybridization. Although hybridization was weaker when compared to *Arabidopsis* genomic DNA, and there was no negative control spot used, these results provide some evidence that *Arabidopsis* probes are able to hybridize to *T. cacao* DNA

and vice versa. It is also possible that one or more copies of each class will hybridize (e.g. *API* and *CAL* may cross-hybridize, as could *AG* with *SHPI* and *SHPI2*) or *T. cacao* may have more than a single copy of each gene.

As a positive control, a gene expected to be constitutively expressed in the growing flower bud was chosen (Amor et al., 1995). A *T. cacao* *CELLULOSE SYNTHASE* cDNA was hybridized to sections of developing *T. cacao* flower bud tissue. The expression pattern of the *CELLULOSE SYNTHASE* gene in *Arabidopsis* documented in the GENEVESTIGATOR microarray gene database showed that *CELLULOSE SYNTHASE* would be expressed in all tissues and at all stages of floral development (Zimmermann et al., 2004) (Appendix D). *In situ* hybridization analysis of the *CELLULOSE SYNTHASE* probe to sections of stage 4 - 9 *T. cacao* flowers revealed that *CELLULOSE SYNTHASE* was expressed in all flower buds and peripheral tissue observed, implying that the *in situ* hybridization protocol used was able to detect *CELLULOSE SYNTHASE* mRNA in the expected pattern (i.e. all cells). Sense probe was generated for all probes and used as a negative control. In all sections observed, there was some weak nonspecific localized hybridization observed due to nonspecific binding of the probe, which was regarded as background.

4.4.2 LEAFY

The product of the *LFY* gene is one of the major floral integrators and regulators of ABC gene activation (see Chapter 1.2.2). In *Arabidopsis*, *LFY* is expressed in the shoot apex and floral meristem tissues as the transition from a vegetative stage to an inflorescence state takes place (Weigel et al., 1992). This observation is supported by *LFY* data contained in the GENEVESTIGATOR microarray gene database (Appendix D) (Zimmermann et al., 2004). Our observations of *T. cacao* *LFY* showed that it was expressed in the floral meristematic tissues during stage 4 -7 flower buds. Similar expression patterns have been observed in *LFY* experiments in *Arabidopsis*. We did not

conduct *in situ* hybridization experiments beyond stage 4 - 7, so it is unknown what the expression patterns of *LFY* are before and after this period. However, these data indicate that relative temporal and spatial expression of the *T. cacao LFY* gene is similar to that of *Arabidopsis* at these stages of development.

LFY homologs have been identified for a number of species including: *Antirrhinum majus* (*FLORICAULA*), tomato (*FALSIFLORA*), *N. tabacum* (*NFLI*), *Populus* (*PTLF*), *Eucalyptus* (*ELF1*, and 2), *Pinus radiata* (*NEEDLY*), and *Vitis vinifera* (*VFL*) (Coen et al., 1990; Mouradov et al., 1998; Southerton et al., 1998; Molinero-Rosales et al., 1999; Rottmann et al., 2000; Ahearn et al., 2001). In *Arabidopsis* and *A. majus* expression begins in the region of the inflorescence meristem and continues through to the formation of the floral meristem, where there is low expression observed. In tobacco *NFLI* expression begins in the inflorescence meristem. In pine, there are no floral organs, however, *NEEDLY* expression is observed in the reproductive meristem. We observed *LFY* expression during organogenesis in *T. cacao*; however, floral transition was not tested. One interesting point is that there have been an increasing number of reports of *LFY* expression in other meristems whose fates are not necessarily floral (Mouradov et al., 1998; Carmona et al., 2002). This could potentially lead to at least two possibilities: the first is that there are, yet, undiscovered *LFY*-like genes that have similar functions in other alternatively fated meristems, for example in pea, *LFY* is important for compound leaf development (Hofer et al., 1997). The second is that there is a single *LFY* gene that functions to integrate signals irrespective of meristem fate. The first hypothesis is somewhat supported by the discovery of two *LFY* orthologs in *Eucalyptus* (Southerton et al., 1998). However, due to the fact that only a single *LFY* ortholog exists in many species (Mouradov et al., 1998; Southerton et al., 1998; Molinero-Rosales et al., 1999; Rottmann et al., 2000; Ahearn et al., 2001; Carmona et al., 2002), the second hypothesis has not been ruled out.

4.4.3 *APETALA1*

The A class gene *API* was reported in *Arabidopsis* to be expressed initially throughout the meristem but in later floral stages was reported to be only found in whorls 1 and 2 as the C gene *AG* prevents *API* accumulation (Mandel et al., 1992; Gustafson-Brown et al., 1994). These observations are supported by the GENEVESTIGATOR microarray gene database, which reported high levels of *API* in the inflorescence and flower (Appendix D) (Zimmermann et al., 2004). High levels of expression were also reported in the pedicel. *T. cacao* showed the same expression patterns consisting of initial expression across the entire meristem (stage 3 – 5), which at stage 6 become limited to only whorls 1 and 2. This is consistent with what has been observed in *Arabidopsis*, and provides some evidence that the expression of *API* is conserved between the two species. Stages beyond 3 – 6 were not tested.

API homologs have been identified in other species including: *Antirrhinum majus* (snapdragon) *SQUA*, *Pisum sativum* (pea) *PEAM4*, and *Silene latifolia* (white campion) *SLM4* (Huijser et al., 1992; Hardenack et al., 1994; Berbel et al., 2001; Gocal et al., 2001; Irish, 2003). Unlike *Arabidopsis*, it has been reported that the *Antirrhinum API* ortholog *SQUAMOSA* appeared not to be required for the specification of floral meristem identity as demonstrated by the inability of *squamosa* mutants being able to produce normal flowers (Huijser et al., 1992). Alternatively, it might function redundantly with other genes.

4.4.4 *APETALA3*

Expression of the B class gene *AP3* has been observed in whorls 2 and 3 during organogenesis in *Arabidopsis* (Weigel and Meyerowitz, 1993) and was reported in the GENEVESTIGATOR microarray gene database as being expressed primarily in the petals and slightly less in the stamen (Appendix D). Surprisingly, *AP3* also seems to be

expressed (according to the GENEVESTIGATOR microarray gene database (Zimmermann et al., 2004)) beyond organogenesis right through to silique formation, whereas *PI* is not. In *T. cacao* we observed sections of stage 4 – 7 floral buds and saw expression exclusively in the second and third whorls where petal and anther primordia were forming. These results suggest that *T. cacao* has similar expression patterns to *Arabidopsis*, implying similarities in gene function. The expression patterns of *AP3* were not tested beyond stage 7 or before stage 4. It is important to note that *AP3* expression was observed in the infertile staminode primordia, providing some evidence that staminode identity could also be conferred by *AP3*.

Orthologs of *AP3* have been identified in many species including *Antirrhinum* (*DEF*), *Petunia hybrida* (*PMADS1*), *Nicotiana tabacum* (*NTDEF*), *Silene latifolia* (*SLM3*), *Gerbera hybrida* (*GDEF2*), and *Zea mays* (*SILKY*) (Schwarzsommer et al., 1990; Irish, 2003). It has been reported that *AP3* orthologs have similar expression patterns for all core dicots investigated (Jack, 2004).

4.4.5 PISTILLATA

The B class gene *PI* is expressed in whorls 2 and 3 of *Arabidopsis* flowers during organogenesis (Weigel and Meyerowitz, 1993). Examination of the GENEVESTIGATOR microarray gene database (Zimmermann et al., 2004) confirmed these observations by reporting expression in the petals and stamens of *Arabidopsis* (Appendix D). Expression of *PI* in *T. cacao* exhibited a similar expression pattern, with expression occurring in the petals and anthers (whorls 2 and 3) for the stages observed (stage 6 – 7). Like *AP3*, *PI* expression was also observed in staminode primordia adding further evidence that *AP3* and *PI* (B class genes) could be implicated with staminode identity. Like *AP3*, *PI* shares similar expression patterns across all species in which it has been investigated. These species include: *Antirrhinum* (*GLO*), *Petunia hybrida*

(*FBP1*), *Nicotiana tabacum* (*NTGLO*), *Silene latifolia* (*SLM2*), and *Oriza sativa* (*OsMADS2*) (Schwarzsommer et al., 1990; Irish, 2003).

One interesting phenomenon reported in *Eranthis hyemalis* (Skipper, 2002) was that expression of *AP3* and *PI* orthologs was observed not only in the petal and anther whorls (2 and 3), but also in developing vascular bundles. It was suggested by Skipper (2002) that *PI* and *AP3* may have a possible involvement in the development of these tissues. This was also observed in the developing vascular bundles of *T. cacao* (Figure 4-8). Although the hypothesis suggested by Skipper (2002), could be correct, we are unable to determine its validity in *T. cacao*, for the following reasons: 1) our use of heterologous probes, which could be cross hybridizing to different genes of similar sequence in *T. cacao*, and 2) the possibility of the probe hybridizing to another gene family member of *AP3* or *PI*.

4.4.6 AGAMOUS

The C class gene *AGAMOUS* is expressed across the central part (whorls 3 and 4) of the meristem in initial developmental stages in *Arabidopsis*. Once the organ primordia form, the expression continues, eventuating in some expression in the fully developed carpel and developing silique (Yanofsky et al. 1992; Zhan pers. Comm.). This report is supported by the *AG* data contained in the GENEVESTIGATOR microarray gene database (Zimmermann et al., 2004) which reports expression primarily in the carpel, then the stamens and silique (Appendix D). Similar expression was observed for *T. cacao*, which showed expression across the central parts of the meristem between stages 3 - 6, where expression was in the central two whorls. *AG* expression is also observed in regions that will eventuate in staminode primordia indicating that C class genes could be involved in specifying staminode identity. The expression patterns of *AG* were not tested in *T. cacao* before stage 3 and after stage 6. However, between stages 3 – 6 these data imply similar expression patterns across the two species.

When comparing other species it is important to consider the lineage of the *AG* gene. It has been recently shown that the functional homolog of *AG* from *Antirrhinum* (*PLE*) arose from separate paralogous lineages rather than orthologous lineages (Irish, 2003; Kramer et al., 2004). These two lineages have been designated the *euAG* and *PLENA* (*PLE*) lineages. The *euAG* lineage includes *Arabidopsis AG* and *Antirrhinum FARINELLI* (*FAR*), while the *PLE* lineage includes the *Antirrhinum PLE* and *Arabidopsis SHATTERPROOF1* and *2* (*SHP1*, *2*) (Irish, 2003). In *Arabidopsis* expression of the *AG* genes are at similar levels in both stamen and carpel organs. However, the *SHP1* and *2* genes show higher expression levels in the carpel organs. In comparison, the *Antirrhinum* equivalents of the *AG* (*FAR*) and *SHP* (*PLE*) genes have equal expression in stamens and carpel organs (Irish, 2003; Kramer et al., 2004). This indicates that the *SHP 1* and *2* genes would be good targets for further study as higher expression of *T. cacao SHP 1* or *2* orthologs in *T. cacao* carpel organs would indicate further conservation in gene expression patterns between stamens and carpels between *T. cacao* and *Arabidopsis*.

4.4.7 Proposal of a *T. cacao* model of ABC flower development

In *T. cacao* and *Arabidopsis* the spatial and temporal expression patterns for the *LFY* gene, and homologs of *AP1*, *AP3*, *PI*, and *AG* are very similar for the stages which were observed during floral development. The only deviation from what was observed in *Arabidopsis* was that the B class genes *AP3* and *PI*, and the C class gene *AG* express in regions where staminode primordia develop, an organ that *Arabidopsis* does not possess. These expression patterns provide some evidence that staminode identity may be determined by combinations of B and C class genes.

Analysis of the literature comparing the expression of these five genes in other species studied was not surprising. While some expression differences were

observed in later stages, primarily due to differences in morphology of the final flower, in general, there is a high degree of conservation of gene expression patterns in earlier stages (Jack, 2004; Ma, 2005). These results imply that although there is a lot of diversity among individual floral morphologies among species, the fundamental unifying principals stated in Jack (2004) are probably correct in that all flowers require, to some degree, the underlying mechanisms allowed by the ABC model and *LFY*. This in turn indicates that genes underlying meristem formation and subsequent organogenesis have similar expression programs.

Based on these results, we are able to propose a version of the ABC model for *T. cacao* based on our observations (Figure 4-10). This model is consistent with what has been described for *Arabidopsis* except we are able to include the staminode organs which like stamens are proposed to require *PI*, *AP3* and *AG* for identity. It is important to note that this model is proposed exclusively on the similarity of gene expression patterns of the observed ABC genes. To validate this model, functional and mutant studies will need to be carried out in *T. cacao* similar to what has been done for *Arabidopsis*.

The *SEPPELLATA 1, 2, 3* and *4* (*SEP1, 2, 3,* and *4*) “E” genes have been included in the model shown in Figure 4-10. Recently, in *Arabidopsis* the ABC model has been extended to include the *SEP 1-4* genes as being important for the identity of all whorls in *Arabidopsis* along with the regular ABC genes (Ma, 2005). *APETALLA2* (*AP2*) is also included in the model in Figure 4-10. Although neither the *SEP* nor *AP2* were tested in these analyses, it is important to indicate their possible role in flower meristem identity based on their involvement in that in many other species (Ma, 2005). It is expected that further analyses will be completed in the future to confirm the role of these genes.

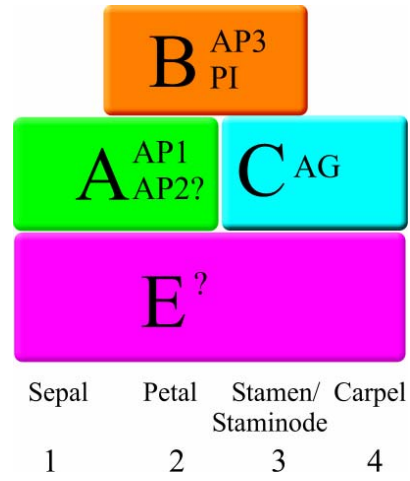


Figure 4-10: ABC model for *T. cacao* (modified from Jack, 2004)

4.5 References

- 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
- Ambrose BA, Lerner DR, Ciceri P, Padilla CM, Yanofsky MF, Schmidt RJ** (2000) Molecular and genetic analyses of the *SILKY1* gene reveal conservation in floral organ specification between eudicots and monocots. *Molecular Cell* **5**: 569-579
- Amor Y, Haigler CH, Johnson S, Wainscott M, Delmer DP** (1995) A membrane-associated form of *SUCROSE SYNTHASE* and its potential role in synthesis of cellulose and callose in plants. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 9353-9357
- Angenent GC, Colombo L** (1996) Molecular control of ovule development. *Trends in Plant Sciences* **1**: 228-232
- Berbel A, Navarro C, Ferrandiz C, Canas LA, Madueno F, Beltran JP** (2001) Analysis of *PEAM4*, the pea *API* functional homologue, supports a model for *API*-like genes controlling both floral meristem and floral organ identity in different plant species. *Plant Journal* **25**: 441-451
- Blazquez MA, Soowal LN, Lee I, Weigel D** (1997) *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**: 3835-3844
- Bowman JL, Smyth DR, Meyerowitz EM** (1991) Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**: 1-20
- Carmona MJ, Cubas P, Martinez-Zapater JM** (2002) *VFL*, the grapevine *FLORICAULA/LEAFY* ortholog, is expressed in meristematic regions independently of their fate. *Plant Physiology* **130**: 68-77
- Chen XM** (2004) A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* **303**: 2022-2025
- Coen ES, Meyerowitz EM** (1991) The war of the whorls - Genetic interactions controlling flower development. *Nature* **353**: 31-37
- Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R** (1990) *FLORICAULA* - A homeotic gene required for flower development In *Antirrhinum majus*. *Cell* **63**: 1311-1322
- Egea-Cortines M, Saedler H, Sommer H** (1999) Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*. *Embo Journal* **18**: 5370-5379
- Fan HY, Hu Y, Tudor M, Ma H** (1997) Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *Plant Journal* **12**: 999-1010
- Fornara F, Marziani G, Mizzi L, Kater M, Colombo L** (2003) MADS-box genes controlling flower development in rice. *Plant Biology* **5**: 16-22
- Freeman WM, Walker SJ, Vrana KE** (1999) Quantitative RT-PCR: pitfalls and potential. *Biotechniques* **26**: 112-122

- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J** (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiology* **124**: 1570-1581
- Gocal GFW, King RW, Blundell CA, Schwartz OM, Andersen CH, Weigel D** (2001) Evolution of floral meristem identity genes. Analysis of *Lolium temulentum* genes related to *APETALA1* and *LEAFY* of *Arabidopsis*. *Plant Physiology* **125**: 1788-1801
- Goto K, Meyerowitz EM** (1994) Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes & Development* **8**: 1548-1560
- Gustafson-Brown C, Savidge B, Yanofsky MF** (1994) Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* **76**: 131-143
- Hardenack S, Ye D, Saedler H, Grant S** (1994) Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. *Plant Cell* **6**: 1775-1787
- Hofer J, Turner L, Hellens R, Ambrose M, Matthews P** (1997) *UNIFOLIATA* regulates leaf and flower morphogenesis in pea. *Current Biology* **7**: 581-587
- Honma T, Goto K** (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**: 525-529
- Huijser P, Klein J, Lonig WE, Meijer H, Saedler H, Sommer H** (1992) *bracteomania*, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene *SQUAMOSA* in *Antirrhinum majus*. *Embo Journal* **11**: 1239-1249
- Immink RGH, Gadella TWJ, Ferrario S, Busscher M, Angenent GC** (2002) Analysis of MADS box protein-protein interactions in living plant cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 2416-2421
- Irish VF** (2003) The evolution of floral homeotic gene function. *Bioessays* **25**: 637-646
- Iscove NN, Barbara M, Gu M, Gibson M, Modi C, Winegarden N** (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nature Biotechnology* **20**: 940-943
- Jack T** (2004) Molecular and genetic mechanisms of floral control. *Plant Cell* **16**: S1-S17
- Jack T, Brockman LL, Meyerowitz EM** (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* **68**: 683-697
- Jackson D** (1992) *In situ* hybridization in plants. In SJ Gurr, MJ McPherson, DJ Bowles, eds, *Molecular Plant Pathology: a practical approach*, Vol 1. Oxford University Press, Oxford
- Jofuku KD, den Boer BG, Van Montagu M, Okamoto JK** (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene *APETALA2*. *Plant Cell* **6**: 1211-1225
- Kelly AJ, Bonnländer MB, Meeks-Wagner DR** (1995) *NFL*, the tobacco homolog of *FLORICAULA* and *LEAFY*, is transcriptionally expressed in both vegetative and floral meristems. *Plant Cell* **7**: 225-234
- Kramer EM, Jaramillo MA, Di Stilio VS** (2004) Patterns of gene duplication and functional evolution during the diversification of the *AGAMOUS* subfamily of MADS box genes in angiosperms. *Genetics* **166**: 1011-1023

- Lenhard M, Bohnert A, Jurgens G, Laux T** (2001) Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* **105**: 805-814
- Lohmann JU, Hong RL, Hobe M, Busch MA, Parcy F, Simon R, Weigel D** (2001) A molecular link between stem cell regulation and floral patterning in *Arabidopsis*. *Cell* **105**: 793-803
- Lynn K, Aida M, Sedbrook J, Masson P, Tasaka M, Barton MK** (1999) The *PINHEAD/ZWILLE* gene acts pleiotropically in *Arabidopsis* development and has overlapping functions with the *ARGONAUTE 1* gene. *Development* **126**: 469-481
- Ma H** (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annual Reviews in Plant Biology* **56**: 393-434
- Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF** (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**: 273-277
- McGonigle B, Bouhidel K, Irish VF** (1996) Nuclear localization of the *Arabidopsis* *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes & Development* **10**: 1812-1821
- Meyerowitz EM** (1987) *In situ* hybridization of RNA in plant tissue. *Plant Molecular Biology Reporter* **5**: 242-250
- Meyerowitz EM, Running MP, Sakai H, Williams RW** (1998) Multiple modes of cell division control in *Arabidopsis* flower development. *Symp Soc Exp Biol* **51**: 19-26
- Molinero-Rosales N, Jamilena M, Zurita S, Gomez P, Capel J, Lozano R** (1999) *FALSIFLORA*, the tomato orthologue of *FLORICAULA* and *LEAFY*, controls flowering time and floral meristem identity. *Plant Journal* **20**: 685-693
- Moon YH, Kang HG, Jung JY, Jeon JS, Sung SK, An G** (1999) Determination of the motif responsible for interaction between the rice *APETALA1/AGAMOUS-LIKE9* family proteins using a yeast two-hybrid system. *Plant Physiology* **120**: 1193-1203
- Mouradov A, Glassick T, Hamdorf B, Murphy L, Fowler B, Maria 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 of the United States of America* **95**: 6537-6542
- Nilsson O, Lee I, Blazquez MA, Weigel D** (1998) Flowering-time genes modulate the response to *LEAFY* activity. *Genetics* **150**: 403-410
- Parcy F, Nilsson O, Busch MA, Lee I, Weigel D** (1998) A genetic framework for floral patterning. *Nature* **395**: 561-566
- Ramonell KM, Zhang B, Ewing RM, Chen Y, Xu D, Stacey G, Somerville S** (2002) Microarray analysis of chitin elicitation in *Arabidopsis thaliana*. *Molecular Plant Pathology* **3**: 301-311
- Riechmann JL, Meyerowitz EM** (1997) MADS domain proteins in plant development. *Biological Chemistry* **378**: 1079-1101
- Rottmann WH, Meilan R, Sheppard LA, Brunner AM, Skinner JS, Ma CP, Cheng SP, 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*. *Plant Journal* **22**: 235-245
- Schwarzsummer Z, Huijser P, Nacken W, Saedler H, Sommer H** (1990) Genetic-control of flower development by homeotic genes in *Antirrhinum majus*. *Science* **250**: 931-936
- Scutt CP** (2001) *In situ* hybridization as a tool for functional genomics. *Ecole thematique Biologie vegetale*: 1-4
- Skipper M** (2002) Genes from the *APETALA3* and *PISTILLATA* lineages are expressed in developing vascular bundles of the tuberous rhizome, flowering stem and flower primordia of *Eranthis hyemalis*. *Annals Of Botany* **89**: 83-88
- Southerton SG, Strauss SH, Olive MR, Harcourt RL, Decroocq V, Zhu XM, Llewellyn DJ, Peacock WJ, Dennis ES** (1998) Eucalyptus has a functional equivalent of the *Arabidopsis* floral meristem identity gene *LEAFY*. *Plant Molecular Biology* **37**: 897-910
- Theissen G, Becker A, Di Rosa A, Kanno A, Kim JT, Munster T, Winter KU, Saedler H** (2000) A short history of MADS-box genes in plants. *Plant Mol Biol* **42**: 115-149
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM** (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843-859
- Weigel D, Meyerowitz EM** (1993) Activation of floral homeotic genes in *Arabidopsis*. *Science* **261**: 1723-1726
- Yang YZ, Xiang HJ, Jack T** (2003) *pistillata-5*, an *Arabidopsis* B class mutant with strong defects in petal but not in stamen development. *Plant Journal* **33**: 177-188
- 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
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W** (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiology* **136**: 2621-2632

Chapter 5

Conclusion



Void

5.1 Flower Development of *T. cacao*

The overarching objective of this thesis research was to explore *T. cacao* flower development. This objective was addressed at three levels; through a detailed morphometric analysis of flower development, a comparison of DNA sequences, and via a high spatial resolution gene expression analysis. We initially analyzed the morphological development of *T. cacao* and concluded that it took 30 days for the average 5.7mm x 3.6mm flower bud to mature (~2x longer and larger than *Arabidopsis*). During development, the flower bud growth rate decreased to 0.088 mmd⁻¹ and completed development at a maximum growth rate of 0.54 mmd⁻¹. We suggest that the switch between the low to high growth rates signified the transition between net differentiation and net elongation of the developing cells. Using data obtained from measuring the organs inside the developing flower bud, we were able to build mathematical models that allow us to predict individual organ growth in terms of time and length of the developing flower bud. Furthermore, we were able to subdivide flower development into 12 stages corresponding to those defined in *Arabidopsis* (Smyth et al., 1990). The duration of these stages were shown to have a similar relationship to the duration of those in *Arabidopsis* in terms of percent of total time. Finally, we determined that the flowers opened between 6pm on the 29th day of flower development and concluded at 6am the following morning.

To assess DNA sequence similarity between *Arabidopsis*, *T. cacao*, and cotton, we used degenerate primers to amplify genes of interest. Out of fifty-nine candidate genes we were able to extract forty-two putative genes from *T. cacao* including the floral integration gene *LEAFY*. In comparing the *T. cacao* gene fragments with their *Arabidopsis* and cotton orthologs it was observed that *Arabidopsis* and *T. cacao*, and *Arabidopsis* and cotton shared similar p-distances from each other (p-dist.=0.189 and p-dist.=185, respectively), while *T. cacao* and cotton had a shorter p-distance (p-dist.=0.095). This suggested that they were more closely related to each other compared

to *Arabidopsis*. Twenty nine genes were selected for phylogenetic analysis and 24 of these showed the expected phylogeny based on chloroplast and mitochondrial DNA sequence data. Furthermore, when compared to other genes from *Arabidopsis* gene families (tribes) there were five cases where multiple putative gene family members were isolated from *T. cacao*. Using the 34 most robust primers, other species in the orders Eurosids I and II were surveyed to see if the PCR primers would be able to amplify product. It was found that in eurosids II, an average of 25 primers were able to produce PCR product and in eurosids I, 17 primers supported amplification. Moreover, when a cluster analysis was performed on the PCR products, it was seen that the respective species' groupings reflected their phylogenetic relationship.

The expression of *LEAFY*, along with the *Arabidopsis* genes *API*, *AP3*, *PI*, and *AG*, was observed in developing *T. cacao* floral tissue primarily during early floral organ developmental stages. In each case the genes temporal and spatial expression was similar to what has been described in other species including *Arabidopsis*. These results may be summarized as follows: *LEAFY* expression is observed across the entire meristem during early developmental stages and expression reduces as organs form. *API* is initially expressed throughout the developing floral meristematic tissues but dissipates to whorls 1 and 2 as *AG* expression inhibits *API* (stage 6). *AP3* and *PI* are observed to express in the meristematic tissue exclusively in whorls 2 and 3, and *AG* are observed to express in the central two whorls (3 and 4) only. These data support the hypothesis that the ABC model is valid for *T. cacao*, thus we are able to propose an ABC model for flowering in *T. cacao* although it requires further functional and mutant analysis to test its validity.

5.2 Is *T. cacao* Really Like *Arabidopsis*?

Another purpose of this thesis was the comparison of the model species *Arabidopsis thaliana* to the less well-studied species, *Theobroma cacao*. We initially

hypothesized that it is expected, due to the evolutionary relatedness of the two species, that the mechanisms and developmental pathways of *Arabidopsis* are highly conserved and comparable to those of *T. cacao*. This hypothesis leads to the prediction that at the molecular level, many of the genes and their expression patterns would be predicted to be very similar and functionally equivalent. The results of this thesis support this hypothesis using floral development as a model pathway.

At the morphological level, it is observed that the 12 equivalent stages of *T. cacao* flower development are similar to stages in *Arabidopsis*, especially when compared with reference to time. At the DNA level, an average 79% DNA sequence similarity is observed. At the expression level, we observe similar temporal and spatial expression patterns for the genes *LEAFY*, *API*, *AP3*, *PI*, and *AG* between what has been previously reported in *Arabidopsis* and the *T. cacao* sections that we studied.

One major morphological difference between the two species is the infertile staminodes in the 3rd whorl of *T. cacao*. Staminodes have been studied in other species including *T. cacao*, primarily as a morphological marker for evolutionary studies (Walker-Larsen and Harder, 2000). Expression analysis revealed that the B class genes *AP3* and *PI* were expressed during the identity of these organs, implying that they may have a role in staminode identity.

5.3 Flower Development: Implications and Reflections

This work supports that hypothesis that the general morphological developmental programs and gene expression patterns (for *LEAFY* and several ABC genes) are similar between *Arabidopsis* and *T. cacao* flowers. This also supports the two unifying principals outlined by Jack, (2004) regarding the conservation of the function of *LEAFY* and the ABC genes across all species. By taking only these results one could infer that there is little difference at all with regard to flower development in *T. cacao*.

However, when looking at the obvious differences in biology and final morphology of flowers in the two species there must be differences in development as well.

Another obvious difference between *Arabidopsis* and *T. cacao* is at the floral transition stage. *Arabidopsis* is an annual plant and forms an apex from which flowers arise in a helical pattern. After flower development, fertilization, seed set and dispersal, the plant will die. *Arabidopsis* occurs naturally throughout the temperate regions of the world and flowers annually in the spring from April – May. Conversely, *T. cacao* is a perennial where a compressed cincinnal cyme forms on the trunk of the tree and continues to produce flowers consistently throughout the year. *T. cacao*'s natural habitat is the tropical rainforests of the Amazon, usually within 20 degrees of the equator. By looking at each species' natural habitat one can infer that there are likely to be potential differences in the signals required to initiate flower development. In Chapter 1 it was stated that the requirements of two pathways, those that enable the floral transition (vernalization, developmental age, maturation) and those that promote the floral transition (hormones, photoperiod, light quality, and temperature), must be met in order for *Arabidopsis* to undergo the floral transition.

Due to *T. cacao* being a tropical crop where there are fewer fluctuations in temperature and light conditions, it can be expected that *T. cacao* would require a different balance of enabling and promoting floral transition factors compared to *Arabidopsis*. It could be expected that there would be little need for vernalization since *T. cacao* does not grow in temperate climates. Due to *T. cacao* living in a very constant environment it could be suggested that the external factors including photoperiod, light quality, or temperature would not play a critical role in floral transition. If this were correct, the major elements important for the floral transition of *T. cacao* would therefore be internal elements such as developmental age, vegetative promoters, and hormones. Once floral transition is complete it has been noted that minor fluctuations in temperature, humidity, and water content can have an effect on the general number of flowers produced on the plant (Sale, 1969, 1970, 1970). It is unknown if these

temperature, humidity, and water content related increases in floral flushing are due to these stimuli affecting the floral transition pathways, or if they are due to physiological factors such as nutrient allocation.

The second potential difference in the developmental pathways between *Arabidopsis* and *T. cacao* is in the gene pathways following organogenesis. There is currently very little known about the genes that are regulated by the ABC genes, however due to different morphologies of the two flowers it can be suggested that different genes would express for different periods of time between the two species. As more knowledge is gained from studies into the targets of the ABC genes, it will be interesting to see the differences in expression patterns and gene functions among species at these later developmental stages.

5.4 What is Next?

Based on the research outlined in this thesis there are several lines of study that can be pursued. Much of this is related to *T. cacao*; however, it is hoped that researchers may take some of this work and use it as a model for investigating their respective organism.

The first avenue for future research would be to compare *T. cacao* flowers before and during pathogenesis of *C. perniciosa* and study the mechanisms involved with the pathogen's ability to revert floral meristems to a vegetative state. It would be important to observe carefully the differing morphological changes that are undergone by floral meristems infected by *C. perniciosa*, especially at the early stages of floral development. These would be related to the observed 12 stages we have defined in *T. cacao*. Once the morphological changes are observed and characterized, hypotheses can be put forth to understand the mechanisms underlying these changes.

One hypothesis that I would test initially arises from the observation in *Arabidopsis* that *AG* and *LEAFY* mutants revert to the vegetative state indicating that these genes are required not only for flower identity, but also for maintenance of the flower meristem (Mandel et al., 1992; Weigel et al., 1992; Okamuro et al., 1996). If this hypothesis were correct, I would predict that the fungus *C. pernicioso* interferes with the expression of these two genes during pathogenesis, thus preventing floral initiation. These experiments would require *in situ* hybridization experiments of these genes in infected meristems. By extending this hypothesis, the interference in gene function could also be proposed to be due to the effects of gibberellins. This could be suggested due to the fact that gibberellins (GA) have been observed as having direct effects on flowering genes (King and Evans, 2003; Ben-Nissan et al., 2004; Yu et al., 2004). It could be hypothesized that the pathogen alters the GA levels in the plant by either inhibiting the production of GA or inhibiting GA itself. Evidence for this hypothesis is found in the flower reversion literature where it has been stated that flower reversion can be averted in *ag* and heterozygous *lfy* mutants by the addition of GA or the inhibition of some phytochromes. Therefore, it seems reasonable to suggest that the fungus could alter the GA concentrations which in-turn prevents *LFY* and/or *AG* from being expressed thus causing the flower to switch to the vegetative state.

The effect of gibberellins on flowering would be interesting to study in themselves. While in annual plants they are known to promote flowering, in perennials such as woody trees they have been recorded to have the opposite effect and are observed to inhibit flowering in several species including cherry, apricot, almond, and lemon (Brunner and Nilsson, 2004). Furthermore, GA inhibitors have been applied to these species and have promoted early flowering, with the exception of juvenile *Eucalyptus nitens* and juvenile *Populus*. It is unknown if GA has a promotional or inhibitory effect on flowering in *T. cacao*. One way to test this would be to examine the *T. cacao* *LEAFY* promoter. When the *Populus* *LEAFY* promoter is compared to the *Arabidopsis* *LEAFY* promoter, it is observed that they only share an 8 bp region of perfect similarity (Brunner and Nilsson, 2004). This region of conservation interacts with a GA-MYB transcription

factor in *Arabidopsis*, suggesting that GA does play a role in *LEAFY* expression. However this could vary between the two species (Brunner and Nilsson, 2004). Examination of the *LEAFY* promoter and subsequent comparison to the respective promoters of these two species may suggest which of the two species may provide the best model of GA interactions in *T. cacao* flowering.

An additional line of research could be to characterize further floral development of *T. cacao*. This would involve determining the full length *T. cacao* sequences of the genes used in this thesis. Full length sequences would be gained by hybridization of *Arabidopsis* orthologs (or *T. cacao* PCR products) to the available *T. cacao* genomic library (Clement et al., 2004). Once the full-length sequence has been determined, it would be useful to identify gene structure and compare it to *Arabidopsis*, as well as to repeat the *in situ* hybridization for some of these genes to verify the results presented in this thesis. The full-length genes would also be transformed into *Arabidopsis* mutants to see if the *T. cacao* genes are able to complement their respective functions in *Arabidopsis*. Research could involve the further characterization of other important flowering genes, especially those upstream of *LEAFY* or down stream of the ABC genes. It would be important to characterize them through DNA sequence comparison, expression, and transformation and to compare them to *Arabidopsis*. These further studies would be expected to identify differences in the flowering pathway between the two species.

It would be possible to compare the two pathways using the morphological models presented in Chapter 2 of this thesis. In using the corresponding stages in *Arabidopsis* and *T. cacao*, it would be possible to compare differing expression levels at the same stages of development between the two species. In general, I would expect genes involved with the ABC model in *Arabidopsis* and *T. cacao* to have similar temporal and spatial expression patterns. However, genes that are involved with sensing the environment (for example genes relating to the promotion of the transition to flowering) should have different gene expression patterns since the two species have

adapted to differing environments. Genes that direct the formation of the actual floral structures would be expected to show different expression patterns between the two species. This would be expected to account for the differing structures of the two flowers.

Another line of research exemplified by the GA example above is to look further at other evolutionarily closely related species that may have more similar morphological characteristics to *T. cacao* than *Arabidopsis*. One species that I would target is *Populus* as it has its entire genome sequenced, is easy to transform, and is a tree like *T. cacao*. It is more evolutionarily distant from *T. cacao* than *Arabidopsis*, being a member of Eurosids I (Soltis et al., 2000). Studies in *Populus* have indicated that plants deficient in phytochromes can speed up bud formation in short days, however no changes in GA levels have been observed in wild type plants under these conditions (Arora et al., 2003). Another species that would be good to compare is cotton, which is in the same family as *T. cacao* and has a large EST database and a relatively large number of researchers working on the species. Currently D. Baum (University of Wisconsin Madison) is in the process of identifying cotton genes involved in flowering and has recently completed a study of the morphology of the cotton flower (pers. comm.).

Finally, it has been reported that in citrus, constitutive expression of *LEAFY* or *API* significantly reduced generation time from six years to as little as one (Pena et al., 2001). Likewise *Arabidopsis* has been reported as having a significantly shorter time to flowering under constitutive *LEAFY* expression (from 3 wks to 1.5-2 wks) (Weigel et al., 1992; Blazquez et al., 1997). It would be interesting to use the transformation protocol in the Guiltinan laboratory (Maximova et al., 2003) to see if the generation time can be reduced for *T. cacao* using this method. If successful in *T. cacao*, a reduction from 3 years to 6-18 months could potentially be observed. One caveat to this experiment is that in *Populus spp.* *API* has been shown to have no effect on flowering time and *LEAFY* has some effect but is variable between individual clones (Brunner and Nilsson, 2004). It is suggested that this occurs in *Populus* because the

phase change of the tree only occurs after time has lapsed allowing establishment in its environment, and energy may not be diverted to flowering until phase change takes place (Brunner and Nilsson, 2004). This is an interesting observation when applied to *T. cacao* jorquette production. Jorquette production is often considered indicative of the juvenile/adult phase change and is followed shortly after by the initial floral transition. Therefore, if flowering time can be reduced by constitutive expression of *LEAFY* this could indicate that *LEAFY* is a very important component in phase change and can effectively over-ride any other mechanisms keeping the tree in its juvenile state. It would be expected however that *LEAFY* would not be the only component in phase change as *lfy* mutants in *Arabidopsis* still produce some flower-like structures.

5.5 References

- Arora R, Rowland LJ, Tanino K** (2003) Induction and release of bud dormancy in woody perennials: A science comes of age. *Horticulture Science* **38**: 911-921
- Ben-Nissan G, Lee JY, Borohov A, Weiss D** (2004) GIP, a *Petunia hybrida* GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. *Plant Journal* **37**: 229-238
- Blazquez MA, Soowal LN, Lee I, Weigel D** (1997) *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**: 3835-3844
- Brunner AM, Nilsson O** (2004) Revisiting tree maturation and floral initiation in the poplar functional genomics era. *New Phytologist* **164**: 43-51
- Clement D, Lanaud C, Sabau X, Fouet O, Le Cunff L, Ruiz E, Risterucci AM, Glaszmann JC, Piffanelli P** (2004) Creation of BAC genomic resources for cocoa (*Theobroma cacao* L.) for physical mapping of RGA containing BAC clones. *Theoretical and Applied Genetics* **108**: 1627-1634
- Jack T** (2004) Molecular and genetic mechanisms of floral control. *Plant Cell* **16**: S1-S17
- King RW, Evans LT** (2003) Gibberellins and flowering of grasses and cereals: prizing open the lid of the "florigen" black box. *Annual Reviews in Plant Biology* **54**: 307-328
- Mandel MA, Gustafson-Brown C, Savidge B, Yanofsky MF** (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. *Nature* **360**: 273-277
- Maximova S, Miller C, Antunez de Mayolo G, Pishak S, Young A, Gultinan MJ** (2003) Stable transformation of *Theobroma cacao* L. and influence of matrix attachment regions on GFP expression. *Plant Cell Reports* **21**: 872-883
- Okamoto JK, den Boer BGW, Lotys-Prass C, Szeto W, Jofuku KD** (1996) Flowers into shoots: Photo and hormonal control of a meristem identity switch in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 13831-13836
- Pena L, Martin-Trillo M, Juarez J, Pina JA, Navarro L, Martinez-Zapater JM** (2001) Constitutive expression of *Arabidopsis* *LEAFY* or *APETALA1* genes in citrus reduces their generation time. *Nature Biotechnology* **19**: 263-267
- Sale PJM** (1969) Flowering of cacao under controlled temperature conditions. *Journal of Horticulture Science* **44**
- Sale PJM** (1970) Growth and flowering of cacao under controlled relative humidities. *Journal of Horticulture Science* **45**: 119-132
- Sale PJM** (1970) Growth, flowering and fruiting of cacao under controlled soil moisture conditions. *Journal of Horticulture Science* **45**: 99-118

- Smyth DR, Bowman JL, Meyerowitz EM** (1990) Early flower development in *Arabidopsis*. *Plant Cell* **2**: 755-767
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS** (2000) Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Botanical Journal of the Linnean Society* **133**: 381-461
- Walker-Larsen J, Harder LD** (2000) The evolution of staminodes in angiosperms: Patterns of stamen reduction, loss, and functional re-invention. *American Journal Of Botany* **87**: 1367-1384
- Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM** (1992) LEAFY controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843-859
- Yu H, Ito T, Zhao Y, Peng J, Kumar P, Meyerowitz EM** (2004) Floral homeotic genes are targets of gibberellin signaling in flower development. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 7827-7832

Appendix A

Degenerate deoxyribonucleotide primers created from consensus sequences of a highly conserved set of *Arabidopsis* and *cotton* genes

i. Clones used to create degenerate primers

Clone Designation	Putative Functional Category	Putative Functional Assignment	Ghi Clone ID	Ath Match (GI:)	Ptr Match
AAA1	Amino acid metabolism and transport	Cystathionine Gamma-Synthase	GA_Ea0023F11	2852454	N/A
AAA2	Amino acid metabolism and transport	Glutamate Decarboxylase	GA_Ea0034H22	4406789	scaffold_44 2232602 2232978
AAA3	Amino acid metabolism and transport	Glutamate Dehydrogenase	GA_Ea0032H9	1336084	LG_XIII 4271909 4272118
AAA4	Amino acid metabolism and transport	Ketol-Acid Reductoisomerase Precursor	GA_Ea0026N04	6735378	scaffold_124 362230 362462
AAA5	Amino acid metabolism and transport	Putative 3-Isopropylmalate Dehydratase Large Subunit	GA_Ea0030B11	4584548	LG_X 5442736 5443894
CMT1	Carbohydrate metabolism and transport	Adenosine Kinase	GA_Ea0032C19	6681336	LG_X 19190448 19190926
CMT2	Carbohydrate metabolism and transport	putative fructokinase	GA_Ea0014D20	4589962	N/A
CMT3	Carbohydrate metabolism and transport	alpha-xylosidase precursor	GA_Ea0035H11	4163997	N/A
CMT4	Carbohydrate metabolism and transport	Nodulin-26	GA_Ea0034C12	4538967	N/A
CMT5	Carbohydrate metabolism and transport	Parafusin	GA_Ea0029N10	4914466	N/A
CDIV1	Cell division and chromosome partitioning	Actin 12	GA_Ea0018N20	1145693	LG_VI 11135617 11136208
CDIV2	Cell division and chromosome partitioning	Actin 2 7	GA_Ea0015L07	1703108	scaffold_185 131678 131955
CDIV3	Cell division and chromosome partitioning	Actin 58	GA_Ea0031A11	1002533	LG_VIII 3057058 3057379
COMET1	Coenzyme Metabolism	farnesyl pyrophosphate synthetase	GA_Ea0031A11	414611	N/A
COMET2	Coenzyme Metabolism	Ubiquitin-Activating Enzyme	GA_Ea0020J06	1703478	LG_IX 5558931 5575848
EPROD1	Energy production and conversion	(S)-2-Hydroxy-Acid Oxidase, Peroxisomal	GA_Ea0024E24	2832641	scaffold_146 434342 434987
EPROD2	Energy production and conversion	Betaine-Aldehyde Dehydrogenase Precursor	GA_Ea0027C08	5882731	LG_XV 5050576 5051801

EPROD3	Energy production and conversion	Isocitrate Dehydrogenase	GA_Ea0035D24	6227018	scaffold_44 2241030 2241376
EPROD4	Energy production and conversion	Pyruvate Dehydrogenase E1 Component Alpha Subunit	GA_Ea0025I05	2829869	LG_VIII 13241860 13242087
EPROD5	Energy production and conversion	Turgor-Responsive Protein 26g	GA_Ea0024O05	4587552	scaffold_158 139417 140786
FU1	Function unknown	Putative Udp-Glucose	GA_Ea0008F01	5080759	scaffold_204 186535 187206
FU2	Function unknown	cysteine proteinase RD19A precursor	GA_Ea0028A16	541856	N/A
GFP1	General function prediction only	Esterase D	GA_Ea0035F07	2618688	LG_VI 3066189 3066638
INORG1	Inorganic ion transport and metabolism	ATP sulfurylase precursor	GA_Ea0015E15	2129540	N/A
INORG2	Inorganic ion transport and metabolism	Nadph-Cytochrome P450 Reductase	GA_Ea0025H01	5730131	N/A
INORG3	Inorganic ion transport and metabolism	Putative Ascorbate Peroxidase	GA_Ea0030M06	6682242	N/A
LM1	Lipid metabolism	4-coumarate:CoA ligase 2	GA_Ea28K14	5702188	N/A
LM2	Lipid metabolism	acetyl-CoA carboxylase	GA_Ea0027F10	1100253	N/A
LM3	Lipid metabolism	Chalcone Synthase 1	GA_Ea0031O06	66543	scaffold_10558 415 650
LM4	Lipid metabolism	Putative Beta-Ketoacyl-Coa Synthase	GA_Ea0035J01	2760830	LG_XVIII 5237551 5237909
LM5	Lipid metabolism	Putative Chalcone synthase	GA_Ea0035J01	2760830	N/A
CHAP1	Molecular chaperones	Dtdp-Glucose 4,6-Dehydratase	GA_Ea0032K16	3738279	LG_II 15055523 15055942
CHAP2	Molecular chaperones	End13 Protein	GA_Ea0035K11	3860272	LG_IX 1499563 1500245
CHAP3	Molecular chaperones	putative ATPase	GA_Ea0035K11	3860272	N/A
CHAP4	Molecular chaperones	T-Complex Protein 1, Epsilon Subunit	GA_Ea0015G13	3024697	LG_VIII 12269596 12270820
CHAP5	Molecular chaperones	Hypothetical 74.3 Kda Protein (PMRD Intergen)	GA_Ea0011L12	6579211	N/A
NMTR1	Nucleotide metabolism and transport	adenylosuccinate synthetase	GA_Ea0031A05	4678286	N/A
NMTR2	Nucleotide metabolism and transport	Amidophosphoribosyltransferase Precursor	GA_Ea0023O03	469193	LG_IV 15205227 15205459
NMTR3	Nucleotide metabolism and transport	Ethylene-Inducible Protein Hever	GA_Ea0033N18	3335371	LG_XVI 11003698 11003966
NMTR4	Nucleotide metabolism and transport	pyruvate decarboxylase-1	GA_Ea0035C02	3688188	N/A

OMWAL1	Outer membrane, cell wall Biogenesis	Cellulose Synthase Catalytic Subunit	GA_Ea0008B17	6714431	scaffold_87 132557 132995
OMWAL2	Outer membrane, cell wall Biogenesis	Udp-Glactose 4-Epimerase	GA_Ea0015P15	2129759	LG_III 11734803 11735412
RRR1	Replication, repair, recombination	Alcohol Dehydrogenase Class III	GA_Ea0005B04	2129562	LG_II 23593252 23593559
SECMOB1	Secretion and motility	Preprotein Translocase Secy Subunit	GA_Ea0031F22	1049293	LG_XVIII 10039584 10040405
STRANS1	Signal transduction	putative cAMP-dependent protein kinase	GA_Ea0025M10	4580468	N/A
TRANS1	Transcription	U90212 DNA binding protein ACBF	GA_Ea0020M08	4835793	N/A
TRANSL1	Translation, ribosomal structure	40s Ribosomal Protein S3	GA_Ea0010P23	4582468	LG_VI 14830036 14831715
TRANSL2	Translation, ribosomal structure	CAB11	GA_Ea0009P07	4262180	LG_XIX 10892104 10892394
TRANSL3	Translation, ribosomal structure	Elongation Factor 2	GA_Ea0018C21	6056373	LG_VII 7296630 7296860
TRANSL4	Translation, ribosomal structure	Hypothetical 47_9 Kda Protein (Elongation Factor, by BLAST)	GA_Ea0025E03	6729016	scaffold_70 893399 893784

Arabidopsis (Ath) and Cotton (Ghi) DNA sequence tgas from www.genome.clemson.edu
Populus (Ptr) genomic DNA sequence tags from <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>

ii. Degenerate deoxyribonucleotide primers

Gene Name	forward primer	reverse primer
Cystathionine gamma-synthase	atgatgttctgctgntgc	aaactgaccacaccnccaaa
Glutamate decarboxylase	atngtcaccggagcnaatgt	gngcaatgaancctccact
Glutamate dehydrogenase	nccnatgaaagggtggaatca	catatctgngcnggaaacat
Ketol-acid reductoisomerase precursor	tcctcatggtganaactgc	agattgacgcaactcnggtc
Putative 3-isopropylmalate dehydratase large subu	gcacttgacacaagaaggta	cattctttccccagcttc
Actin 58	tgcncttgactatgancagga	ttgatctcatgctgctnng
Actin 12	cgaangctaaccgtgagaag	gtgtggctnacaccatctcc
Actin 2 7	ggagctgagagattccgtg	atcctccgatccagacactg
Dtdp-glucose 4,6-dehydratase	gcaacgttaatcccattggt	tcctgtctgnttcccatca
End13 protein	aagtggatgggtgaaagtga	gcctgatcaagagcataggg
Putative AAA-type atpase	aagtggatgggtgaaagtga	gcctgatcaagagcataggg
T-complex protein 1, epsilon subunit	tngctgtgagcatttgag	gcatcttcaantgctttgg
Adenosine kinase	gctgctgcaanaacaagg	acaaangcatcacctgcac
Fructokinase	tggatcaataaagttgatngtg	ccctttcaccagaagngtgac
Farnesyl pyrophosphate synthetase 1	cggttggtgnattgaatgg	agattcccatgtcaacaagaa
Ubiquitin-activating enzyme e1 1	tggacattggnccangctaa	ggtcncctgaggcacnta
Actin	tgggtcatggttggnatg	ggcacagtgtgaganacacc
B-tubulin	caaatgtgggatgcnaagaa	ccngtgtaccaatgcaagaa
Cellulose synthase	tggagaaatgaacagtttgg	tcccatcaaaccttcaaga
Eukaryotic initiation Factor	tgcaaggtgtngangtttc	atctgttctcncctatngc
Ubiquitin	tgcatgnttctgaaaacnc	gnagacgagcatancacttgc
(S)-2-hydroxy-acid oxidase	gctactccagcgttganga	tcaattgncagcaacata
Betaine-aldehyde dehydrogenase precursor	ctgaagnttggatgcnaag	ctgtcaacatgnggatng
Isocitrate dehydrogenase [nadp]	agngatttctngcncagg	agngcngatcctntggtcat
Pyruvate dehydrogenase e1 component alpha subunit	gattgttggngctcaggctc	tggactctngcggctct
Turgor-responsive protein 26g	gagatgtggaatccnctgg	gcnattgcttnccaatttc
Ethylene response factor AP2 domain transcrip fact	gaattgccgttatgtgacga	agagattcggcgaagaacg
Putative UDP-glucose sterol glucosyltransferase	ncnatggctgatcctgaact	ccacaccnctagcncctca
Esterase d	aagtggagaactggcgtatg	ggcatcncatctctccaag
Gibberellin regulatory protein	cngacatggctgangtgg	gagcatgctntcgancca
4-coumarate--coa ligase 1	tcgtngatcgattgaangaa	ntcaccacaaatgcaacagg
Acetyl-coa carboxylase	tngtgcagttgagacacagac	ccttctggctcaagnacatt
Chalcone synthase 1	tacatggctccttncctgga	gccaagcaaccttngtgta
Putative beta-ketoacyl-coa synthase	gntacaactgggagggaatgg	tttagagcatcacnngcaac
Adenylosuccinate synthetase precursor	tcctttgtnacttctccag	cagttgtngtccaaactcc
Amidophosphoribosyltransferase precursor	gngttgncacaaatggnaa	acagcnccttctnctct
Ethylene-inducible protein hever	gttgangcncagatcctga	tgaaaacctcatcatcca
Pyruvate decarboxylase isozyme 1	gcnggtgcttacagnaga	ttgctntcttcaagcngtt
Cellulose synthase catalytic subunit [udp-forming]	acngaagatgtggtcacagg	actgctcgtttcncacca
Udp-glucose 4-epimerase	ccaanaancatgctccta	aaacngctgngcatctct
Alcohol dehydrogenase class iii	tgttggtgaaagngtaactgaa	gaacaccacancgaagaagg
Preprotein translocase secy subunit	gaaggtgaagctggnagaaa	tcctctctgcttctgaacat
Rna-binding post-transcriptional regulator csx1	gtgatcgnganttgacag	ngccaagtctccaacaana
40s ribosomal protein s3	gttctnggtgagaagggna	cggagnttccactcactnat
60s ribosomal protein l10	gaagctgncgtattgcttg	ncgaccaggaactgaact
Elongation factor 2 (ef-2)	tcctgggtgagaagaang	ggaagntcagacgcaacctt
Hypothetical 47_9 kda protein m021b04_12	cctggnagnacngttcagat	ttgctttagntgancagn

Appendix B

Sequences of *T. cacao* PCR fragments amplified with the degenerate primer set

>Tca_ Glutamate Decarboxylase

attgtcaccggagcgaatgtgcaggttgctgggagaagtttgctagatacttcgaggttgagctgaagg
 aagtgaagctgaaagaggatattatgtgatggatcctgtaaaagcagttgatatggtgatgaaaatac
 catttggtgcagccattctgggatctaccctgactggggagtttgagaatgtaaagctccttcatac
 ctctcaccaacaagaataaggagactggtgggacacccaatacatggtgatgctgccagtgaggat
 tcattgccc

>Tca_ Glutamate Dehydrogenase

atagtttgattctgatcctatgaataatctaggtgacctgatgaagtgaatgcttagcacaac
 taatgacatggaagactgcagttgccaatcccataatggtggggctaaaggtgggataggatgtaacc
 aggagagtaagcatatctgaactggaacggcttactcagtttcaccagaagatacatgacctgatt
 ggagttcacacagatgtcccgtccagatatg

>Tca_ Ketol-Acid Reductoisomerase Precursor

tccttcattggtgatactgctctacaacagcacgattgggtcaaggaaatgggcccctcgattcgatta
 catccttaccacagcgtttggtggcagttgataagggcaccatcaaccaagatctgatcagcaac
 ttctgtcagaccgggtgcacgggtgcccattgaagtgtgtcagttgagcccacagttgacattcag
 tgnccccagatgctgattttgtgccgaccnngagttgcgtaactc

>Tca_ Putative 3-Isopropylmalate Dehydratase Large Subunit

gcacttgacaagaaggtcactgcaggcctggggaggtttgttagggacagattctcacacctgcactg
 ctggagcatttgtaattgtctacaggaataggaaatacagatgcaggccttcttgggcactgggaa
 gcttctactcaaggtgccaccgactttgagattgtgatggatggtaaatgctcactattgcttga
 aaggatttgatttacagatcattggtgaaatctgtagctggtgcaacatacaaatccatggagttg
 ttggaacaactgttgaaagttaaacatggaagaacgtatgactttatgcaatatggttgtaagctgg
 gggaaagaatg

>Tca_ Actin 12

cgaatgctaaccgtgagaagatgacccaattatgtttgagacctcaatactcctgctatgtacgtagc
 cattcaggccgtgcttctacttatgccagtggtcgtacaaccggattgtttggactctggagatggt
 gtnagccacac

>Tca_ Actin 2 7

ggagctgagagattccgtgtccagaagtcctctccagccatctctcattgggatggaagctgctggaa
 tccatgagactacctacaactctatcatgaagtgtgacgtggatatcaggaaagatctctatgtaacat
 tgtctcagtggtggtcaactatgtccctggtattgcagaccgaatgagcaaggagatcactgccctt
 gctccaagcagcatgaagattaaggtcgttcaccaccagagagaaagtacagtgctggtcggaggat

>Tca_ Actin 58

ttgctctgactatgatcaggaactgagactgctaagagcagctcatctgttgagaagaactatgagtt
 gcctgatggacaagtcattactattggagctgagagattccgtgtccagaagtcctctccagccatc
 ctactgggatggaagctgctggaatccatgagactacctacaactctatcatgaagtgtgacgtggata
 tcaggaaagatctctatgtaacattgtgctcagtggtggtcaactatgtccctggtattgcagaccg
 aatgagcaaggagatcactgccctgctcccagcagcatgaagatca

>Tca_ Dtdp-Glucose 4,6-Dehydratase

cacgttatccattggtgtccgcagttgctacgatgagggaaagcgtactgctgaaacgttgaccatgga
 ctatcacagagggcgccggcgttgaggttaggattgccagaattttcaatacttacggacctgatatg
 atcgatgatggtcgtgctgtagcaatttcgtgctcaggcttaaggaaggagcctttgactgtatatg
 gtgatgggaaccagacaagga

>Tca_ End13 Protein

aagtggatgggtgaaagtgaaagctggttcaaaccttttcagatggctcgcgatagtctctcca
 tcatattcattgatgaaatagattccttgtgtggccagcgtggagagggcaatgagagtgaagcatcaag
 acgatcaaaaccgaacttctgtgcagatgcagggtgttgacataatgatcagaaagcctagtctt
 gcagcaaccaatactcctatggtcttgatcaggc

>Tca_ T-Complex Protein 1, Epsilon Subunit

tggctgttgagcatttgagcatatagctcagaagtttgattttggaccaacaaatagagcctttggt
 tcaaaactgcatgaccactctgtctcgaaaattgtgaatcgatcaagcgcctttggctgagatttct
 gttaaagcagttcttctgttgcgtgatctagagaggaaagatgtcaatttagatttgattaaggtgagg
 gaaaagttggggggaactgggaagatactgagctaataatcggaatcactgttgacaaggacatgagtca
 tccacagatgcaaagcacattgaagatgc

>Tca_ Adenosine Kinase

gctgctgcaanaacaaggtgttctcaatgaatcttctgctccatttattgtgagttctcaaggacg
 cacaggagaaagttttaccgtatatggactttgtctttggcaatgagacagaggcaagaaccttcaaa
 ggttcattggctgggagactgatgatgttgcggagatagcctgaagatttctcagtgccaaaagcatct
 ggaacattcaagaggattactgttattaccaggggtgcagatcctgttgttctgctgaggatgggaaag
 tgaagcaattcctgtaataattgttaccgaagaaaacttgttgataccaatggtgcaggtgatgcttt
 tgt

>Tca_ Ubiquitin-Activating Enzyme

tggaacattggtcaggctaaatcaactgttcagcttctgctgctgcatctataaatcctcagctcaaga
 tcgaagccttgcaaaatcgtgtgggtcctgaaactgagaatgtttaatgacaccttctgggagaacct
 aacagtggtcattaatgcattagataatgtcaatgctaggctgtatgttgcagaggtgcttgatttc
 cagaaaccacttctgaatcaggaactcttggctgctaaatgcaacaccagatggtgattcctcatctaa
 ctgagaactaaggtgcctcaagtacc

>Tca_(S)-2-Hydroxy-Acid Oxidase, Peroxisomal

ctattccagcgttgacnaggtgcttcaacaggacctggccttcgcttttccaactctatgtatgcaa
 gcacaggaatgtggttcncagctagtgagaagagctgaaaggctggttccaggcaattgcccttact
 gtggatactccaagacttggcggagggaagctgatatcncnnaatantatttactctgnnaccnattct
 tgantttgannaantttganggnacggaccnccggcncaantgnataannaccngatgactccggancta
 nattcanangttgcangngacannnnga

>Tca_ Betaine-Aldehyde Dehydrogenase Precursor

ctgaaggtttgatgcaagcaaaaggctcctgtctccctccaatgggtacatttaagagttatgtgct
taaggaaccaattgggggttggattgattactccatggaattatccactattgatggctgatggaaa
gtagctcctgccttggtgcaggntgtgctgcaatactgaagccttctgagtnggcatctcacctgtt
tggagctggctgaantgtgtagggatgttggcttccctcctggtgncctgnacattcctgactgggntg
gggcctgaantggnccccttgcatnacatcctcatgttgacaanaa

>Tca_ Isocitrate Dehydrogenase

agagattcttagcccaagggttgatctcttgattgatgacatctgtgctggtggtgccagatcggaa
gacaatagaagcagaggctgctcatggtacggttacgcgacattacagagttcacaaaagggtggtaa
accagcacaacagcatagcatccatcttggctggacacgtgggctgacacagggcaaaattggatg
acaatatcagactcttgagttcactgagaaactgagggcgtctgtattggagctgtggaatctgggaa
gatgaccaaatctagcact

>Tca_ Pyruvate Dehydrogenase E1 Component Alpha Subunit

gattgttggtgctcaggtccgttaggatgtggcttggcgtttgcgcaaaagtactctaaagacgagaat
gtgaccttgccttgtatgggtgacggggctgctaataagggcagttgttgaggcttgaatatctctg
cgcttgggatctgctgcaatgttggctgagagaataactatggtatggggacggcagagtggag
agccgcaaagagtcca

>Tca_ Turgor-Responsive Protein 26g

gagatgtggaatcctcttggaaatagtggtgtaatacggcatttaactcccatgtgctgttcttggat
ggaacgctgcatcgcactggtgtgtggcaactgcattgttggaaaggtgctccaaccactccttggat
caccattgcaatgacaaagctagtggtggagtttggaaaagaataactgccaggtgcaatctcact
tcatttgggtggtgctgaaattggcgaagcaatagc

>Tca_ Putative Udp-Glucose

ctatggctgatcctgactagaagctattagacaaagaaggatgcaagagctaatggctcaacatggcgcg
gggagtaacagaaccctgatcaacagaaagctcaggaagacgccaagagggagcagatgaacgaagcc
aatgatgctcagtcagatttgcctctgaagctcgtgaaagagttgctgaattgccttggatgaacc
tgagaacgctagaggtgtgg

>Tca_ Esterase D

aagtggaagaactggcgtatgtacgattatgttgcacaagaaatgcaaaactcctgagtgaatctcc
ctcaactgatacatcaaaagcatctatttctggtcattctatgggtgggcatggtgctctgacagtcta
tctgaaaaacctgataagtataagtcagtgtctgctttgacaccattgccaatctataaattgccc
tggggtcagaaggcttccaaaactatcttgtaacgacaagctgcttgggaggaataagatgcc

>Tca_ Chalcone Synthase 1

tnatggcttccctggatcattccaatcctcatttgagcaagagctgatagtctggaattanat
gatgattccctcagcaactccattcattctaaatcagatgnagaacaacctgaagacattctg
ctaagctcacaactaataatctcgggagaacagattcactaccatcaaggttgcctcgaa

>Tca_ Putative Beta-Ketoacyl-Coa Synthase

gttacaactggggaggaatgggctgtagtcaggtttaatttcaatagatcttgctaaaaatcttctca
 agtccatccaactcctatgcattggttatcagcatggagaacatcaccttgaactggactttgggaat
 gatcgctcaaaactggttcaactgcttgttaggatgggagggcgtgcaatattgctctcgaacaaac
 gctctgacagaagaagatccaaatgaattggttcacactgtgcgactcacaaggggtgctgatgataa
 gtgcttttctgtgttaccaggaagaagactctcgggggaagggttggtttacattgcaaaaggatctg
 atggcagttgccggtgatgctctaaa

>Tca_ Amidophosphoribosyltransferase Precursor

gtgttcacacaatggtaacttagtaaattatagactttgagagctatgcttgaagataatggttcaat
 ttttaactagttctgatactgaggttgcttcatttgattgctatttcaaaggctaggcctttcttt
 ttgaggattgtgatgcttgtgagaagctgaagggccttactccatggtgtttgcgactgaggataagc
 ttgttcagtagctgaccctatgggttaggcctttgtaatgaggaggagtaatggggctgt

>Tca_ Ethylene-Inducible Protein Hever

ttgacgcgcagatcctgaagccatcggaatcgattatgttgacgagagcgaagtgcttaccagccga
 cgaagagaaccacattaacaacataatttcggattccttctgttgcgggtgtcgaattgggggaa
 gcgcttcggaggatccgggaaggagctgctatgattcgaacaaaggagaagctgggaccggcaacatta
 tcgaagccgttaggcacgtgaggtccgtgatgggagacattagggtgcttagaaacatggatgatgatga
 ggtttc

>Tca_ Cellulose Synthase Catalytic Subunit

tatgaagatgtggcacagggtatagaatgcacaacagaggatggaatcagttattgtgaccaagc
 gagatgcttccgtggaactgctcaatcaatcactgatgggcttcaccaagtcctgcgttgggctac
 tggttctgtgagattttttcccgtacaatgccttccttctagcccaagaatgaagtgtgacg
 aggattgcgatcttaattgttgaatctatcccctcacctccgttttctgattgtactgtttcctcc
 ctgcactttcctttctctggccagttcattgttcagaccctcaatgtcactttccttactacctct
 caccattacagtactctttgttctagctgtgcttgaattaagtgcttggcattgagtagaagaa
 tgggtgtgaaacgagcagt

>Tca_ Udp-Glucose 4-Epimerase

cccaagaactcatgccttacatacagcaagtggctgttggcagattgcctgaacttaatgtatggtc
 atgattatccaactaaagatggcagtgcgattcgagactacattcatggtatggacttagcagatggta
 tattgtgcacttaggaagcttttactacacagaataggttgattgcctataaccttggaaactggc
 tgtgtacatctgtgcttgaatgggtgctgcttttggaggagcttctggaagaaaatccctattaac
 tatgccaagaagaccaggagatgctagagccgtt

>Tca_ Alcohol Dehydrogenase Class III

tgtgtgtaaggggtaactgaagtcaaccaggtgaccatgcatccctgttaccaggcagagtgtaga
 gaatgcaagtctgcaaatcaggaaagacaaacctatcgcgcaaaagttagggcagccactggagctggag
 tcatgatgaatgatcgcaagagtcgcttctcaataatggaaaaccaatataccacttcatgggcacctc
 gacattcagccagttcacgggtgtacatgatgttagtggtaagattgatccacaagctcctttggac
 aaagtaagccttcttgggtgtggttc

>Tca_ Preprotein Translocase Secy Subunit

gaaggtgaagctgggagaagaagattcttcagtatactagatatgcttcagttggtttgctatagtac
 aggcaattggccaagtactataccttcgtccatagtcaatgactttagcacgcagtgggttctctcctc
 tgtcactatattgacgcttgggtgctgtagttacgacatatattggggaagaattctgatctaaaactt
 ggaaatggcacatctcttaataattacaacattttatcctacttggcagcatcttttgtaggacag
 ttgtgcaggcatatcaggatggtaactacattggacttgtaccatcattatctcattctctctgctggt
 tcttggcattgtctatgttcaggaagcagagaggaa

>Tca_40s Ribosomal Protein S3

gttctcggtgagaaggggagaaggattagagagctgacttctgttgttcagaaacggttcaagttccag
 agaacagtgtagaactctatgctgagaaggtaacaacaggggtcttggccattgctcaggcagagtc
 cctgcgttacaagctcctggaggccttgcgttcggagggttgcctatggtgttctcagatttatcatg
 gagagtggggccaagggatgtgaggtcatggtgagtggaaanctccg

>Tca_ CAB11

gaagctgcgcgtattgcttgaagatggggcttgggtcgcacagcctgggttctttatgtcttccacct
 tctgatctccacatagtggaacaaatgaactcaataacgaagagggtgatgaggaagcaaagtattct
 gcttccagcatcaccattgaggagcattgattatccaatcttagtgaaaactcaggcaagagca
 tcctgcaacaccaacatagcccaacgcgagttcacaagctcagcttgcacataccagttcaagttccc
 tggtcg

>Tca_ Elongation Factor 2

tcctggtgagaagaaggattgtatgtaagagtgtacagagaactgttattggatgggaaagaggca
 ggaaactgtggaggatgtgccttggtaacacagttgctatggttggtttgatcagttattaccaag
 aatgctactttgactaatgagaaggaagttgatgcccaccaattcgtccatgaagttctctgtctc
 ctgttctcgtgttgcgttcag

>Tca_ Hypothetical 47_9 Kda Protein (Elongation Factor, by BLAST)

ggaaggtcagacgcaaccttgcaactgaacagcaacacgaacaacaggagagacagagaacttcatggcac
 gaattgggtgggcatcaactccttctcattagtcaaagtagcattcttggtaataaactgatccaaacc
 aaccatagcaactgtgtaccacaaggcacatctccacagttcctccttccatccaataaca
 gttctctgtacactcttaacatacaaatcattcttctcaccaggga

> Tca_Leafy

cagcttcanaattaccctcantaaggactagtctcaggtttaaacgaattcgccttagcttggggaaacatac
 caaatagcaagacggcgatgagcattgaaaattgcatcaatgtcccaacctggcgcgccgctatggccacaagag
 gctttagcaggcctgtctccaagccccaacattctcgcctctctcttgaagctctcctcagcgcattgatgcctct
 cgtcaaggcagtgaggcatagcaatgcacgtagtgcacatctgggttggtaatgtagcttccccagccttctc
 gcgtacctgaagggcgaattcgcggccgctaaattcaattcgcctatagtgagtcgtattacaattcatgccgnn
 ncnntttta

Appendix C

Phylogenetic trees created by Mega 2 to compare genes in *Arabidopsis*, cotton, *Populus* and *T. cacao*

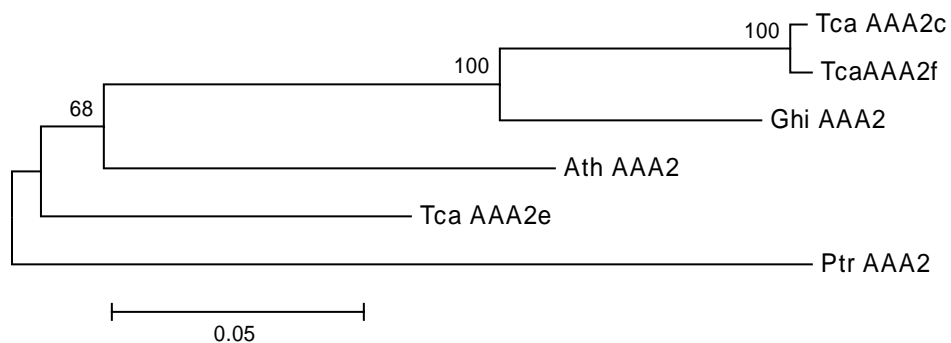
Key:

Tca = *Theobroma cacao*
Ghi = *Gossypium hirsutum*
Ath = *Arabidopsis thaliana*
Ptr = *Populus trichocarpa*
Os = *Oryza sativa*

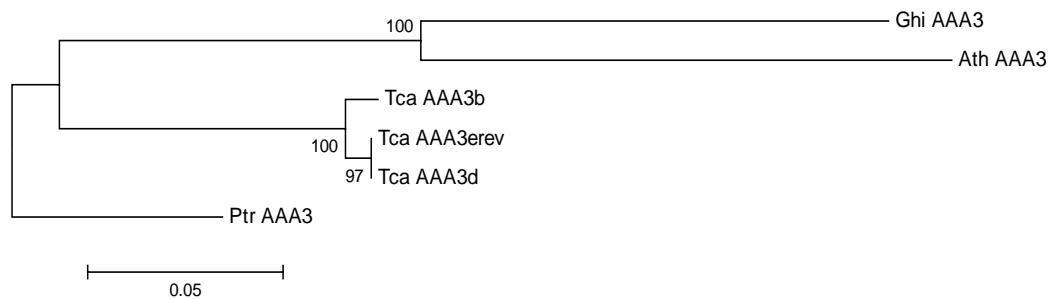
- Each tree has undergone 100 bootstrap replications with replacement and are indicated at their respective branches.
- Scale at the bottom of each tree represents p-distance for Tca, Ghi and Tca comparisons, while gene family trees have a scale in Kumar 2 Parameters
- Tca genes are recorded by their plasmid name. This is derived by their respective functional category (please see Table 3-1 for functional categories).
- All Ath and Ghi sequences were used to design the *T. cacao* heterologous primers.

i. Phylogenetic trees of *T. cacao*, cotton, *Arabidopsis* and *Populus*

Glutamate Decarboxylase



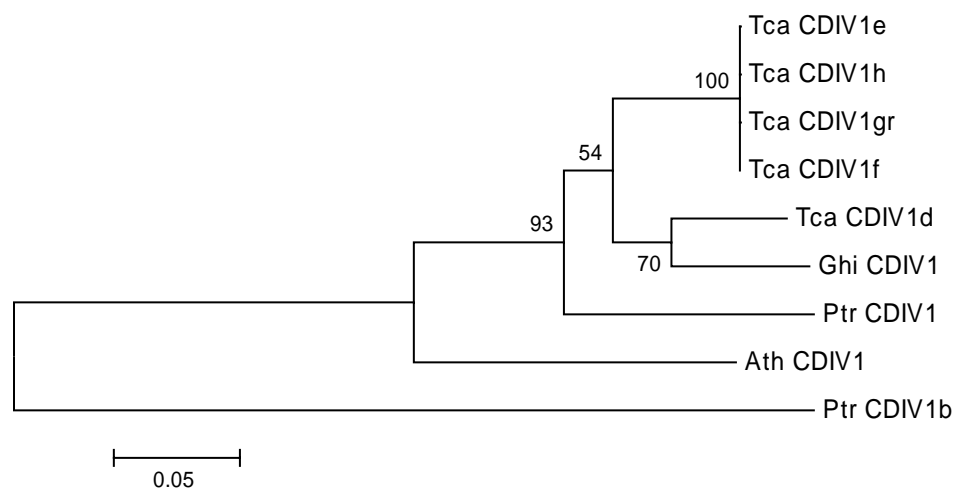
Glutamate dehydrogenase



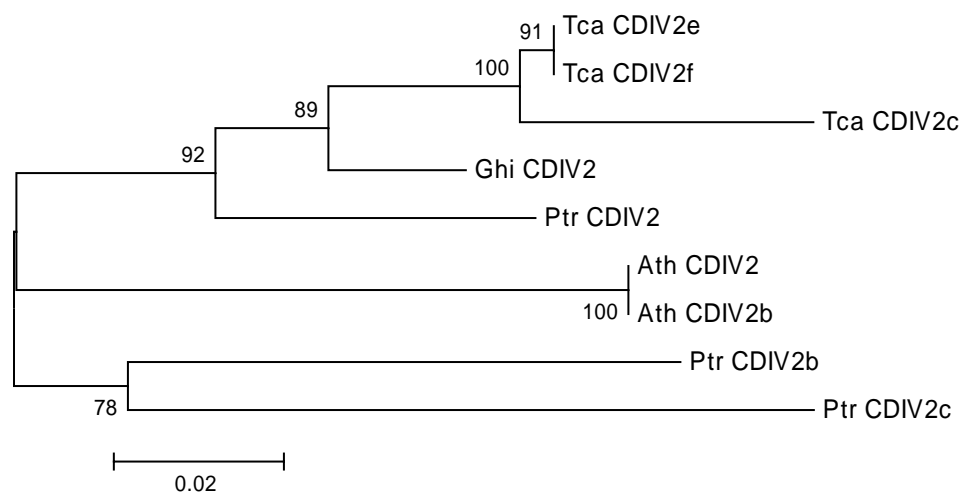
Ketol-Acid Reductoisomerase Precursor



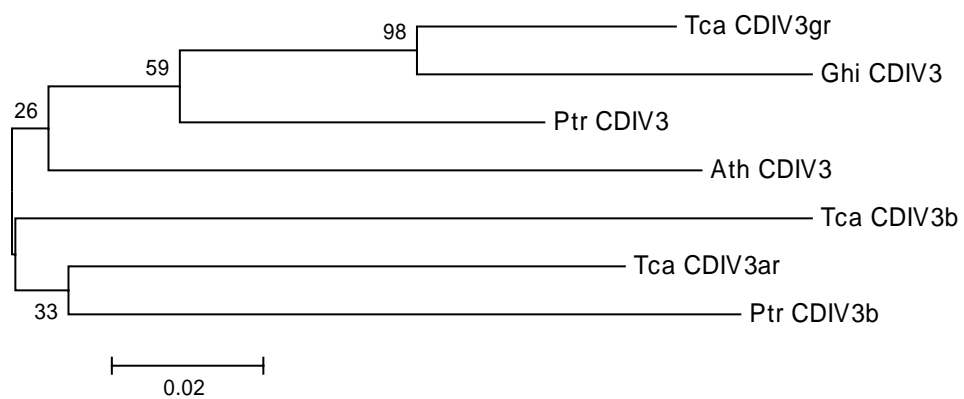
Actin 12



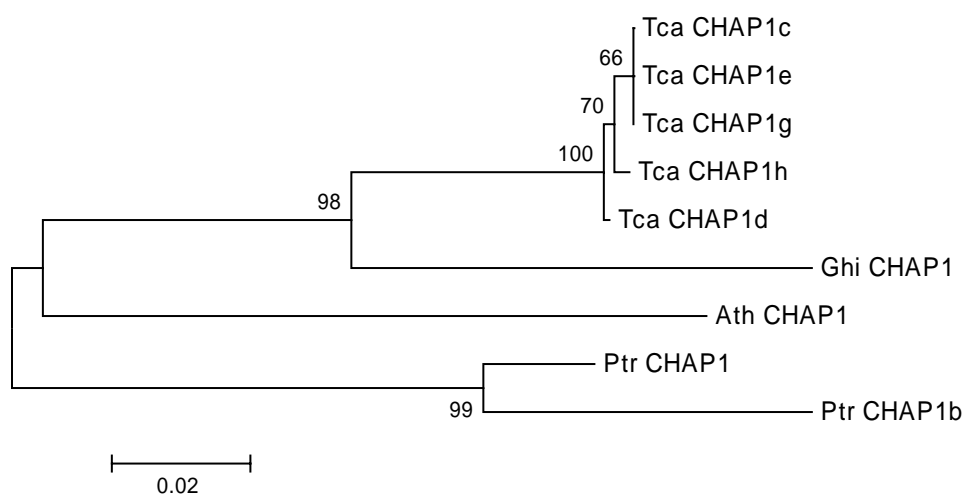
Actin 27



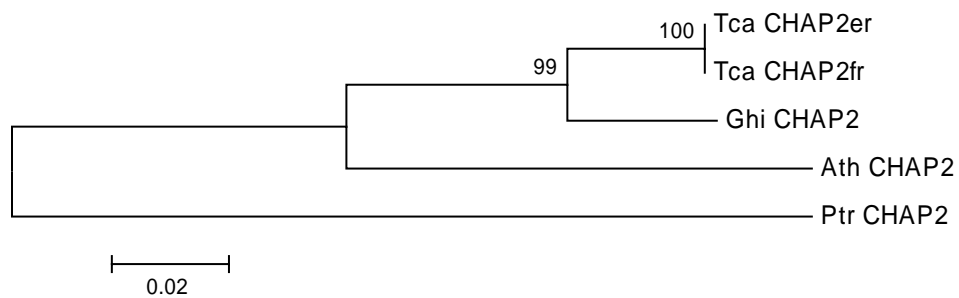
Actin 58



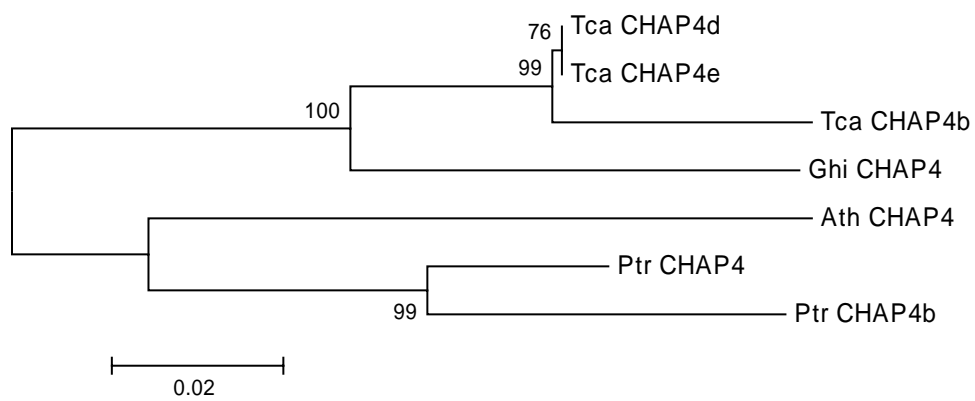
Dtdp-Glucose 4,6-Dehydratase



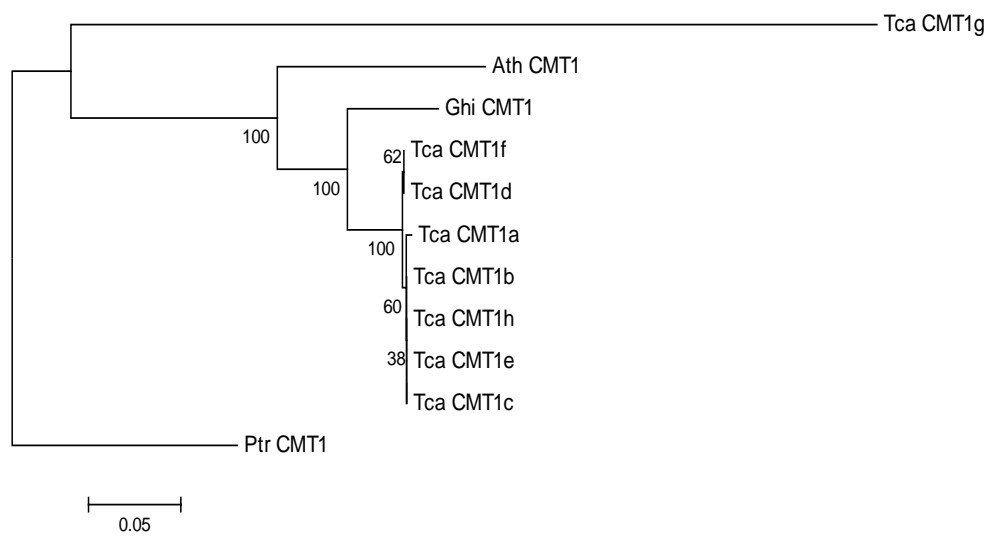
End13 Protein



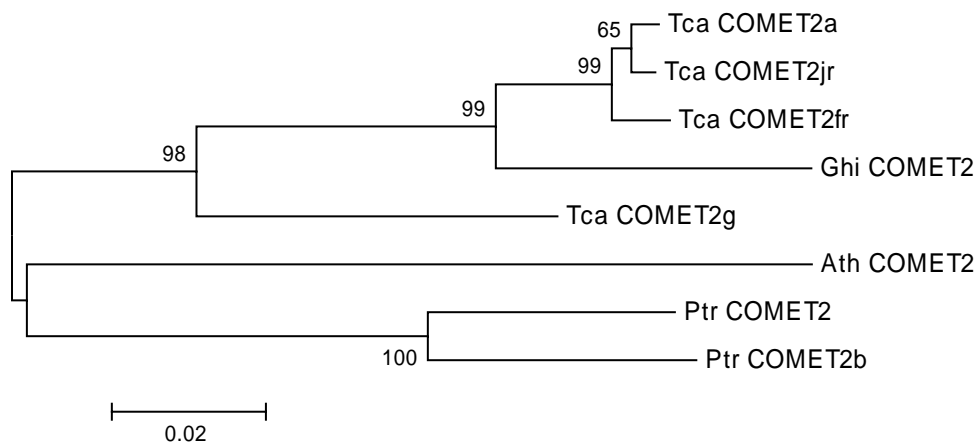
T-Complex Protein 1, Epsilon Subunit



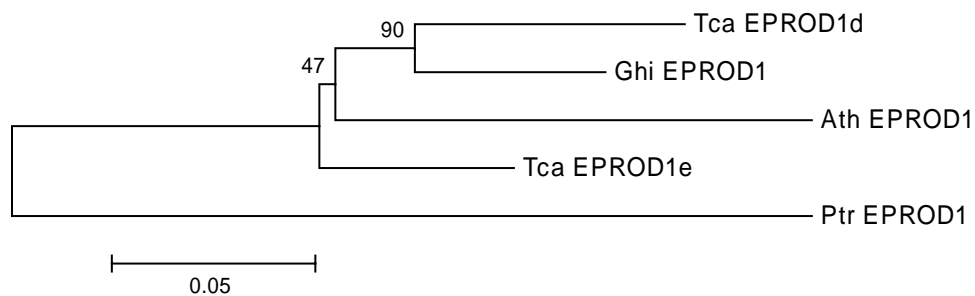
Adenosine Kinase



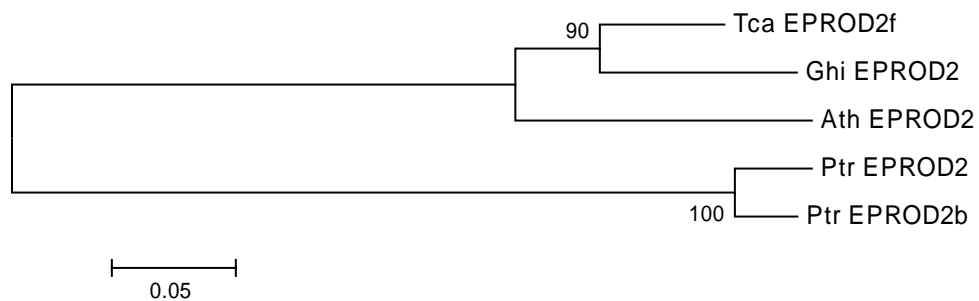
Ubiquitin-Activating Enzyme



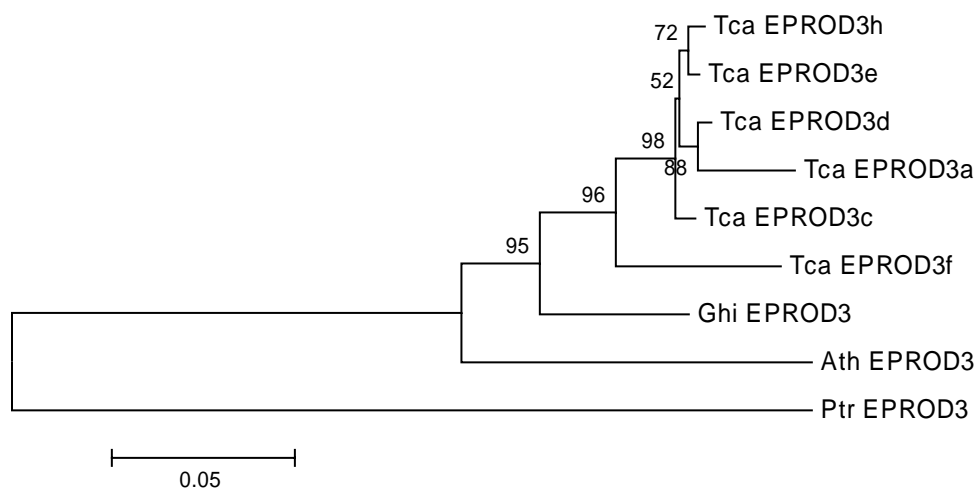
(S)-2-Hydroxy-Acid Oxidase, Peroxisomal



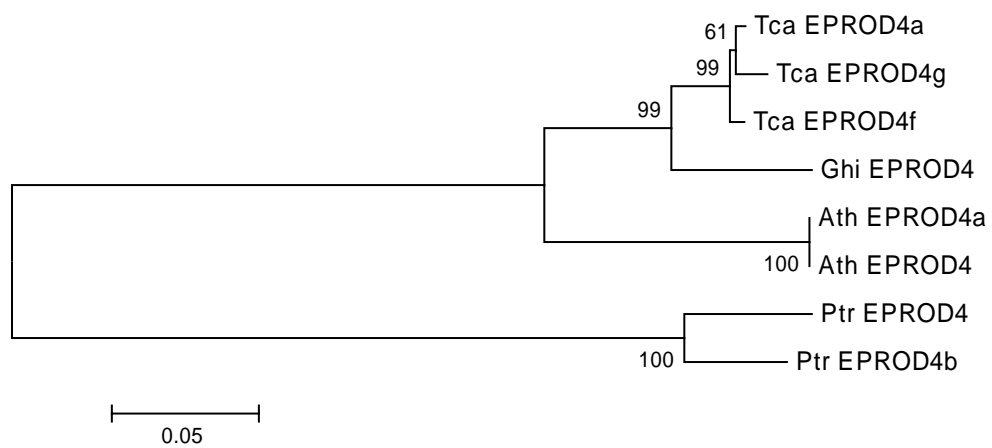
Betaine-Aldehyde Dehydrogenase Precursor



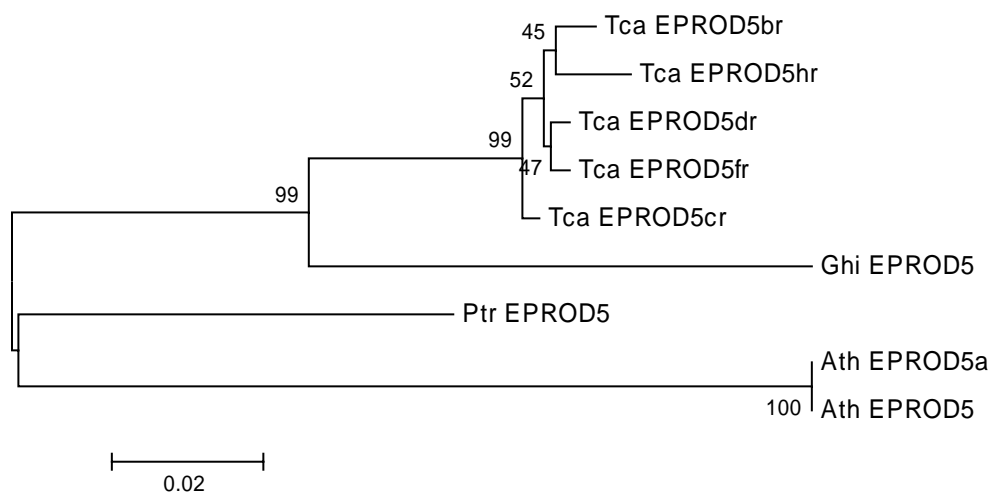
Isocitrate Dehydrogenase



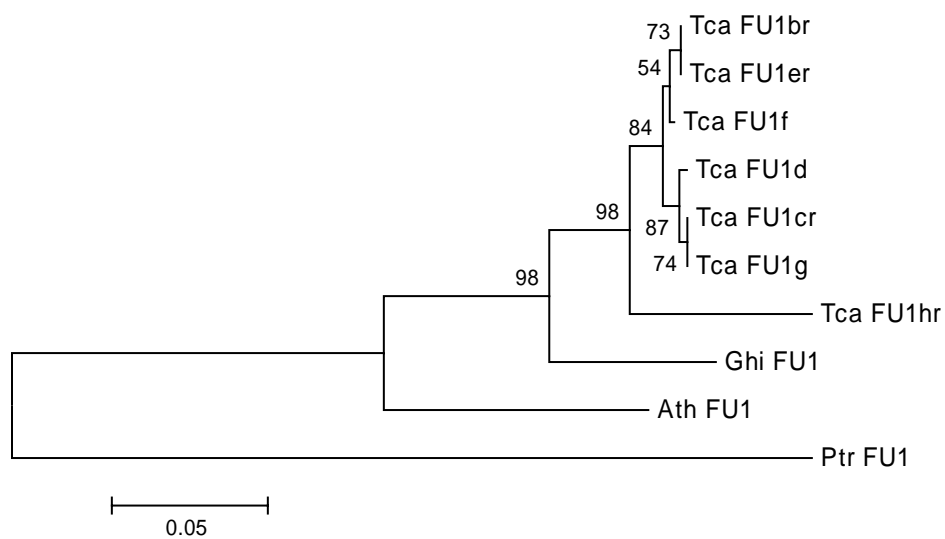
Pyruvate Dehydrogenase E1 Component Alpha Subunit



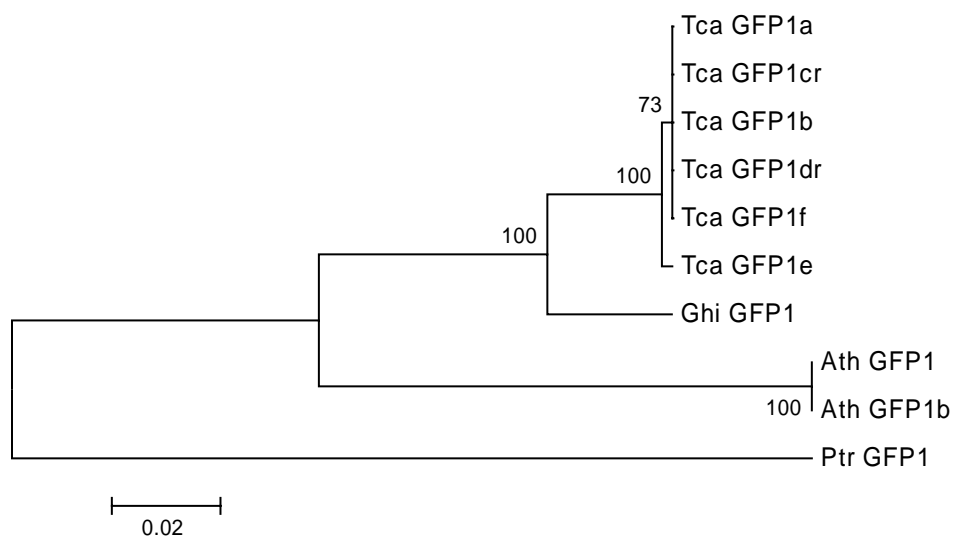
Turgor-Responsive Protein 26g



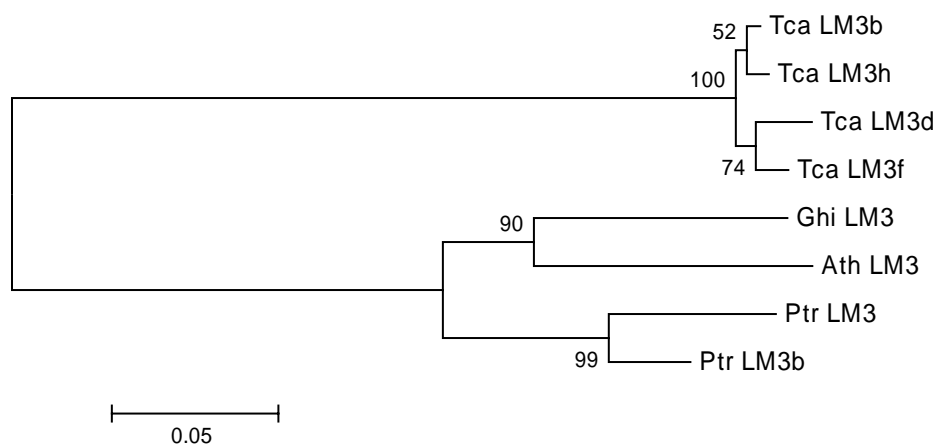
Putative Udp-Glucose



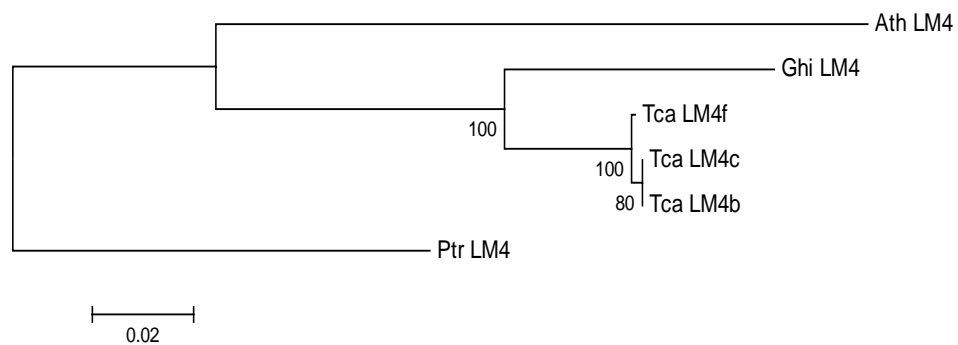
Esterase D



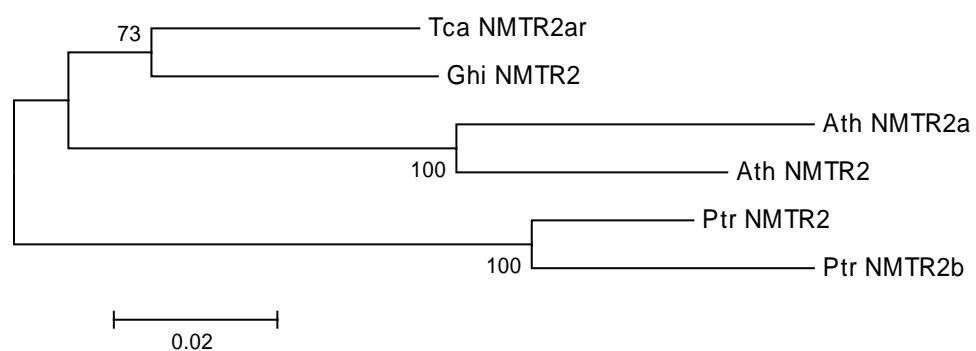
Chalcone Synthase 1



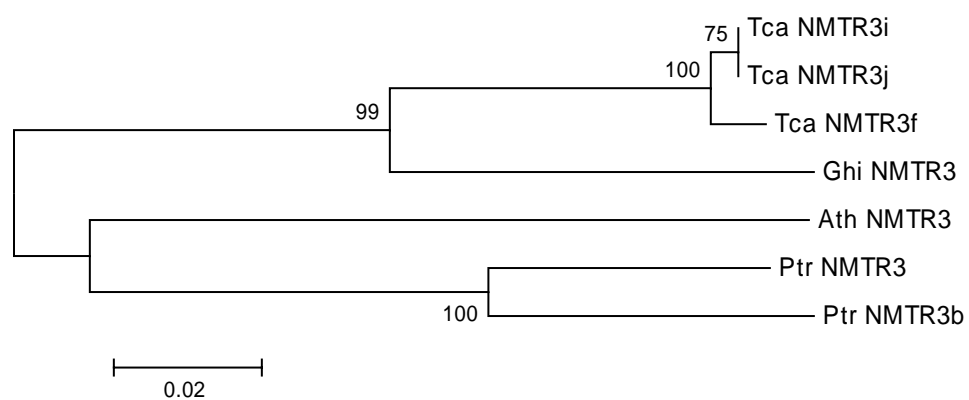
Putative Beta-Ketoacyl-Coa Synthase



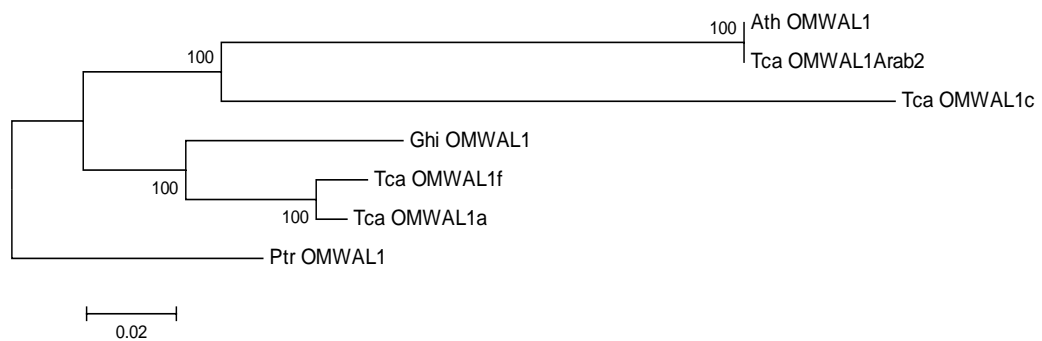
Amidophosphoribosyltransferase Precursor



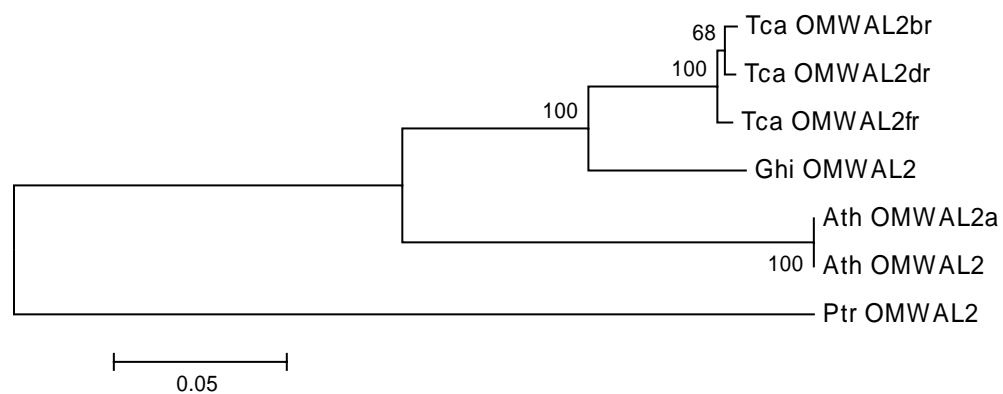
Ethylene-Inducible Protein Hever



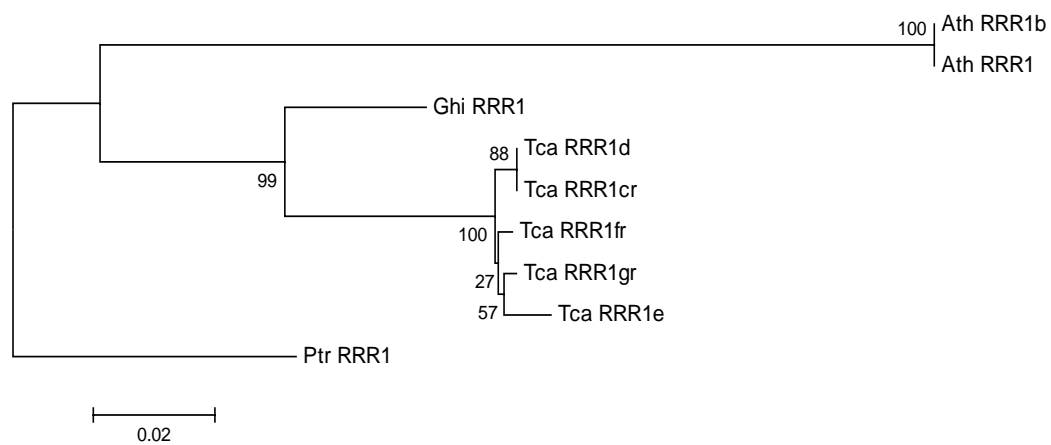
Cellulose Synthase Catalytic Subunit



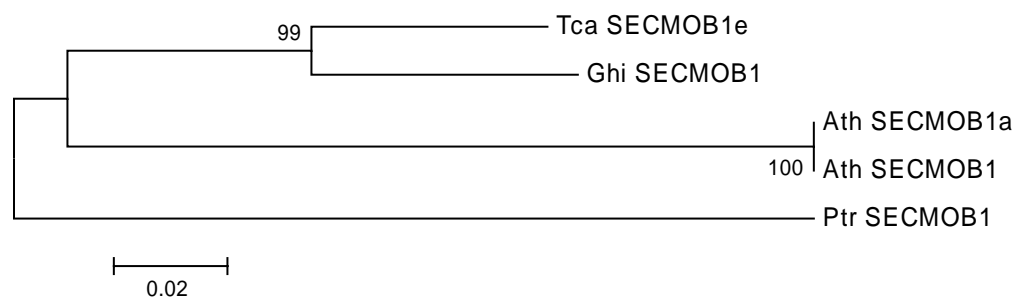
Udp-Glactose 4-Epimerase



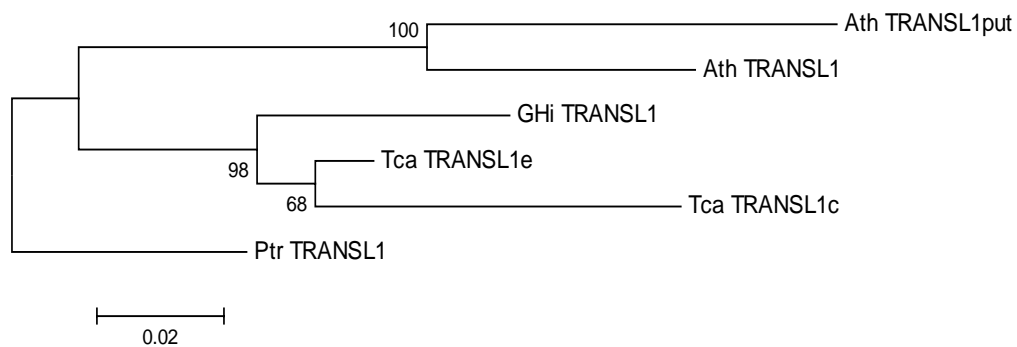
Alcohol Dehydrogenase Class III



Preprotein Translocase Secy Subunit



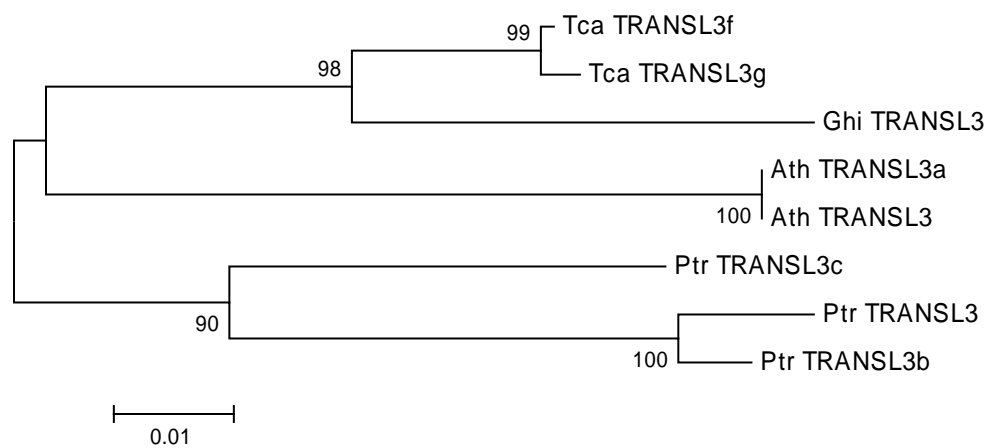
40S Ribosomal Protein S3



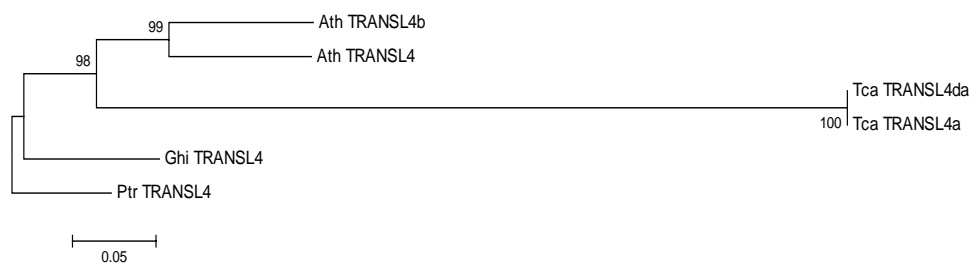
CAB11



Elongation Factor 2

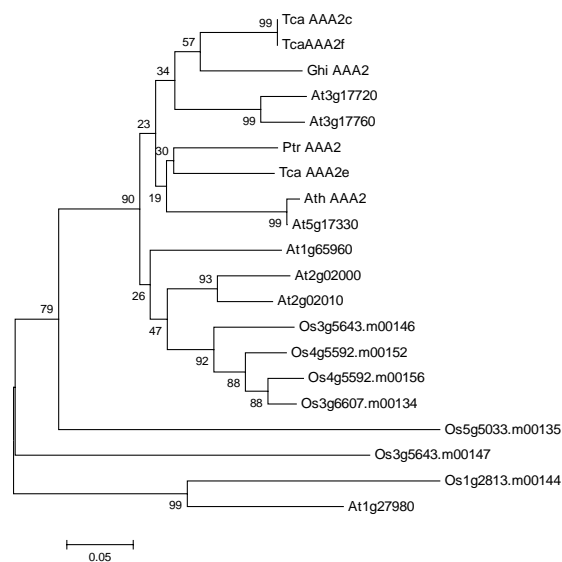


Hypothetical 47_9 Kda Protein

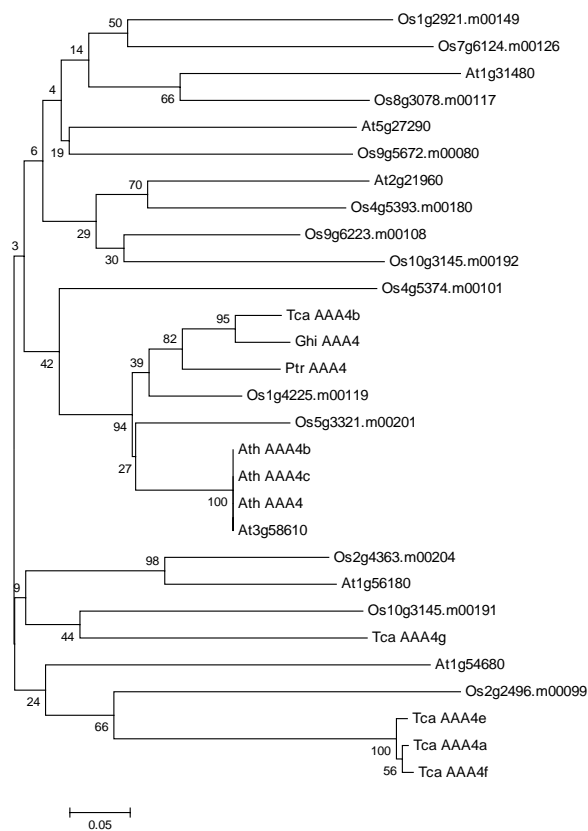


ii. Phylogenetic trees of *T. cacao*, cotton, *Populus*, and gene family members of *Arabidopsis* and *O. sativa*

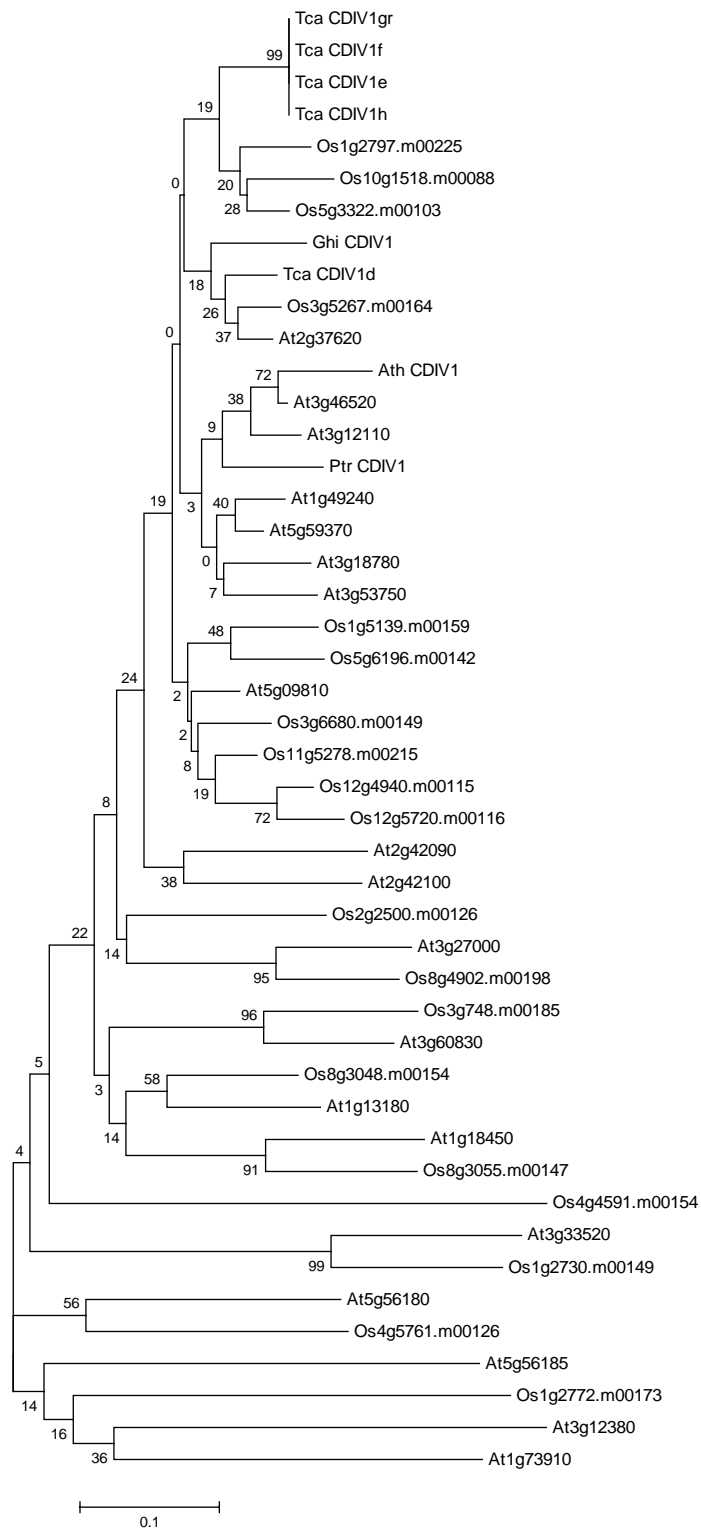
Glutamate Decarboxylase



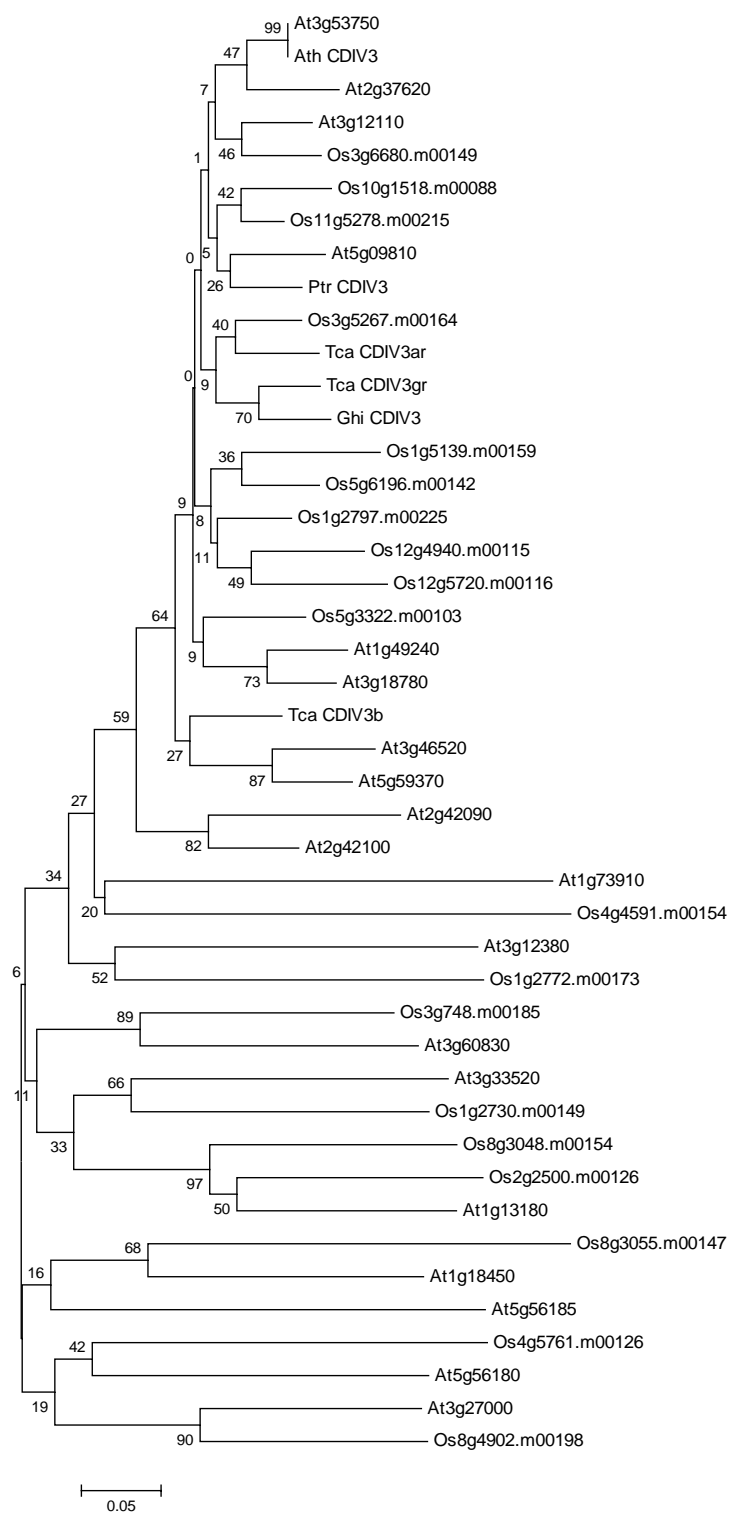
Ketol-Acid Reductoisomerase Precursor



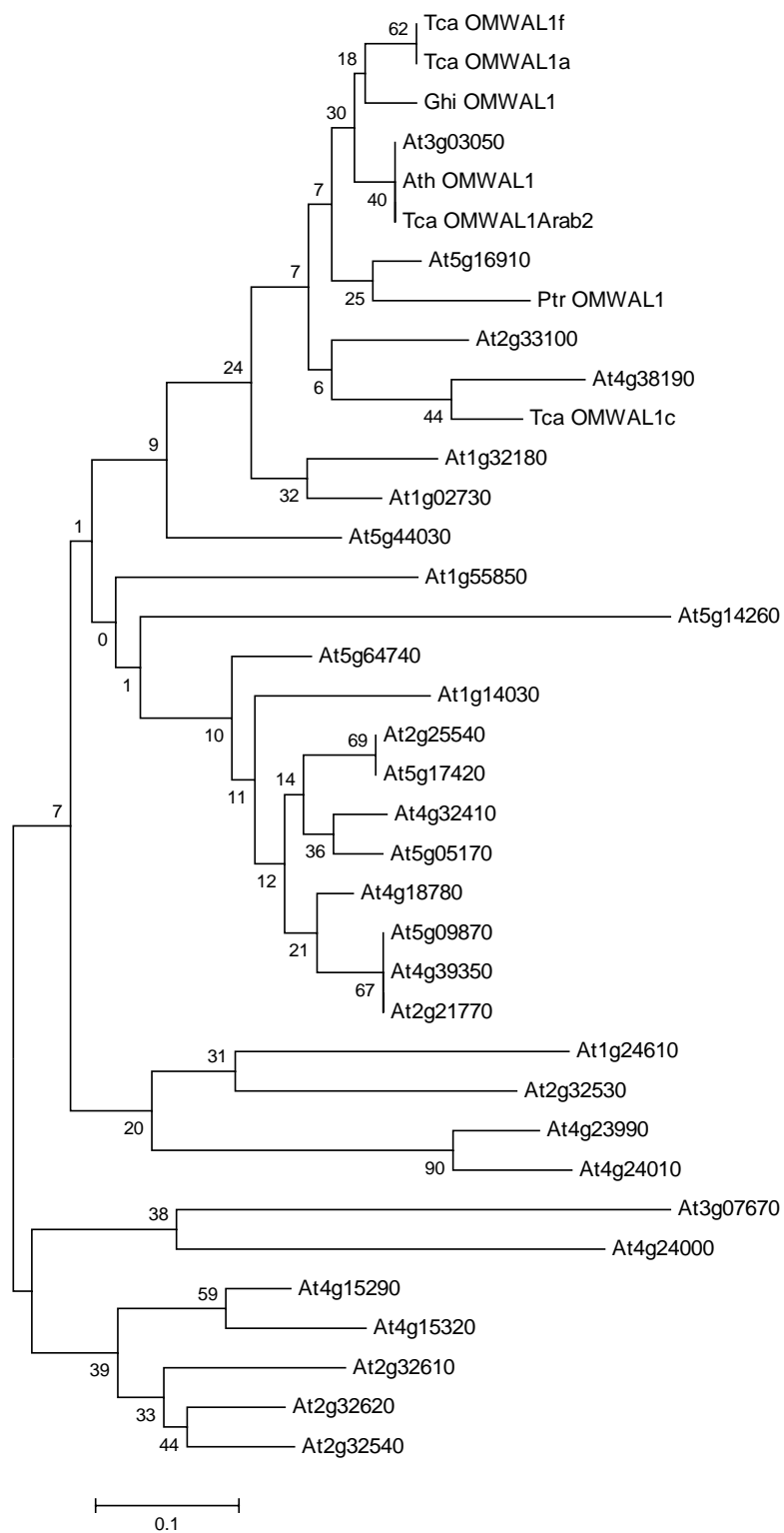
Actin 12



Actin 58

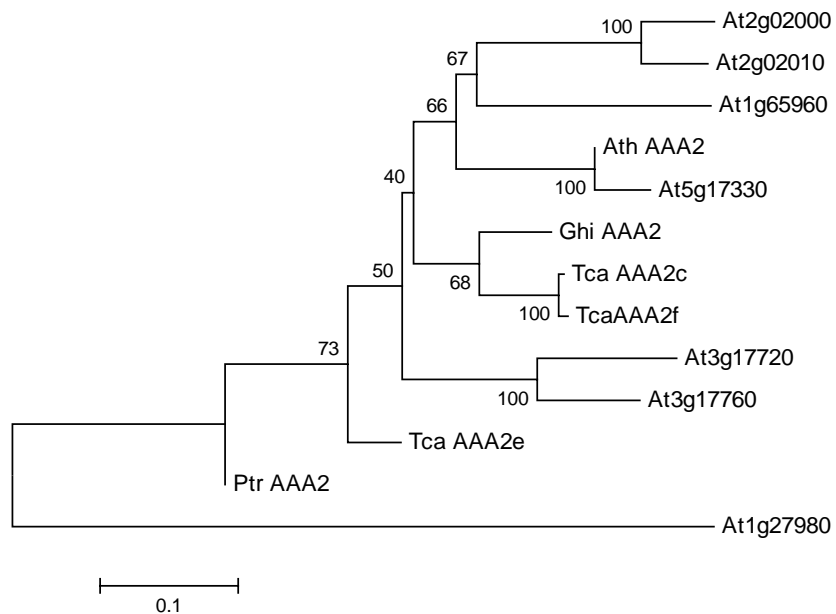


Cellulose Synthase

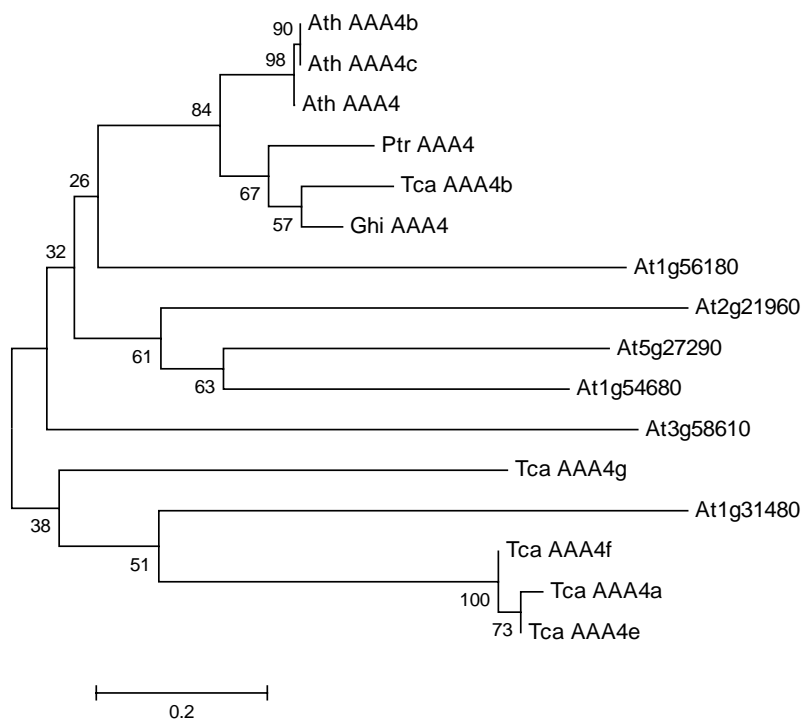


iii. Phylogenetic trees of *T. cacao*, cotton, *Populus*, and gene family members of *Arabidopsis*

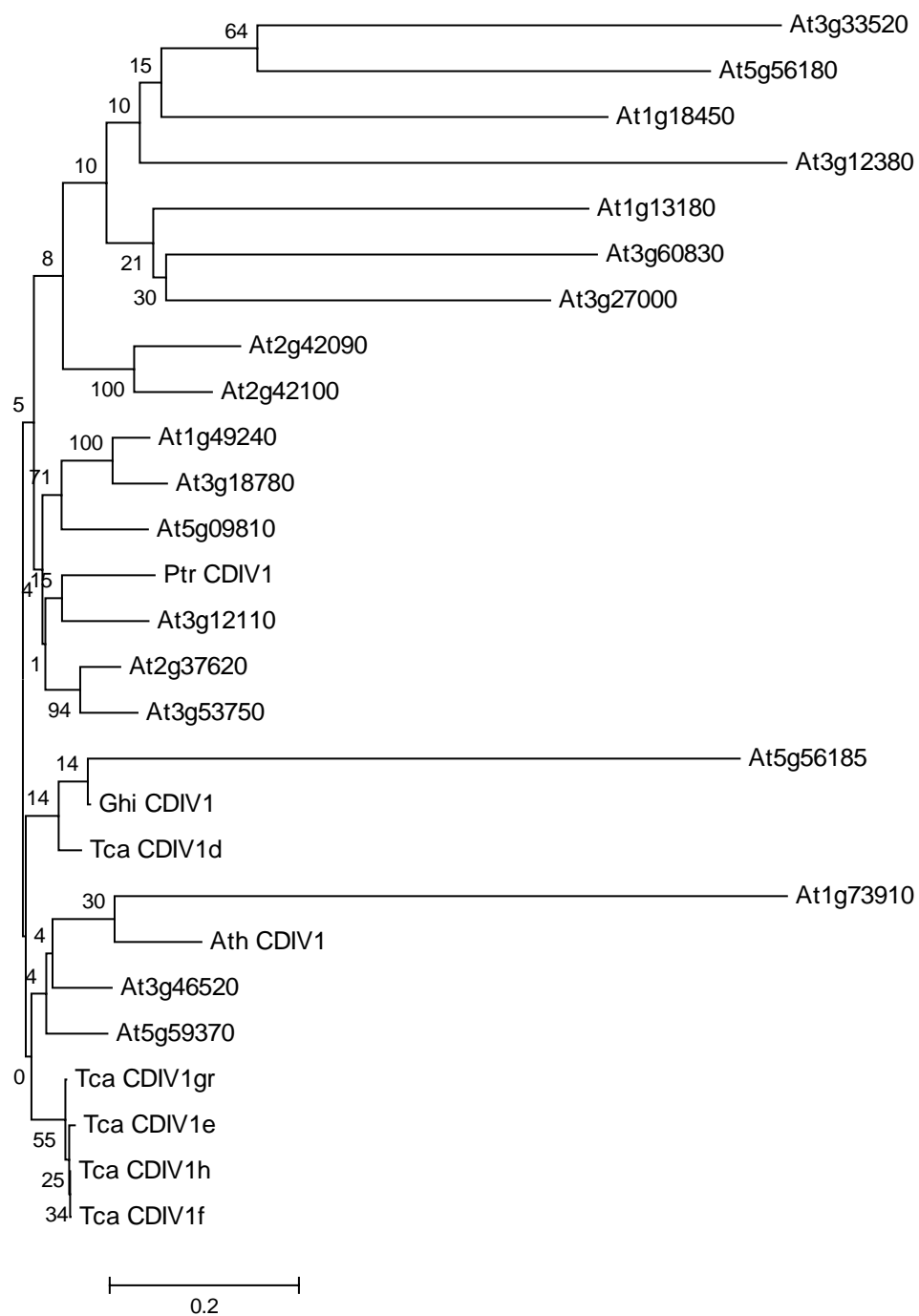
Glutamate Decarboxylase



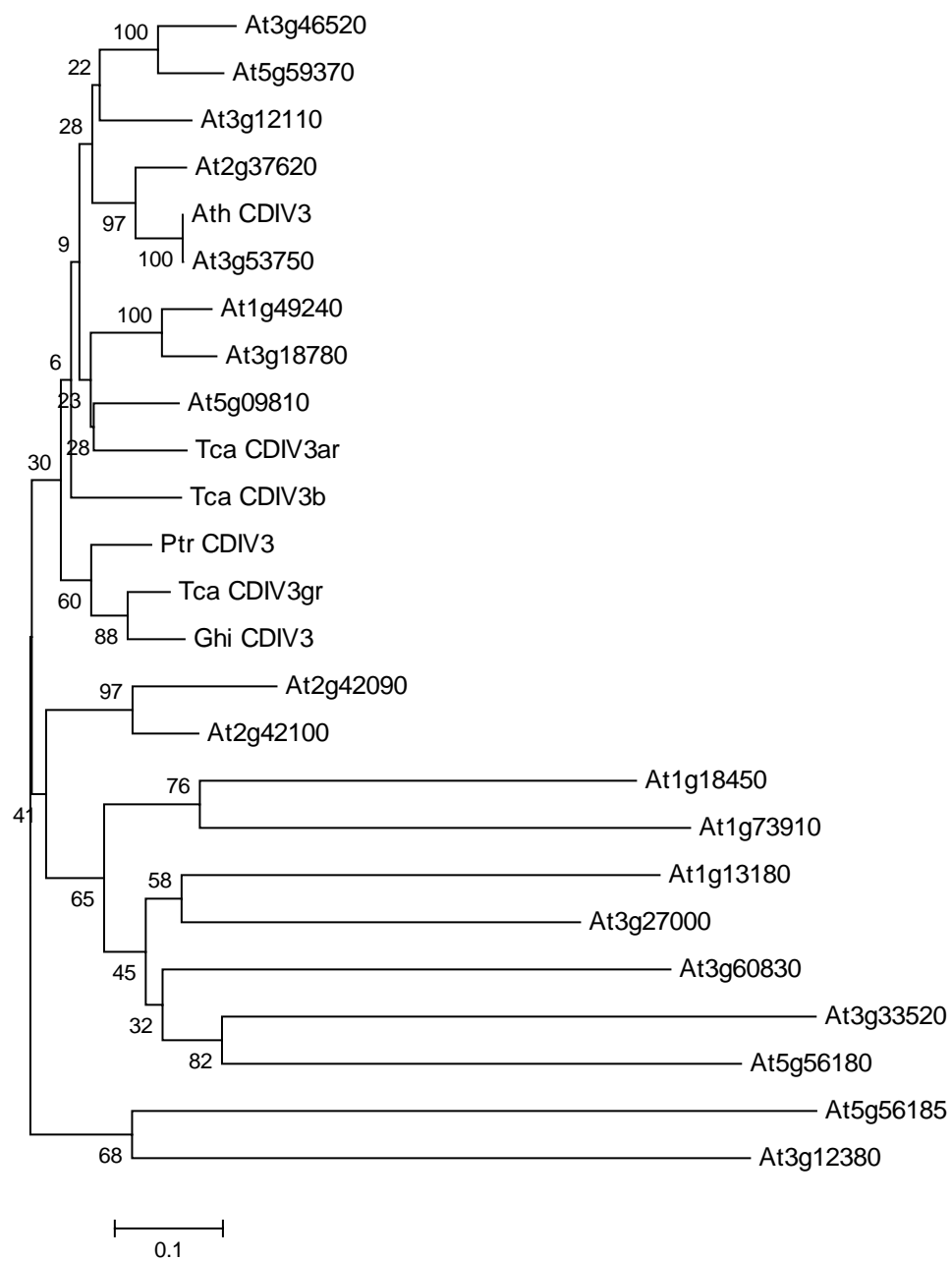
Ketol-Acid Reductoisomerase Precursor



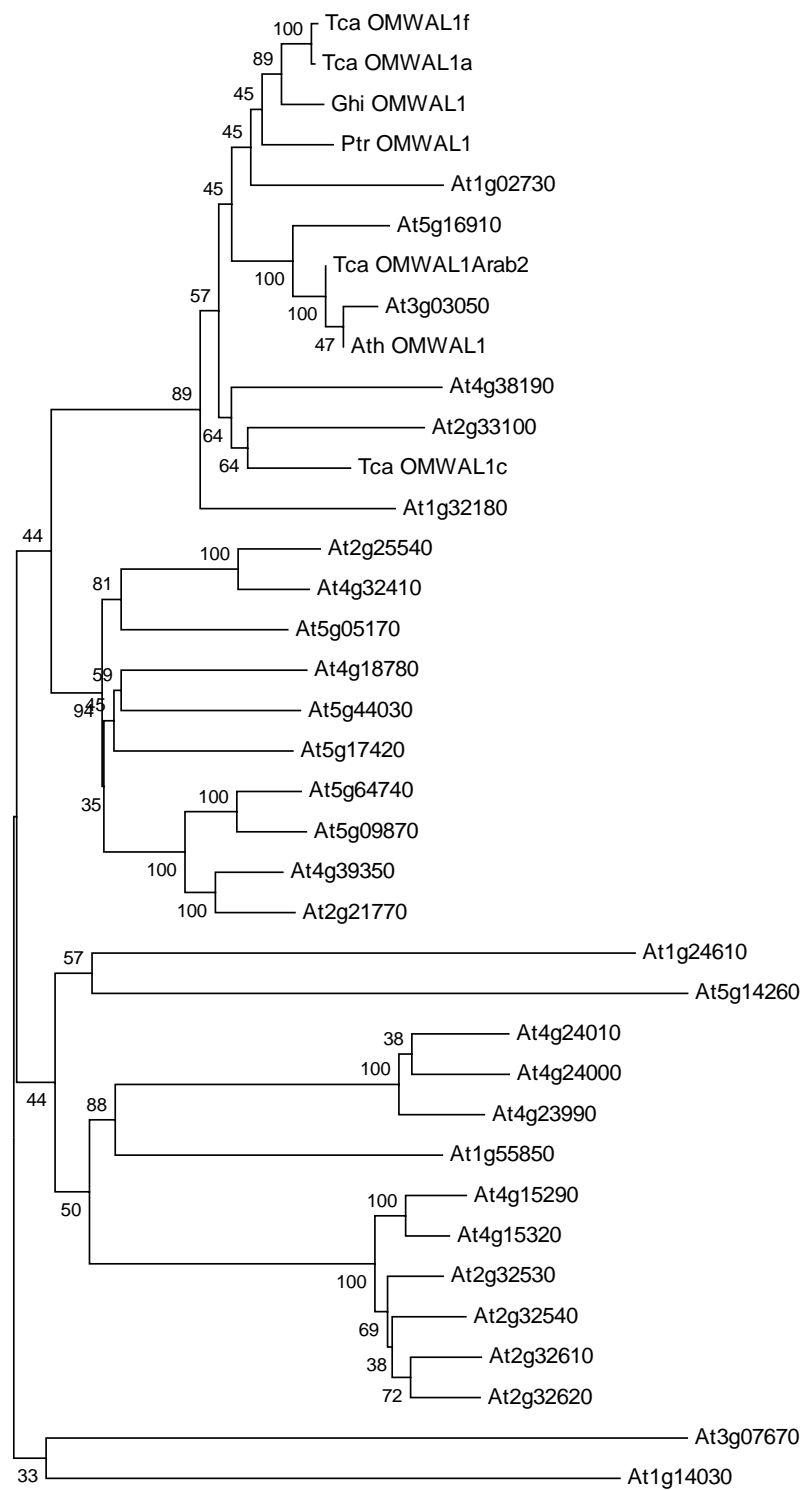
Actin 12



Actin 58



Cellulose Synthase



Appendix D

Analysis of the gene expression of *LEAFY*, *API*, *AP3*, *PI*, *AG*, and *CELLULOSE SYNTHASE* in *Arabidopsis* using the GENEVESTIGATOR database

(<https://www.genevestigator.ethz.ch>)

The order of the tables are:

LEAFY

AP1

AP3

PI

AG





























CELLULOSE SYNTHASE

Expression Relative to organ

Expression relative to stage in development

These tables were compiled in Nov 2004.

Anatomy	# of Chips	Mean	Std-Dev	Std-Err	247490_at/At5g61850 (linear)	
					0	1000
0 callus	6	125	41	18		
1 cell suspension	46	70	50	7		
2 seedling	341	104	58	3		
21 cotyledons	3	56	38	27		
22 hypocotyl	3	86	19	13		
23 radicle	3	94	19	13		
3 inflorescence	204	256	416	29		
31 flower	74	192	142	17		
311 carpel	6	64	28	12		
312 petal	6	98	14	6		
313 sepal	6	77	14	6		
314 stamen	14	299	185	51		
315 pedicel	3	125	37	26		
32 silique	17	87	34	9		
33 seed	54	214	230	32		
331 embryo	18	366	339	82		
34 stem	19	140	92	22		
35 node	3	46	27	19		
36 shoot apex	26	773	926	185		
37 cauline leaf	3	48	28	20		
4 rosette	494	128	77	3		
41 juvenile leaf	69	142	62	7		
42 adult leaf	237	117	67	4		
43 petiole	30	153	63	12		
44 senescent leaf	3	76	13	9		
5 roots	208	116	97	7		
51 primary root	0	0	0	0	n/a	
52 lateral root	6	168	87	39		
53 root hair	0	0	0	0	n/a	
54 root tip	0	0	0	0	n/a	
55 elongation zone	6	106	38	17		

Anatomy	# of Chips	Mean	Std-Dev	Std-Err	259372_at/At1g69120 (linear)	
0 callus	6	79	33	15		
1 cell suspension	46	81	80	12		
2 seedling	341	125	65	4		
21 cotyledons	3	52	49	34		
22 hypocotyl	3	80	23	16		
23 radicle	3	76	11	8		
3 inflorescence	204	2022	2742	192		
31 flower	74	4065	3051	357		
311 carpel	6	118	31	14		
312 petal	6	6912	1489	666		
313 sepal	6	6688	4077	1823		
314 stamen	14	257	140	39		
315 pedicel	3	4727	215	152		
32 silique	17	291	216	54		
33 seed	54	361	606	83		
331 embryo	18	827	876	212		
34 stem	19	1258	2118	499		
35 node	3	52	15	10		
36 shoot apex	26	1674	2706	541		
37 cauline leaf	3	44	20	14		
4 rosette	494	136	117	5		
41 juvenile leaf	69	133	69	8		
42 adult leaf	237	111	119	8		
43 petiole	30	151	89	16		
44 senescent leaf	3	43	6	4		
5 roots	208	100	104	7		
51 primary root	0	0	0	0	n/a	
52 lateral root	6	151	101	45		
53 root hair	0	0	0	0	n/a	
54 root tip	0	0	0	0	n/a	
55 elongation zone	6	105	35	16		
























Anatomy	# of Chips	Mean	Std-Dev	Std-Err	251898_at/At3g54340 (linear)					
					0	4250	8500	12750	17000	
0 callus	6	95	51	23						
1 cell suspension	46	118	66	10						
2 seedling	341	141	100	5						
21 cotyledons	3	170	18	13						
22 hypocotyl	3	169	18	12						
23 radicle	3	138	31	22						
3 inflorescence	204	3002	3751	263						
31 flower	74	5934	4586	537						
311 carpel	6	945	109	49						
312 petal	6	16007	1583	708						
313 sepal	6	1717	187	84						
314 stamen	14	5488	6021	1670						
315 pedicel	3	137	16	11						
32 silique	17	1888	838	209						
33 seed	54	1418	1580	217						
331 embryo	18	791	425	103						
34 stem	19	841	1342	316						
35 node	3	40	45	32						
36 shoot apex	26	1132	1912	382						
37 cauline leaf	3	85	16	11						
4 rosette	494	204	169	8						
41 juvenile leaf	69	189	87	11						
42 adult leaf	237	168	129	8						
43 petiole	30	211	93	17						
44 senescent leaf	3	109	35	25						
5 roots	208	159	140	10						
51 primary root	0	0	0	0		n/a				
52 lateral root	6	240	74	33						
53 root hair	0	0	0	0		n/a				
54 root tip	0	0	0	0		n/a				
55 elongation zone	6	204	67	30						

Anatomy	# of Chips	Mean	Std-Dev	Std-Err	254595_at/At4g18960 (linear)				
					0	975	1950	2925	3900
0 callus	6	3767	667	298					
1 cell suspension	46	2357	1997	298					
2 seedling	341	238	298	16					
21 cotyledons	3	130	8	6					
22 hypocotyl	3	106	13	9					
23 radicle	3	140	4	3					
3 inflorescence	204	3740	5887	413					
31 flower	74	8819	7110	832					
311 carpel	6	139	34	15					
312 petal	6	18957	3043	1361					
313 sepal	6	835	221	99					
314 stamen	14	9107	10852	3010					
315 pedicel	3	137	39	27					
32 silique	17	336	202	50					
33 seed	54	274	277	38					
331 embryo	18	485	377	92					
34 stem	19	1305	2027	478					
35 node	3	97	13	9					
36 shoot apex	26	1171	2088	418					
37 cauline leaf	3	78	15	11					
4 rosette	494	212	189	9					
41 juvenile leaf	69	186	84	10					
42 adult leaf	237	164	112	7					
43 petiole	30	214	89	16					
44 senescent leaf	3	94	6	4					
5 roots	208	293	150	10					
51 primary root	0	0	0	0	n/a				
52 lateral root	6	136	43	19					
53 root hair	0	0	0	0	n/a				
54 root tip	0	0	0	0	n/a				
55 elongation zone	6	125	48	21					

Anatomy	# of Chips	Mean	Std-Dev	Std-Err	254595_at/At4g18960 (linear)				
					0	975	1950	2925	3900
0 callus	6	128	48	21					
1 cell suspension	46	47	30	4					
2 seedling	341	75	49	3					
21 cotyledons	3	82	20	14					
22 hypocotyl	3	49	5	3					
23 radicle	3	43	15	10					
3 inflorescence	204	1206	1626	114					
31 flower	74	2418	2058	241					
311 carpel	6	2344	620	277					
312 petal	6	69	39	17					
313 sepal	6	118	55	24					
314 stamen	14	2905	3399	943					
315 pedicel	3	26	26	18					
32 silique	17	1554	415	104					
33 seed	54	283	411	56					
331 embryo	18	439	657	159					
34 stem	19	482	726	171					
35 node	3	34	14	10					
36 shoot apex	26	304	521	104					
37 cauline leaf	3	64	12	9					
4 rosette	494	81	55	2					
41 juvenile leaf	69	110	80	10					
42 adult leaf	237	85	66	4					
43 petiole	30	113	110	20					
44 senescent leaf	3	95	15	10					
5 roots	208	51	45	3					
51 primary root	0	0	0	0	n/a				
52 lateral root	6	107	67	30					
53 root hair	0	0	0	0	n/a				
54 root tip	0	0	0	0	n/a				
55 elongation zone	6	59	34	15					

Anatomy	# of Chips	Mean	Std-Dev	Std-Err	252886_at/At4g39350 (linear)				
					0	2750	5500	8250	11000
0 callus	6	888	87	39					
1 cell suspension	46	1136	452	67					
2 seedling	341	4127	1563	85					
21 cotyledons	3	3229	41	29					
22 hypocotyl	3	4232	171	121					
23 radicle	3	3526	55	39					
3 inflorescence	204	2789	1928	135					
31 flower	74	1760	657	77					
311 carpel	6	1397	246	110					
312 petal	6	1592	580	259					
313 sepal	6	1053	495	221					
314 stamen	14	1489	540	150					
315 pedicel	3	3266	137	97					
32 silique	17	4497	1321	330					
33 seed	54	2431	1399	192					
331 embryo	18	1681	686	166					
34 stem	19	6152	2995	706					
35 node	3	6556	267	189					
36 shoot apex	26	2396	823	165					
37 cauline leaf	3	1890	98	69					
4 rosette	494	3772	1413	64					
41 juvenile leaf	69	3715	1178	143					
42 adult leaf	237	4022	1685	110					
43 petiole	30	4543	1406	261					
44 senescent leaf	3	1142	11	8					
5 roots	208	3294	1945	135					
51 primary root	0	0	0	0	n/a				
52 lateral root	6	8329	3042	1361					
53 root hair	0	0	0	0	n/a				
54 root tip	0	0	0	0	n/a				
55 elongation zone	6	10192	907	406					

Probeset/AGI	0 callus	1 cell suspension	2 seedling	21 cotyledons	22 hypocotyl	23 radicle	3 inflorescence	31 flower	311 carpel	312 petal	313 sepal	314 stamen	315 pedicel	32 silique	33 seed	331 embruo	34 stem	35 node	36 shoot apex	37 cauline leaf	4 rosette	41 juvenile leaf	42 adult leaf	43 petiole	44 senescent leaf	5 roots	52 lateral root	55 elongation zone	Annotation
251898_at/At3g54340																												AP3..	
246072_at/At5g20240																												PI ...	
259372_at/At1g69120																												AP1 ...	
247490_at/At5g61850																												LFY...	
254595_at/At4g18960																												AG ...	
252886_at/At4g39350																												cellulose synthase	

Probeset/AGI																											Annotation
254595_at/At4g18960																											AG
251898_at/At3g54340																											AP3
246072_at/At5g20240																											PI
259372_at/At1g69120																											AP1
247490_at/At5g61850																											LFY
252886_at/At4g39350																											cellulose synthase

VITA

John-David Swanson

Education

Doctor of Philosophy
 The Pennsylvania State University Jun 1999 - Present
 Integrative Biosciences, Ecological and Molecular Plant Physiology
 Minor in Statistics

Master of Science and Technology (Hons), Biology
 The University of Waikato, New Zealand 1998

Diploma of Management Studies
 The University of Waikato, New Zealand 1996

Bachelor of Science and Technology, Biology
 The University of Waikato, New Zealand 1996

Professional Experience

Lock Haven University of Pennsylvania,
 Lock Haven, PA Jan 2005 – May 2005

Visiting Assistant Professor of Biology
 Pfizer Inc. Groton, CT Jul 2001 – Sept 2001

Summer Intern
 The Pennsylvania State University Jun 1998 – Jun 1999

Senior Research Technologist
 Fletcher Challenge Forests, Te Teko, NZ Sep 1996 – May 1998
 Technician

Tasman Forestry, Te Teko, NZ Nov 1994 – Jul 1995
 Student Intern

New Zealand Forest Research, Rotorua, NZ Nov 1993 – Mar 1994
 Student Intern